



ICCBEI 2025 Hong Kong, China

Proceedings of the 6th
International Conference on Civil
and Building Engineering
Informatics

Edited by
Prof. Jack C.P. Cheng, Prof. Nobuyoshi Yabuki, Prof. Yantao Yu
ICCBEI 2025 Organizing Committee

All rights reserved © 2025

No part of this work may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without written permission from the Publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work.

Published at the Hong Kong University of Science and Technology, Hong Kong, by the Hong Kong University of Science and Technology.



Board of Directors

Asian Group of Civil Engineering Informatics (AGCEI)

President:

Nobuyoshi Yabuki, Osaka University

Members:

Jack Chin Pang Cheng, Hong Kong University of Science and Technology

Patrick Shang-Hsien Hsieh, National Taiwan University

Sang-Ho Lee, Yonsei University

Veerasak Likhitrungsilp, Chulalongkorn University

Zhiliang Ma, Tsinghua University

Koji Makanae, Miyagi University

Xiangyu Wang, Curtin University

Yang Zou, Auckland University

Qti cph lpi Committee

Chair:

Jack C.P. Cheng (Chair), The Hong Kong University of Science and Technology

Vice chair:

Nobuyoshi Yabuki (Co-Chair), Osaka University

Yantao Yu (Co-Chair), The Hong Kong University of Science and Technology

Members:

O qwo kc'F cu."Vj g"J qpi "Mqpi "Wpkxgtuk\ "qh'Uelgpeg"cpf "Vgej pqm {
Vwt dg"Z lpi { w"Vcq."Vj g"J qpi "Mqpi "Wpkxgtuk\ "qh'Uelgpeg"cpf "Vgej pqm { "
Lgppkgt'Lk { lpi \ j cpi ."Vj g"J qpi "Mqpi "Wpkxgtuk\ "qh'Uelgpeg"cpf "Vgej pqm { "
Tcppkg'l vs lpi "Z w."Vj g"J qpi "Mqpi "Wpkxgtuk\ "qh'Uelgpeg"cpf "Vgej pqm { "
Gtle\ j gp { w"Nkcpi ."Vj g"J qpi "Mqpi "Wpkxgtuk\ "qh'Uelgpeg"cpf "Vgej pqm { "
Y kpe { "Y gprk'Nkw."Vj g"J qpi "Mqpi "Wpkxgtuk\ "qh'Uelgpeg"cpf "Vgej pqm { "
\ j cqzlp\ j cpi ."Vj g"J qpi "Mqpi "Wpkxgtuk\ "qh'Uelgpeg"cpf "Vgej pqm { "
\ cqrp'Rcp."Vj g"J qpi "Mqpi "Wpkxgtuk\ "qh'Uelgpeg"cpf "Vgej pqm { "
J cqlwp'Nwq."Vj g"J qpi "Mqpi "Wpkxgtuk\ "qh'Uelgpeg"cpf "Vgej pqm { "
[wpr gpi "Y cpi ."Vj g"J qpi "Mqpi "Wpkxgtuk\ "qh'Uelgpeg"cpf "Vgej pqm { "
Y gk j cq"Vcpi ."Vj g"J qpi "Mqpi "Wpkxgtuk\ "qh'Uelgpeg"cpf "Vgej pqm { "
[wugp"Y cpi ."Vj g"J qpi "Mqpi "Wpkxgtuk\ "qh'Uelgpeg"cpf "Vgej pqm { "

Scientific Committee

Chair:

Jack C.P. Cheng (Chair), The Hong Kong University of Science and Technology

Vice chair:

Nobuyoshi Yabuki (Co-Chair), Osaka University

Yantao Yu (Co-Chair), The Hong Kong University of Science and Technology

Members:

Robert Amor, University of Auckland

Sutee Anantsuksomsri, Chulalongkorn University

Farrukh Arif, NED University of Engineering and Technology Karachi

Habeb Astour, Erfurt University of applied sciences

Niels Bartels, TH Köln Technische Universität

André Borrmann, Technical University of Munich

Frederic Bosche, University of Edinburgh

Conrad Botton, École de Technologie Supérieure

Alexander Buttgeit, Jade University of Applied Sciences

Jiayu Chen, Tsinghua University

Albert Y. Chen, National Taiwan University

Weiwei Chen, University College London

Hung lin Chi, The Hong Kong Polytechnic University

Fabiano Correa, Universidade de Sao Paulo

Hiroaki Date, Hokkaido University

Nashwan Dawood, Teesside University

Yichuan Deng, South China University of Technology

Weili Fang, Huazhong University of Science and Technology

Ian Flood, University of Florida

Tomohiro Fukuda, Osaka University

Mani Golparvar-Fard, University of Illinois at Urbana-Champaign

Hongling Guo, Tsinghua University

Brian Guo, University of Canterbury

Timo Hartmann, TU Berlin

Patrick Shang-Hsien HSIEH, National Taiwan University

Zhenzhong Hu, Tsinghua University

Nan Hu, South China University of Technology
Ryuichi Imai, Hosei University
Vincent J.L. Gan, National University of Singapore
Vineet Kamat, University of Michigan
Satoshi Kanai, Hokkaido University
Constantine Katsanis, École de Technologie Supérieure
Hongjo Kim, Yonsei University
Markus König, Ruhr University Bochum
Satoshi Kubota, Kansai University
Sang-Ho Lee, Yonsei University
Ghang Lee, Yonsei University
Heng Li, The Hong Kong Polytechnic University
Xiao Li, The University of Hong Kong
Nan Li, Tsinghua University
Haijiang Li, Cardiff University
Charinee Limsawasd, Chulalongkorn University
Jiarui Lin, Tsinghua University
Wilson Lu, The University of Hong Kong
Xinzheng Lu, Tsinghua University
Xiaowei Luo, City University of Hong Kong
Hanbin Luo, Huazhong University of Science and Technology
Jun Ma, The University of Hong Kong
Koji Makanae, Miyagi University
Chao Mao, Chongqing University
Jürgen Melzner, Bauhaus-Universität Weimar
Tomohiro Mizoguchi, Nihon University
Ivan Mutis, Illinois Institute of Technology
Keiji Nagatani, The University of Tokyo
Wei Pan, The University of Hong Kong
Mi Pan, University of Macau
Pher Errol B. Quinay, University of the Philippines Diliman
Jeff Rankin, University of New Brunswick
Sérgio Scheer, Universidade Federal do Paraná
Joonoh Seo, The Hong Kong Polytechnic University
Brian Sheil, Cambridge University
Kay Smarsly, Hamburg University of Technology

Lucio Soibelman, University of Southern California
Borja García de Soto, New York University Abu Dhabi
Yi Tan, Shenzhen University
Takahiro Tanaka, Hiroshima University
Jingyuan Tang, Chongqing University
Walid Tizani, University of Nottingham
Žiga Turk, University of Ljubljana FGG
Qian Wang, Southeast University
Mingzhu Wang, City University of Hong Kong
Hsi Hsien Wei, The Hong Kong Polytechnic University
Yiwei Weng, The Hong Kong Polytechnic University
Frank Xue, The University of Hong Kong
I-Tung Yang, National Taiwan University of Science and Technology
Xianzhong Zhao, Tongji University
Ying Zhou, Huazhong University of Science and Technology
Zhenhua Zhu, Wisconsin University
Yang Zou, University of Auckland

Foreword

It is our great pleasure to welcome you to the 6th International Conference on Civil and Building Engineering Informatics (ICCBEI 2025), hosted by the Hong Kong University of Science and Technology. We are deeply honored to hold this prestigious event, which continues a long tradition of fostering collaboration and advancing research in civil and building engineering informatics.

The ICCBEI conference stands as a premier global event focusing on advancements in virtual reality, augmented reality, and building information modeling. This year, over 280 unique participants from 70 governments have gathered to explore and share cutting-edge innovations and practical applications of information technologies within the Architectural, Engineering, Construction, and Operation (AECO) sectors. Esteemed keynote speakers from both academia and industry are set to deliver engaging talks on trending topics, offering invaluable insights and knowledge.

Out of 149 submitted papers, a meticulous review process has resulted in the selection of 101 accepted papers. These will be showcased in parallel sessions across eleven key themes: Digital Twins and BIM Technologies, Construction Safety and Worker Well-being, Innovations for Operation and Maintenance, XR (VR/AR/MR) Visualization and Simulation, Construction Robotics and Human-Robot Collaboration, Large Model and Intelligent Building Operation and Maintenance, Sustainable Construction and Development, OpenBIM and Generative Design, AI for Safety and Health in Building Environment, Information Management in Design and Maintenance, AI in Construction: Innovations, Risks, and Best Practices, Education and Training in Construction, Digital Twins for Building and Infrastructure Maintenance, Digital Campus.

The conference also celebrates its tradition of fostering strong connections among participants. While many attendees are loyal veterans of the event, this year welcomes a significant number of newcomers, and efforts are being made to create a welcoming and interactive environment for them to network and engage.

As part of the organizing committee, our mission is to highlight innovative practices, address pressing challenges, and explore opportunities that AECO professionals encounter. We aim to ensure that attendees leave the conference inspired, enriched with knowledge, brimming with fresh ideas, and equipped with new professional connections. We are certain that you will find the conference enriching and that you'll enjoy the vibrant and sophisticated ambiance of Hong Kong's city center.



Prof. Jack C.P. Cheng

Conference Chair
The Hong Kong University of
Science and Technology



Prof. Nobuyoshi Yabuki

Conference Co-chair
Osaka University



Prof. Yantao Yu

Conference Co-chair
The Hong Kong University of
Science and Technology

Sponsors and Supporting Organizations

Special Thanks to



Special Acknowledgement



Sponsors



Governmental Bodies



Professional and Academic Institutes



Table of Contents

Part 1: Digital Twins and BIM Technologies

HETEROGENEOUS GRAPH NEURAL NETWORKS - BASED PAPER DARWINGS AUTOMATIC LAYERING METHOD

Xun Lu (The Hong Kong University of Science and Technology), Yantao Yu (The Hong Kong University of Science and Technology)

2

INVESTIGATING SOLUTIONS TO THE SHORTAGE OF CONSTRUCTION SITE MANAGERS IN JAPAN: A RESEARCH STUDY ON THE INTRODUCTION OF YOUNG BIM ENGINEERS FROM CHINA

Rui Sucheng (Kanazawa University), Koji Makanae (Miyagi University), Liu Jun (University of Shanghai for Science and Technology), Makoto Fujiu (Kanazawa University), Yuma Morisaki (Kanazawa University)

12

BIM TO FEM MODEL CONVERSION OF LOW - RISE RESIDENTIAL BUILDINGS FOR DEBRIS FLOW DAMAGE ANALYSIS

Xuhong Huang (The Hong Kong University of Science and Technology), Yantao Yu (The Hong Kong University of Science and Technology)

20

BIM ADOPTION AND IMPLEMENTATION CHALLENGES IN MONGOLIA

Khuvilai Erdene (Yonsei University), Sung Hoon Kim (TAESUNG SNI,), Sang - Ho Lee (Yonsei University)

32

INTELLIGENT GEOMETRIC QUALITY INSPECTION FOR MULTIPLE PREFABRICATED COMPONENTS IN LARGE-SCALE STORAGE YARD BASED ON BIM-LiDAR-UAV/UGV

Limei Chen (Shenzhen University), Zhigang Guo (Shenzhen Municipal Group Co., Ltd.), Yi Tan (Shenzhen University), Penglu Chen (Shenzhen University)

42

BIM-BASED DISCRETE EVENT SIMULATION FOR EMBODIED CARBON ASSESSMENT

Yuqing Xu (The Hong Kong University of Science and Technology), Xingbo Gong (The Hong Kong University of Science and Technology), Helen H.L. Kwok (The Hong Kong University of Science and Technology), Jack C.P. Cheng (The Hong Kong University of Science and Technology)

54

ENHANCING SCAN-TO-BIM IN BRIDGE ENGINEERING THROUGH POINT CLOUD COMPLETION

Tao Yang (University of Auckland), Yang Zou (University of Auckland), Enrique del Rey Castillo (University of Auckland)

62

DEVELOPING DEFECT EMBEDDED MASONRY H-BIM USING DEEP LEARNING - BASED DETECTION AND SEGMENTATION

Xinyu Tong (Xi'an Jiaotong-liverpool University,), Jingyu Zhou (Xi'an Jiaotong-liverpool University,), Cheng Zhang (Xi'an Jiaotong-liverpool University,), Yiping Dong (Xi'an Jiaotong-liverpool University,)

71

A DIGITAL TWIN - DRIVEN APPROACH FOR DETECTING OVERBREAK AND UNDERBREAK IN DRILL-AND-BLAST TUNNELS

Zhiyang Pan (Southwest Jiaotong University), Yichong Zhang (Tongji University), Yujie Lu (Tongji University), Bing Tang (Zhejiang Quzhou-Lishui Railway Co., Ltd.,)

79

Part 2: Construction Safety and Worker Well-being

CREATION OF A WEB-BASED TOOL FOR THE VISUAL INSPECTION OF BUILDING EQUIPMENT

Angelina Aziz (Ruhr University Bochum), Markus König (Ruhr University Bochum)

89

AUTOMATIC ERGONOMIC ASSESSMENT CONSIDERING AWKWARD POSTURE AND EXTERNAL LOAD

Fei Tian (The Hong Kong University of Science and Technology), Yantao Yu (The Hong Kong University of Science and Technology)

98

EFFECTS OF A PASSIVE BACK SUPPORT EXOSKELETON ON BALANCE WHILE WALKING

Masood Khan (The Hong Kong Polytechnic University), JoonOh Seo (The Hong Kong Polytechnic University), Maxwell Fordjour Antwi-Afari (Aston University), Yuan Zhou (The Hong Kong Polytechnic University)

108

BACK EXOSKELETONS FOR CONSTRUCTION WORKERS: INVESTIGATION OF HUMAN - EXOSKELETON INTERACTION FORCES

Ting Lei (The Hong Kong Polytechnic University), Kelvin Holam Heung (The Hong Kong Polytechnic University), Joon Oh Seo (The Hong Kong Polytechnic University)

115

EEG TECHNOLOGY IN CONSTRUCTION SAFETY RESEARCH: CURRENT TRENDS, CHALLENGES, AND FUTURE DIRECTIONS

Tanghan Jiang (Monash University), Jiayu Chen (Tsinghua University), Yihai Fang (Monash University)

124

ENHANCING TRAFFIC SAFETY IN SMART CITIES THROUGH PERCEIVED RISK ANALYSIS FOR VULNERABLE ROAD USERS

Yuting Zhang (Tsinghua University), Xinyu Pan (Tsinghua University), Jiayu Chen (Tsinghua University)

132

ENHANCING THE PERFORMANCE OF MULTI-OBJECT TRACKING IN TRAFFIC STREAM VIDEOS THROUGH INITIAL VELOCITY AND FRAMESKIPPING STRATEGIES

Chia-Chun Lin (National Taiwan University), Albert Y. Chen (National Taiwan University)

140

AN IMPROVED SAFETY MANAGEMENT METHOD FOR WELDING OPERATIONS BASED ON GAZE ESTIMATION AND OBJECT DETECTION

Ying Liang (Shenzhen University), Yi Tan (Shenzhen University), Dong Jia (China Construction Fifth Engineering Division COPR.LTD.)

148

APPROCHING AN INTERACTIVE RESOLUTION ON NEW CITIZENS' JOB-HOUSING BALANCE AND VACANT HOUSING REDUCTION IN GUANGZHOU

Ziyao Dong (South China University of Technology), Yangbing Zhang (South China University of Technology), Yuejiang Su (South China University of Technology), Yingqiao Fang (South China University of Technology), Yilin Huang (South China University of Technology), Siyang Zheng (South China University of Technology)

159

Part 3: Innovations for Operation and Maintenance

LIGHTWEIGHT TRANSFER LEARNING FOR WATER BODY SEGMENTATION USING ADAPTOR-BASED FINE-TUNING

Jiapei Zhao (Osaka University), Nobuyoshi Yabuki (Osaka University), Tomohiro Fukuda (Osaka University) 168

SEMANTIC-ENRICHED IMAGE RETRIEVAL FOR BRIDGE DAMAGE ASSESSMENT

Chengzhang Chai (Cardiff University), Jiucui Liu (Cardiff University), Yan Gao (Cardiff University), Guanyu Xiong (Cardiff University), Haijiang Li (Cardiff University) 178

ADVANCING BLOCKCHAIN AND SMART CONTRACTS FROM ACADEMIA TO INDUSTRY: INSIGHTS INTO THE CONSTRUCTION SECTOR

Xuling Ye (Ruhr University Bochum), Liu Liu (Ruhr University Bochum), Markus König (Ruhr University Bochum) 187

A STUDY ON THE PERFORMANCE COMPARISON OF SEDIMENT MOVEMENT DETECTORS USING UNET-BASED MODELS AND THE SLIDING PARTITIONING METHOD

Xiaosong Liu (Yamaguchi University), Makoto Oda (Yamaguchi University), Kei Kawamura (Yamaguchi University), Tsuyoshi Wakatsuki (National Research Institute for Earth Science and Disaster Resilience (NIED)) 197

DEVELOPMENT OF 'VOICES OF BRIDGES' FOR INFRASTRUCTURE MANAGEMENT

Yota Chinen (Yamaguchi University), Kei Kawamura (Yamaguchi University), Shuji Sawamura (Yamaguchi University) 206

CB-YOLOV5: STREETLIGHT DETECTION BASED ON LOW-LIGHT IMAGES IN HIGH-INTERFERENCE ENVIRONMENT

Shiqi ZHANG (Chongqing University), Jingyuan TANG (Chongqing University), Jingke HONG (Chongqing University) 216

DEVELOPING A MARKOV DECISION PROCESS MODEL TO SUPPORT ESG DECISION-MAKING IN THE INTERNATIONAL CONSTRUCTION MARKET

Yuting Duan (Southwest Jiaotong University), Meng Ye (Chongqing University), Penglin Dai (Southwest Jiaotong University) 222

INTELLIGENT DESIGN FOR COMPONENT SIZE OF REINFORCED CONCRETE FRAME STRUCTURES USING DIFFUSION MODELS

Yi Gu (Tsinghua University), Sizhong Qin (Tsinghua University), Wenjie Liao (Southwest Jiaotong University), Siqi Chen (Tsinghua University), Xinzhen Lu (Tsinghua University) 231

CONSTRUCTABILITY-BASED MULTI-OBJECTIVE OPTIMIZATION FOR REINFORCING BAR DESIGN IN RECTANGULAR CONCRETE BEAMS

Luis F. Verduzco (The Hong Kong University of Science and Technology), Jack C.P. Cheng (The Hong Kong University of Science and Technology) 239

Part 4:XR (VR/AR/MR), Visualization, and Simulation

AUTOMATIC DETECTION OF BUILDING EQUIPMENT USING LASER INTENSITY OF MLS POINT CLOUD

Arisa Kobayashi (Nihon University), Tomohiro Mizoguchi (Sanyo-Onoda City University)

250

BIM-AR DRIVEN COLLABORATIVE MANAGEMENT OF COMPLEX MEP EQUIPMENT INSTALLATION

Youde Zheng (Shenzhen University), Liyun Yao (Shenzhen University), Silin Li (Shenzhen University), Yugui Zhou (China Construction Fifth Division Southern Construction Subsidiary CORP.LTD), Yi Tan (Shenzhen University)

260

ENHANCED CURTAIN WALL CONSTRUCTION PROGRESS MONITORING VIA TOWER CRANE PERSEPTIVE AND INTEGRATED 3D RECONSTRUCTION

Tao Zhong (Tongji University), Yujie Lu (Tongji University), Shuo Wang (Tongji University), Xuanjun Li (Tongji University)

270

3D INDOOR RECONSTRUCTION BASED ON SPATIAL LAYOUT ESTIMATION USING PANORAMIC INSPECTION VIDEO FROM BUILDING SITES

Shuo Wang (Tongji University), Yujie Lu (Tongji University), Tao Zhong (Tongji University), Lijian Zhong (Tongji University), Yufan Chen (Tongji University)

280

A REAL-WORLD VISUAL SLAM DATASET FOR INDOOR CONSTRUCTION SITES

Wenyu Li (The Hong Kong University of Science and Technology), Xinyu Chen (The Hong Kong University of Science and Technology), Yantao Yu (The Hong Kong University of Science and Technology)

289

SEMANTIC SEGMENTATION OF MEP WITH DEEP LEARNING AND SYNTHETIC POINT CLOUDS

Hongzhe Yue (Southeast University), Qian Wang (Southeast University)

297

HTUNNEL-HLS: A 3D POINT CLOUD DATASET FOR BENCHMARKING CONSTRUCTION HIGH-WAY TUNNEL SEGMENTATION METHODS

Li-zhuang Cui (Shandong University), Lei Kou (Shandong University), Hanming Zhang (Shandong University), Yusen Wang (The Hong Kong University of Science and Technology), Feng Guo (Shandong University), Jian Liu (Shandong University)

305

A SCAN-TO-BIM APPROACH FOR RENOVATING EXISTING BUILDING ROOMS USING SURFACE RECONSTRUCTION AND DEEP LEARNING

May Thin Zar Soe (Osaka University), Nobuyoshi Yabuki (Osaka University), Tomohiro Fukuda (Osaka University), Yoshiro Hada (Tokyu Construction Co., Ltd)

313

VIRTUAL SCANNING METHOD BASED ON BIM AND RAY TRACING FOR OCCLUSION EVALUATION AND GEOMETRY EXTRACTION

Siwei Lin (Shanghai Jiao Tong University), Liping Duan (Shanghai Jiao Tong University), Bin Jiang (Shanghai Jiao Tong University), Jincheng Zhao (Shanghai Jiao Tong University)

323

Part 5: Advancing Construction through Robotics and Human Machine Collaboration

AN EXPERIMENTAL STUDY FOR MULTI-ROBOT COORDINATION IN MULTISTORY BUILDING CONSTRUCTION SITES

Haojun Luo (The Hong Kong University of Science and Technology), Yantao Yu (The Hong Kong University of Science and Technology)

332

UNDERSTANDING WORKERS' ACCEPTANCE OF ROBOTICS IN CONSTRUCTION

Zhida Zhang (University of Macau), Fanfan Meng (University of Macau), Mun On Wong (University of Macau), Mi Pan (University of Macau)

343

ASSESSMENT OF AUTOMATION POTENTIAL FOR ON-SITE CONSTRUCTION TASKS

Bing Sun (Chongqing University), Chao Mao (Chongqing University), Wen Yi (The Hong Kong Polytechnic University), Tingpeng Wang (Chongqing University)

353

A SIMULATION-BASED GENETIC ALGORITHM SCHEDULE OPTIMIZATION METHOD FOR BRIDGE CONSTRUCTION

Yuan Yao (Western Sydney University), Vivian WY Tam (Western Sydney University), Jun Wang (Western Sydney University), Khoa N Le (Western Sydney University), Anthony Butera (Western Sydney University), Wenchi Shou (Western Sydney University)

361

TOWARDS HIGHER INFORMATION DENSITY IN IMAGE TRANSMISSION: LEARNED IMAGE COMPRESSION FOR CONSTRUCTION SITE MONITORING

Jiucan Liu (Cardiff University), Chengzhang Chai (Cardiff University), Linghan Ouyang (Cardiff University), Haijiang Li (Cardiff University)

367

TEACHING ROBOT END EFFECTORS TO GRASP CONSTRUCTION TOOLS BASED ON DEEP REINFORCEMENT LEARNING

Xiaohu Yu (The Hong Kong University of Science and Technology), Yantao Yu (The Hong Kong University of Science and Technology)

377

MULTI-OBJECTIVE OPTIMIZATION OF TOWER CRANE LAYOUT PLANNING IN MODULAR INTEGRATED CONSTRUCTION CONSIDERING EFFICIENCY, COST, AND LIFTING SAFETY

Fangzheng LI (The Hong Kong Polytechnic University), Hung-Lin CHI (The Hong Kong Polytechnic University), Rongyan LI (The Hong Kong Polytechnic University), Roy Dong Wang (The Hong Kong Polytechnic University), Tarek Zayed (The Hong Kong Polytechnic University), Bo Xiao (Michigan Technological University)

386

A STUDY OF HUMAN-ROBOT COLLABORATIVE CONSTRUCTION TASK ALLOCATION PROBLEM CONSIDERING THE LEVEL OF TASK DANGER

Yilin Wang (Chongqing University), Chao Mao (Chongqing University), Tingpeng Wang (Chongqing University)

395

ENERGY CONSUMPTION MODELING FOR A ROBOT ARM IN 3D CONCRETE PRINTING

Tong Liu (The Hong Kong Polytechnic University), Song Du (The Hong Kong Polytechnic University), Fei Teng (The Hong Kong Polytechnic University), Yiwei Weng (The Hong Kong Polytechnic University)

404

Part 6: Large Model and Intelligent Building Operation and Maintenance

ENHANCING KNOWLEDGE CAPTURE AND REUSE FOR OFFSITE CONSTRUCTION THROUGH A CONTAINER-BASED KNOWLEDGE APPROACH

Zhen Zhang (University of Auckland), Yang Zou (University of Auckland), Brian H.W.Guo (University of Canterbury), Lixin Jiang (University of Auckland), Johannes Dimyadi (Codify Asset Solutions Ltd), Roy Davies (University of Auckland)

410

AUTOMATED TEXT CLASSIFICATION OF CONSTRUCTION INSPECTION REPORT: A SMALL SAMPLES TRAINING APPROACH

Kai Li (Wuhan University of Technology), Chao Dong (Wuhan University of Technology), Xueqing Fang (Wuhan University of Technology), Da Li (Henan University of Science and Technology)

420

AN LLM AGENT-BASED APPROACH FOR CALCULATING FLOOR AREAS VIA BIM IFC FILES

Xueying Zhu (The University of Hong Kong), Jun Ma (The University of Hong Kong)

430

HERITAGE BUILDING INFORMATION MANAGEMENT AND INTELLIGENT QUERYING BY MULTIMODAL LARGE LANGUAGE MODELS AND KNOWLEDGE GRAPH

Jiaying ZHANG (The Hong Kong University of Science and Technology), Jeff Chak Fu CHAN (The Hong Kong University of Science and Technology), Ziyu ZHAO (The University of Western Australia), Jack C.P. CHENG (The Hong Kong University of Science and Technology)

439

BRIDGE INSPECTION USING A MULTI-MODAL VISION LANGUAGE MODEL

Zhengxing Chen (University of Auckland), Yang Zou (University of Auckland), Vicente A.Gonzalez (University of Alberta), Jason Ingham (University of Auckland), Liam M.Wotherspoon (University of Auckland)

447

AN INTERACTIVE LLM-BASED FRAMEWORK FOR SPATIAL RELATIONSHIP QUERY ON OPEN-BIM ELEMENTS WITH IFC4 SCHEMA

Ang Li (The Hong Kong University of Science and Technology), Peter Kok-Yiu Wong (The Hong Kong University of Science and Technology), Xingyu Tao (The Hong Kong University of Science and Technology), Jack C.P.Cheng (The Hong Kong University of Science and Technology)

455

OPTIMIZED LANGUAGE-EMBEDDED 3DGS FOR REALISTIC MODELING AND INFORMATION STORAGE OF HISTORICAL BUILDINGS

Zhenyu LIANG (The Hong Kong University of Science and Technology), Jeff Chak Fu CHAN (The Hong Kong University of Science and Technology), Jiaying ZHANG (The Hong Kong University of Science and Technology), Zhaolun LIANG (The Hong Kong University of Science and Technology), Boyu WANG (The Hong Kong University of Science and Technology), Mingzhu WANG (City University of Hong Kong), Jack C.P. CHENG (The Hong Kong University of Science and Technology)

463

SUSTAINABLE DEVELOPMENT MANAGEMENT METHOD FOR CONSTRUCTION PROJECTS USING LARGE LANGUAGE MODEL

Xingbo Gong (The Hong Kong University of Science and Technology), Yuqing Xu (The Hong Kong University of Science and Technology), Helen H.L. Kwok (The Hong Kong University of Science and Technology), Xingyu Tao (The Hong Kong University of Science and Technology), Jack C.P. Cheng (The Hong Kong University of Science and Technology)

471

AUTOMATIC COMPLIANCE CHECKING OF BIM MODELS FOR ARCHITECTURE AND FIRE PROTECTION BASED ON KNOWLEDGE GRAPHS AND MACHINE LEARNING

Sihao Li (Southeast University), Guangyao Chen (Southeast University), Yangze Liang (Southeast University), Zhao Xu (Southeast University)

480

Part 7: Sustainable Construction and Development

DEVELOPING GNN-BASED SURROGATE MODELS FOR MULTI-OBJECTIVE SUSTAINABLE PERFORMANCE PREDICTIONS OF RESIDENTIAL BLOCKS

Zhaoji Wu (The Hong Kong University of Science and Technology), Wenli Liu (The Hong Kong University of Science and Technology), Jack C.P. Cheng (The Hong Kong University of Science and Technology), Zhe Wang (The Hong Kong University of Science and Technology), Helen H.L. Kwok (The Hong Kong University of Science and Technology), Cong Huang (The Hong Kong University of Science and Technology), Fangli Hou (The Hong Kong University of Science and Technology)

489

GEOSPATIAL GRAPH ATTENTION NETWORK FOR HIGH-RESOLUTION BUILDING FACADE PHOTOVOLTAIC POTENTIAL PREDICTION

Zheng Li (The University of Hong Kong), Jun Ma (The University of Hong Kong)

499

REAL-TIME SENSOR-BASED CHARACTERIZATION OF RECYCLED COARSE AGGREGATES (RCA): ADVANCING SUSTAINABLE CONSTRUCTION THROUGH AUTOMATED QUALITY ASSESSMENT

Cheng Chang (Delft University of Technology), Francesco Di Maio (Delft University of Technology), Peter Rem (Delft University of Technology)

505

EXPLORING RIVERINE LITTER DETECTION BY DEVELOPING COMPREHENSIVE DATASET AND DEEP LEARNING

Xiaohan Xu (Xi'an Jiaotong-Liverpool University), Cheng Zhang (Xi'an Jiaotong-Liverpool University), Yunfei Xia (University of Liverpool), Xiaohui Zhu (Xi'an Jiaotong-Liverpool University), Peter Burgess (University of Liverpool)

514

MAPPING AIR POLLUTION CONCENTRATIONS ON SIDEWALKS IN CENTRAL BUSINESS DISTRICTS BASED ON MOBILE MONITORING

Xujing Yu (The University of Hong Kong), Jun Ma (The University of Hong Kong)

523

BLOCKCHAIN-DRIVEN TRANSPARENCY IN CONSTRUCTION TENDERS: SMART CONTRACTS AND SUSTAINABLE WASTE MANAGEMENT

Silvia Meschini (University of Turin), Daniele Accardo (University of Turin), Lavinia Chiara Tagliabue (University of Turin), Stefano Rinaldi (University of Brescia), Rosa Meo (University of Turin), Andrea Bracciali (University of Turin), Giuseppe Martino Di Giuda (University of Turin)

530

ENHANCING RESILIENCE AND ADAPTABILITY OF INFRASTRUCTURE TRANSMISSION AND DISTRIBUTION NETWORKS: A REVIEW STUDY

S Mohammad Saleh Nikoopayan Tak (New Jersey Institute of Technology), Reyhaneh Moazami Goodarzi (Iran University of Science and Technology), Yanxiao Feng (New Jersey Institute of Technology)

540

Part 8: OpenBIM and Generative Design

STREAMING SENSOR REGISTRATION AND UPDATING PROCESS IN BIM-BASED DIGITAL TWINS

Cheng-Wen Fu (National Taiwan University), Tzong-Hann Wu (National Taiwan University), Yun-Tsui Chang (National Taiwan University), Shang-Hsien Hsieh (National Taiwan University)

551

LEVERAGING TRANSFORMERS FOR IMPROVED DECISION-MAKING IN ROAD MAINTENANCE

Rui Kang (University of Cambridge), Stephen Green (University of Cambridge), Ioannis Brilakis (University of Cambridge)

561

EFFICIENT ACCESSIBILITY INTEGRATED MULTIOCCUPANCY BUILDING LAYOUT GENERATION USING CONSTRAINED DIFFUSION MODELS

Haolan Zhang (Virginia Polytechnic Institute and State University), Ruichuan Zhang (Virginia Polytechnic Institute and State University)

570

MULTI-STORY FLOOR PLAN GENERATION FROM BUILDING VOLUMETRIC DESIGN USING GRAPH NEURAL NETWORK

Atsuya Ogura (Osaka University), Tomohiro Fukuda (Osaka University), Nobuyoshi Yabuki (Osaka University), Taro Narahara (New Jersey Institute of Technology)

579

IFC-BASED INDOOR SPACE EXTRACTION WITH TOPOLOGICAL GRAPH MODELING

Zheng Zhao (Huazhong University of Science and Technology), Weiya Chen (Huazhong University of Science and Technology), Zhiyuan Guo (Huazhong University of Science and Technology), Wenming Jiang (Huazhong University of Science and Technology)

589

THE ROLE OF DIGITAL ECONOMY IN THE HIGH-QUALITY DEVELOPMENT OF THE CONSTRUCTION INDUSTRY: EVIDENCE FROM CHINA

Aoqi Huang (Chongqing University), Hongjuan Wu (Chongqing University), Taozhi Zhuang (Chongqing University)

598

Part 9: AI for Safety and Health in Building Environment

TRAJECTORY FORECASTING FOR WORKER SAFETY IN CONSTRUCTION USING TRANSFORMER AND GRAPH ATTENTION NETWORKS

Mohammed Alduais (University of Alberta), Xinming Li (University of Alberta), Qipei Mei (University of Alberta)

609

A SELF-ATTENTION FUSION-BASED BI-LSTM FRAMEWORK FOR OCCUPANT-CENTRIC PREDICTION OF INDOOR ENVIRONMENTAL FACTORS

Min Jae Lee (Virginia Polytechnic Institute and State University), Ruichuan Zhang (Virginia Polytechnic Institute and State University)

618

INTEGRATED FRAMEWORK FOR COMPLIANCE CHECKING AND PERFORMANCE EVALUATION IN BUILDING DESIGN

Jin Han (Tsinghua University), Zhe Zheng (Tsinghua University), Zhen Xu (University of Science and Technology Beijing), Xin-Zheng Lu (Tsinghua University), Jia-Rui Lin (Tsinghua University)

627

VISUAL-DRIVEN INSPECTION FOR COLLECTING THE STATUS OF FIRE SAFETY EQUIPMENT IN BUILDING SPACES

Fangzhou Lin (The Hong Kong University of Science and Technology), Zhengyi Chen (The Hong Kong University of Science and Technology), Xiao Zhang (The Hong Kong University of Science and Technology), Boyu Wang (The Hong Kong University of Science and Technology), Jack C.P. Cheng (The Hong Kong University of Science and Technology)

636

VISION LLM-DRIVEN OPERATIONAL HAZARD RECOGNITION FOR BUILDING FIRE SAFETY COMPLIANCE CHECKING

Dayou Chen (Loughborough University), Long Chen (City University of Hong Kong), Yiheng Zeng (University College London), Craig Hancock (Loughborough University), Russell Lock (Loughborough University), Simon Sølvsten (University of Southern Denmark)

646

A KNOWLEDGE GRAPH BASED COMMON DATA ENVIRONMENT SOLUTION FOR BUILDING SEISMIC LOSS ESTIMATION

Zeyu PAN (Shanghai Jiao Tong University), Jianyong SHI (Shanghai Jiao Tong University)

655

Part 10: Information Management in Design and Maintenance

PROBABILITY MAP GUIDED POINT RENDERING TECHNIQUE FOR REFINED SEGMENTATION OF HIGH-RESOLUTION CRACK IMAGES

Honghu Chu (Hunan University), Weiwei Chen (University College London), Lu Deng (Hunan University) 666

MAINTENANCE MANAGEMENT - A MODULE IN ASSET MANAGEMENT IN A MEDIUM-SIZED GERMAN CITY AS AN EXAMPLE

Alexander Buttgerit (Jade University of Applied Sciences), Stefan Gomolluch (City of Münster) 676

EXAMINING FACTORS OF HUMAN-AI TRUST: COMPARING HUMAN- AND AI-RELATED FACTORS

Xinyue Zhou (Chang'an University), Sheng Xu (Chang'an University) 686

A DOMAIN KNOWLEDGE-ENHANCED LARGE VISION-LANGUAGE MODEL FOR CONSTRUCTION SITE SAFETY MONITORING

Chak-Fu Chan (The Hong Kong University of Science and Technology), Xiaowen Guo (The Hong Kong University of Science and Technology), Peter Kok-Yiu Wong (The Hong Kong University of Science and Technology), Jolly Pui-Ching Chan (Drainage Services Department), Jack C.P. Cheng (The Hong Kong University of Science and Technology), Pak-Him Leung (AutoSafe Ltd), Xingyu Tao (The Hong Kong University of Science and Technology) 694

EXPLORING CONSTRUCTION 5.0 PARADIGMS IN THE AEC SECTOR OF NEW ZEALAND: CONCEPTUAL FOUNDATIONS AND IMPLICATIONS FOR PRACTICE

Xichen Chen (Auckland University of Technology), Ali Ghaffarianhoseini (Auckland University of Technology), Amirhosein Ghaffarianhoseini (Auckland University of Technology) 704

RESEARCH ON A METHODOLOGY OF AI-DRIVEN SUBSTATION LAYOUT DESIGN BASED ON LARGE LANGUAGE MODEL

Jianyong SHI (Shanghai Jiao Tong University), Zeyu PAN (Shanghai Jiao Tong University), Longfei PAN (Shanghai Jiao Tong University), Junpeng Hu (Shanghai Jiao Tong University) 714

Part 11: AI in Construction: Innovations, Risks, and Best Practices

CLOUD TYPE CLASSIFICATION THROUGH SEMANTIC SEGMENTATION FOR ANALYSIS OF EARTH'S RADIATIVE BALANCE

Yu-Wen Wu (National Taiwan University), Nofel Lagrosas (Kyushu University, National Central University), Albert Y. Chen (National Taiwan University)

723

TRANS-DIMENSIONAL BAYESIAN INVERSION FOR ESTIMATING EARTH PRESSURES ON IN-SERVICE UNDERGROUND STRUCTURES

Zhiyao Tian (City University of Hong Kong), Xianfei Yin (City University of Hong Kong)

731

PROPOSAL OF THE USE OF DIFFERENTIAL RENDER IN REBAR INSPECTIONS

Yoshihiro Yasumuro (Kansai University), Youta Tanaka (Kansai University), Takeshi Takai (Takenaka Corporation), Satoshi Kubota (Kansai University)

740

MACHINE LEARNING-ASSISTED PREDICTION OF STEAM-CURED CONCRETE FOR 3D PRINTING APPLICATIONS

Junbo Sun (Curtin University), Yufei Wang (Curtin University), Xianda Liu (East China Jiao Tong University), Jiajie Shang (Curtin University), Wenfu Zhang (East China Jiao Tong University), Xiangyu Wang (Nanjing Institute of Technology, East China Jiao Tong University)

746

THE GREAT MISALIGNMENT IN CONSTRUCTION CONTRACTING: BEST PRACTICES AND SOFTWARE

Matt Stevens (Western Sydney University), Wei Zhou (Western Sydney University), Babatunde Fatai Ogun-bayo (University of Johannesburg)

752

EXPLORING THE RISKS OF INTEGRATING GENERATIVE ARTIFICIAL INTELLIGENCE INTO CONSTRUCTION RISK MANAGEMENT: INSIGHTS FROM A SYSTEMATIC LITERATURE REVIEW

Mohamed Mohamed (Teesside University), M.K.S Al-Mhdawi (Teesside University, Trinity College Dublin), Farzad Rahimian (Teesside University), Udechukwu Ojiako (University of Strathclyde, University of Hull, Johannesburg Business School), Charf Mahammedi (Teesside University)

761

Part 12: Digital Twins for Building and Infrastructure Maintenance

OPTIMIZING HYDROPOWER OPERATIONS: A TAILORED DIGITAL TWIN FRAMEWORK FOR HYDROPOWER PLANT

Weiwei Chen (University College London), Yang Su (University College London)

767

PROPOSAL FOR A SIMPLE INSPECTION TOOL USING HOLOLENS2

Tomohiro Mizoguchi (Sanyo-Onoda City University)

778

A PRACTICE IN CHINA FOR COMPUTER AIDED DESIGN AND CONSTRUCTION OF PREFABRICATED ENCLOSURE WALLS

Shen Wei (Beijing Yjk Building Software Co., Ltd.), Liu Tao (Beijing Yjk Building Software Co., Ltd.), Wang Guangjian (Beijing Yjk Building Software Co., Ltd.), Yu Hu (Beijing Yjk Building Software Co., Ltd.), Gao Hang (Beijing Yjk Building Software Co., Ltd.)

784

A BIM COMPONENT-CENTRED BRIDGE DIGITAL TWIN FOR SMART AND PRACTICAL BRIDGE MAINTENANCE

Guanyu Xiong (Cardiff University), Yan Gao (Cardiff University), Chengzhang Chai (Cardiff University), Kehong Chen (Cardiff University), Haijiang Li (Cardiff University)

795

A DISTRIBUTED TWIN SYSTEM FOR MANAGING ASSET-PROCESS INTERACTIONS IN HIGH-WAY INFRASTRUCTURE SYSTEMS

Mengtian Yin (University of Cambridge), Tom Kelly (University of Cambridge), Pieter Pauwels (Eindhoven University of Technology), Liming Xu (University of Cambridge), Brian Sheil (University of Cambridge), Junxiang Zhu (University of Cambridge), Yue Xie (University of Cambridge), Ioannis Brilakis (University of Cambridge)

802

FINITE ELEMENT MODEL UPDATING ENHANCED BY DEEP LEARNING FOR REINFORCED CONCRETE CONTAINMENT VESSEL DIGITAL TWIN

Jiaji Wang (University of Hong Kong), Si Fu (University of Hong Kong), Chawit Kaewnuratchadasorn (University of Hong Kong)

813

Part 13: Education and Training in Construction

LIMITED DATA-ORIENTED WORKER INTENTION RECOGNITION METHOD IN WORKER-ROBOT COLLABORATION FOR CONSTRUCTION

Zhaoxin Zhang (The Hong Kong University of Science and Technology), Zaolin Pan (The Hong Kong University of Science and Technology), Yantao Yu (The Hong Kong University of Science and Technology), Liang Liu (Hong Kong INDICS International Information Technology Co., Ltd)

820

SIMULATION OF RESPONSE TIME REDUCTION OF OHCA CONSIDERING DRONE BASE SELECTION FOR AED DELIVERY

Yu Wu (National Taiwan University), Kuan-Chen Chin (Taipei Hospital, Ministry of Health and Welfare), Jen-Tang Sun (Far Eastern Memorial Hospital), Ying-Chih Ko (National Taiwan University Cancer Center), Wen-Chu Chiang (National Taiwan University Hospital, Yun-Lin Branch), Mathew Huei-Ming Ma (National Taiwan University Hospital, Yun-Lin Branch), Albert Y.C. Chen (National Taiwan University)

829

EVALUATING THE IMPACT OF PROBLEM-BASED LEARNING AND DIGITAL MODELS ON CIVIL ENGINEERING EDUCATION

Hamidreza Alavi (University of Cambridge), Brian Sheil (University of Cambridge), Ioannis Brilakis (University of Cambridge)

837

ENHANCING TOWER CRANE OPERATION SKILLS IN AERIAL CONSTRUCTION FACTORIES THROUGH VIRTUAL REALITY TRAINING: A CONTROLLED STUDY

Jingfeng Yuan (Southeast University), Lingxiao Wang (Southeast University), Feixiang Hu (Southeast University), Yang Su (Southeast University), Chen Li (Southeast University), Hao Tian (Southeast University), Weidong Lin (Southeast University)

842

CONSTRUCTION EQUIPMENT MOTION RECOGNITION USING BLOCKCHAIN CYBERSECURITY-AWARE DECENTRALIZED MACHINE LEARNING FRAMEWORK FOR CONSTRUCTION

Chengliang Zheng (Wuhan University), Xingyu Tao (The Hong Kong University of Science and Technology), Jiarui Lin (Tsinghua University), Moumita Das (The Hong Kong University of Science and Technology), Wenchi Shou (Western Sydney University), Jack C.P. Cheng (The Hong Kong University of Science and Technology)

853

BUILDING DECAY IN HONG KONG: ASSESSING AND MAPPING BUILDING CONDITION INDEX USING CHATGPT AND AIRBORNE POINT CLOUDS

Wai Hung LEE (The University of Hong Kong), Fan XUE (The University of Hong Kong)

861

Part 14: Digital Campus

A SCHEME AND APPLICATION OF A DIGITAL TWIN-BASED ROAD PAVEMENT MANAGEMENT SYSTEM

Koji Makanae (Miyagi University), Yogo Kurokawa (Ikee Co., Ltd.), Kiyotaka Suda (Kankyo Fudo Techno Co., Ltd), Jevica (Construction IoT Research Lab), NurulNajwa Binti Khamis (Construction IoT Research Lab)

872

DEVELOPMENT AND VALIDATION OF THE GAME ENGINE-BASED DIGITAL TWIN APP "CONNECTIA" FOR PRACTICAL USE IN THE CONSTRUCTION PHASE

Tomohide Yuasa (Obayashi Corporation), Kodai Yamamoto (Obayashi Corporation)

879

DESIGN STRATEGIES FOR ENHANCING SPATIAL VITALITY THROUGH THE INTEGRATION OF SPORTS VENUES AND COMMERCIAL SPACES

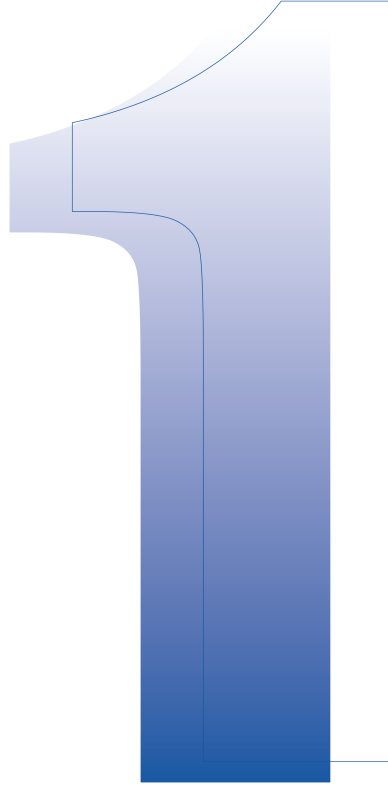
Xuefei Yuan (Jiaxing University, Shanghai Cuiyunyuan Architectural Design Consulting Co., Ltd), Benchen Fu (Shenzhen University, Shenzhen Key Laboratory of Architecture for Health & Well-being), Shan Man (Jiaxing University), Yuanyuan Wei (Jiaxing University)

889

REAL-TIME ON-SITE INDOOR MODELING AND QUALITY VISUALIZATION USING MIXED REALITY DEVICE WITH TOF SENSOR

Norihiko Goto (Hokkaido University), Hiroaki Date (Hokkaido University), Satoshi Kanai (Hokkaido University)

899



Part I

Digital Twins and BIM Technologies

HETEROGENEOUS GRAPH NEURAL NETWORKS-BASED PAPER DRAWINGS AUTOMATIC LAYERING METHOD

Xun Lu¹, Yantao Yu^{2,*}

1) Ph.D. Student, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: xun.lu@connect.ust.hk

2) Ph.D., Assis. Prof., Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: ceyantao@ust.hk

(* Corresponding author)

Abstract: Building Information Modelling (BIM) can significantly improve aging buildings' smart operation and maintenance (O&M) efficiency, and support advanced equipment maintenance and emergency response. However, many aging buildings lack BIM due to the era in which they were built, and the manual reconstruction process requires experts with specialized knowledge and consumes considerable modeling time. Many existing studies utilize deep learning methods based on image features to extract component information from paper drawings for BIM reconstruction, however, the paper drawing layering pre-processing, which is essential for improving the information extraction accuracy, still requires inefficient manual drafting and line classification. Although existing studies have proposed methods for drawings layering by line classification, these methods perform poorly on construction engineering drawings due to the complexity and line dense in drawings. To fill this gap, we propose a heterogeneous graph neural networks (GNNs) based method that predicts the category of line elements on paper drawings to achieve automatic layering with three modules: 1) paper drawing vectorization by line extraction and duplicate elements merging to detect the line elements representing component contour and annotation; 2) graph structure construction by considering the different topological relationships among lines to represent the line properties; 3) heterogeneous graph nodal classification model to predict the line category and realize automatic layering. The proposed method was tested on an actual engineering drawing dataset, and the results show that the method has an overall F_1 score more than 0.74 and exceeds the baseline model by over 0.1. This research improves paper drawing pre-processing efficiency and provides a new solution for information extraction in ageing buildings BIM reconstruction.

Keywords: Building information modeling, Graph Neural Networks, Drawing processing, Information extraction, 3D reconstruction

1. INTRODUCTION

Building information model (BIM) integrates building structural architectural and facility management data from diverse data sources and it is the foundation of and information of building smart operation (Hu et al., 2022). Many existing studies apply BIM to energy retrofit (Kamel & Memari, 2019), emergency response (Feng et al., 2021) and facility management (Hu et al., 2016), etc., which shows BIM's great potential to improve ageing building O&M efficiency and reduce cost. Despite it benefits, the ageing buildings built more than 20-30 years ago often lack BIM due to the technology limitations, and the manual BIM reconstruction process is knowledge- and labor- intensive (de Wilde, 2019), because it requires experts with specific knowledge to collect building information from original building documents. To improve ageing building's BIM reconstruction efficiency, the automatic information extraction and BIM reconstruction methods are required.

Building drawings contain accurate geometric and semantic information about building components (Yin et al., 2020). For ageing building BIM reconstruction, paper drawings are an ideal data source, because they are stored throughout the ageing building's life cycle as part of the deliverables of building completion. What's more, paper drawings have lower acquisition costs and there are no data privacy issues caused by on-site data collection processes. Many existing studies have been devoted to extracting information from raster drawings obtained by scanning or photocopying paper drawings for BIM reconstruction (Pan et al., 2023). Since different components are represented on the same plan by discrete pixels, some factors such as occlusion and overlapping between components will affect the accuracy of information extraction. It is necessary to divide building components of different categories and sizes into layers, thereby improving BIM reconstruction performance (Pan et al., 2023; Zhao et al., 2021). However, the layering pre-processing requires converting paper drawings into vector drawings (such as CAD drawings) by manual drafting, which is low efficient. Considering that most ageing buildings typically lack vectorized format drawings, there is an urgent need for an effective method to layer paper drawings according to component categories automatically.

Line elements represent most of the building component contour and annotations. Substantial research (Fan et al., 2021; Kim et al., 2021; Xie et al., 2022) has been conducted to classify line elements categories on drawings. Some studies focus on the classification of lines on the Piping and instrument diagram (P&ID) (Kim et al., 2021), where the line element category can be analyzed by the attached specific symbols. However, these

methods do not adopt to construction drawings because component shapes are represented by proportional contours rather than abstract symbols. And the image features-based Convolutional Neural Networks (CNNs) methods cannot be directly used to predict pixel categories due to lack of texture. Other studies map the line elements in mechanical drawings to homogeneous graph structures where the line elements are represented by nodes, the edges represent adjacent relationships between lines, and Graph neural networks (GNNs) models are used to predict line categories. The GNN-based methods do not suffer from feature sparse issues. However, the line density in construction drawings exceeds that in mechanical drawings. The topological relationship of lines in a drawing cannot be expressed solely in terms of adjacency. It should consider the heterogeneous relationships among lines. Moreover, except for the line geometric attributes, the distribution of nearby texts is also the basis for determining the line category. Therefore, there is still room for improvement in the line classification of construction paper drawings.

Considering these limitations, this study proposes a paper drawings line elements vectorization and layering method to automate the ageing building paper drawings automatic layering process of BIM reconstruction. For improving the paper drawing layering efficiency, this study 1) automates the drawing line elements vectorization process through image processing; and 2) accurately classifies line elements by a heterogeneous GNN-based nodal classification method. This study will provide the ageing building BIM reconstruction field with a new information extraction method, which will help enhance the ageing building O&M efficiency.

2. LITERATURE REVIEW

This section reviews information extraction methods for BIM reconstruction based on paper drawings to demonstrate the motivation for this study. Moreover, this section introduces related line classification and layering methods, these methods are not adaptable for ageing building paper drawing layering due to the difference in research object or scope of application of these methods. Based on the review, the research gaps are identified.

2.1 BIM reconstruction information extraction method based on paper drawings

Building drawings can be divided into CAD vector drawings, raster drawings, and paper drawings. CAD drawings store drawing elements in vectorized format, which can provide accurate geometric information of building components and the standardized layer name can also provide semantic information such as component categories (Yin et al., 2020). However, due to technological development, many ageing buildings can only obtain hand-drawn or paper drawings from existing completion documents. These paper drawings are generally converted to raster format through electronic scanning or photography for data storage and transmission. The information extracted from drawings for BIM reconstruction mainly includes geometric and semantic information. Geometric information typically refers to the components' locations, shapes, and the drawing coordinate system (Pan et al., 2023); semantic information typically refers to components' categories (Xu et al., 2024) or other drawing elements' categories (such as symbols and tags (Moon et al., 2021; Pan et al., 2023)).

Existing research uses various deep learning-based methods to extract geometric information and semantic information from drawings, because these methods are more robust to different drawing scales or rotations. For deep learning-based methods, drawing layering pre-processing is the necessary step to improve information extraction accuracy. Regarding geometric information extraction, the overlapping and obscuring of different categories of lines will affect the geometric information extraction performance. Since the instance masks generated by instance segmentation methods usually has a rough outline which cannot be directly used for BIM reconstruction, some studies combine instance segmentation results with image processing methods, such as connected domain algorithm (Pan et al., 2023), to simultaneously extract the accurate geometric contour and component category. Drawing layering is necessary for these methods because the overlapped line elements can segment complete contour and lead to incomplete contour detection results. Regarding semantic information extraction, many components on drawings have a similar appearance (for example, many components are represented by similar geometric primitives such as rectangles or polygons), which will mislead the deep learning model (Zhao et al., 2021). Therefore, drawing layering can improve the semantic information extraction performance. Moreover, drawing layering can enrich the intuitive representation of semantic data in paper drawings, such as component categories, etc., and enhance the computer's perception and understanding of drawings (Lin et al., 2023), thus supporting more complex BIM reconstruction tasks.

In summary, paper drawings are important information sources for ageing building BIM reconstruction. Automatic information extraction from drawings through deep learning methods can improve modeling efficiency. Although drawing layering is the key to improving geometric and semantic information extraction accuracy, few studies have focused on this issue. Motivated by the current research status, this study aims to propose an automatic paper drawing layering method, including drawing vectorization and line classification.

2.2 Automatic line classification and layering method

The objective of paper drawing layering is to identify drawing elements that represent different components and export them to different layers. In construction engineering drawings, line elements represent most building components and annotation information, extensive drawing layering research focuses on drawing line classification. Line classification methods can be divided into conventional methods and deep learning

methods. Similar to drawing information extraction, conventional methods such as template matching (Kang et al., 2019) or heuristic rule-based detection (Moreno-García et al., 2017) have problems of poor adaptability and poor generalization to drawing rotation and scale changes respectively. As for those machine learning-based methods, the feature types used for classification mainly include visual features and graph features.

The visual feature-based methods employ CNNs to extract pixel-scale drawing visual features. Based on these features, object detection or instance segmentation methods classify individual pixels that makeup drawing line elements, and pixel areas of the same classification are marked as component contours. However, line classification methods based on visual features perform poorly due to the pixel sparsity and lack of texture of line elements (Fan et al., 2021; Xie et al., 2022). Some relevant studies combine drawing symbol detection and line extraction to predict the line types through heuristic matching rules (Kim et al., 2021; Moon et al., 2021). This kind of method, however, only adapts to the drawings representing information by lots of abstract symbols and lines, such as P&ID. In construction engineering drawings, in addition to the axis and annotation lines, other line elements generally represent the actual location and contour of the component. Therefore, when classifying component line elements such as wall contours, the symbol detection-based methods are not adaptable.

The GNN constructs graph structure according to the message-passing mechanism between the graph nodes and generate the embedding vectors that represent the node attribute and connection relationship with neighbors for the downstream classification tasks. Compared with visual features, the GNN-based methods are not affected by similar appearance. Existing studies encode the position, size, and other attributes of line elements into nodal features, and constructs edges between nodes according to the adjacent relationship to construct a homogeneous graph structure (Xie et al., 2022; W. Zhang et al., 2023), then the GNN model is used to conduct line classification on mechanical drawings, which achieved good results. However, the density of lines in construction engineering drawings is greater, and adjacent contour lines and annotation lines often have similar attributes, making it hard to predict the categories solely by relying on the geometric attributes of lines. Moreover, in engineering practice, human engineers often refer to the lines' nearby annotation texts to understand drawings. Therefore, it is necessary to consider the impact of the semantic information of the annotation text when conducting line classification tasks. Regarding the relationship between line elements, in addition to the adjacency relationship, construction engineering drawings also include many parallel lines (such as wall contours represented by parallel lines). However, current homogeneous graph-based line classification research fails to consider this relationship. Heterogeneous graph is a type of graph structure that encompasses multiple types of nodes and edges, which can represent more complex data systems with different relationships. Therefore, this study aims to propose a heterogeneous graph structure to represent the complex line relationships of construction engineering drawings.

In summary, the GNN-based method can encounter the challenge of sparse visual features compared to the CNN-based methods, but existing GNN-based research failed to accurately classify dense and complex line elements in construction engineering drawings. As a result, this study designs a type of heterogeneous graph structure to accurately represent the complex engineering drawings. The line elements in the drawings are treated as graph nodes, and the line location, size attributes, as well as information of nearby annotation texts are encoded as nodal features. And the graph edges are used to represent the adjacency and parallel relationships between lines.

2.3 Research gaps

It is observed that most paper drawing layering methods 1) only consider the geometric attributes of line elements and fail to simultaneously encode the annotation text information as node features, and 2) are unable to encompass complex topological relationships between line elements. As a result, the line elements in ageing building paper drawings cannot be accurately classified, affecting the paper layering pre-processing and information extraction accuracy for BIM reconstruction.

3. Methods

The proposed paper drawings' automatic layering method is shown in Figure 1. Given a paper drawing as input, the drawing vectorization module will automate the drawing vectorization process by extracting line elements. Specifically, the line thinning processing will be conducted to reduce subsequent line duplication detection due to different line widths, then the LSD line detection method (Grompone von Gioi et al., 2010) will extract the vectorized format line elements by detecting line endpoints. In the graph construction module, a nodal feature list is proposed to consider the line geometric attributes and their relationship with adjacent annotation texts, and two types of edges, including adjacent and parallel edges are constructed according to the topological relationship with adjacent line elements. This graph construction strategy can accurately represent the characteristics of line elements by considering line attributes and surrounding texts simultaneously, and represent more complex relationships among line elements in construction engineering drawings. In the GraphSAGE (Graph SAmple and aggreGatE) (Hamilton et al., 2017)-based line classification module, a GraphSAGE model architecture suitable for the heterogeneous graph structures proposed in this study is developed. The GraphSAGE model is an inductive representation learning method that is not limited to fixed graph structure, which enables the line classification model learned from training set drawings to generalize to different drawings. The following sections detail each module.

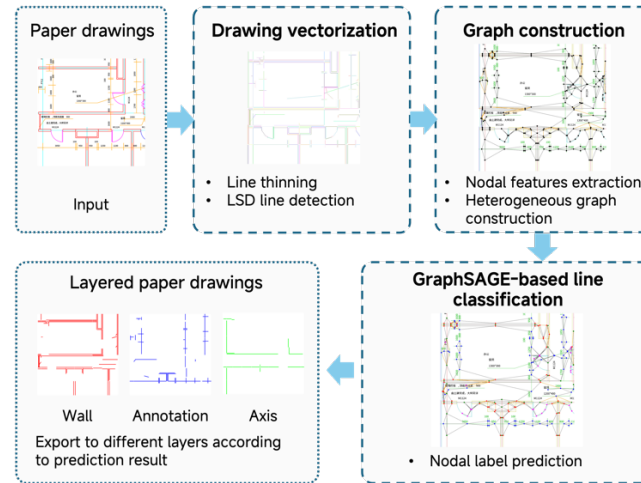


Figure 1. The roadmap of the proposed method

3.1 Drawing vectorization

The purpose of drawing vectorization is to convert line elements from a combination of discrete pixels to vectorized elements represented by endpoints, which are used to construct the graph structure. Given that different categories of drawing elements typically have different line thicknesses, for example, contour lines of walls are usually wider than annotation lines, which will cause different distance between line pairs in the next line extraction step. Therefore, the line thinning step is necessary. This study employs line thinning algorithm proposed by Zhang and Suen (T. Y. Zhang & Suen, 1984) to get the skeletons of line elements, the input drawing and result of line thinning is shown in Figure 2.(a) and (b) separately.

Based on the line element skeletons, the vectorized line elements are obtained through line detection. In this step, LSD line detection method is utilized due to its superior adaptation to high-resolution raster drawings compared with Hough transform method, which is commonly used in other studies (Kim et al., 2021; Moon et al., 2023). The LSD detector generates parallel lines on both sides of original line elements. To ensure that parallel lines belonging to the same original elements are correctly paired, a line grouping algorithm is devised in this study. This algorithm assesses the parallelism of current lines, pairing those of similar length and within a specified threshold distance. The formed line pairs are shown in Figure 2 (c) with each line pair represented by a distinct color. Note that the prior line thinning ensures uniform thickness across all original line elements, facilitating the establishment of distance thresholds for pairing. Finally, to get the line vectors represent original line elements, two lines in a line pair are merged by averaging the endpoint coordinates. The result of drawing vectorization is shown in Figure 2 (d). The drawing vectorization module automates the manual drafting involved in paper-based BIM reconstruction and generates vectorized line elements for graph construction in the subsequent phase.

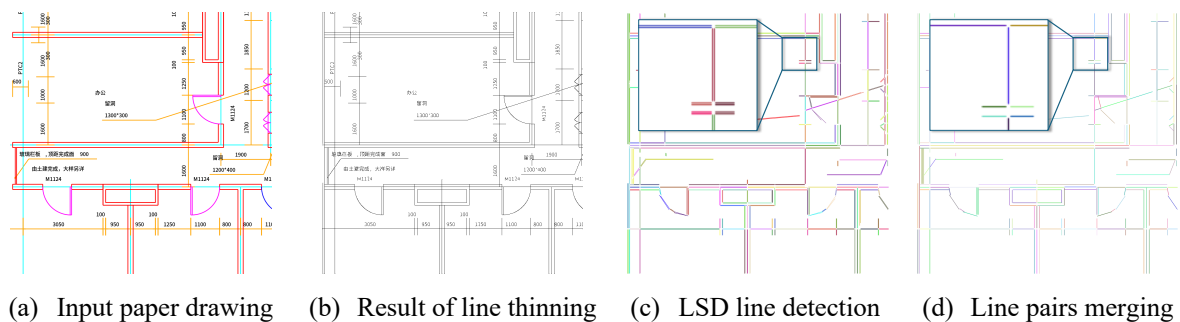


Figure 2. Drawing vectorization processing

3.2 Graph construction

The heterogeneous graph structure plays an important role in embedding the line attributes and contextual topological relationships. In the proposed graph structure, nodes correspond to line elements extracted in the vectorization processing, while edges represent the adjacent and parallel relationship within line elements.

This study introduces a nodal feature list to encode the line element's geometric attributes and the relationship with nearby annotation texts. As illustrated in Table 1, the features include the position, dimension,

angle and relationship with texts of each line elements. Specifically, the coordinates and lengths of the line endpoints are normalized to the dimensions of the drawings (in this study, the dimension of raster format paper drawings is 1000 * 1000) to ensure that the position and dimension features remain invariant across varying drawing sizes. The angle feature is the angle with horizontal direction. The angle feature denotes the orientation with respect to the horizontal axis. Regarding the textual relationships, the annotation text type distinguishes between numeric and non-numeric content. Text type encoding is crucial as numeric text typically signifies dimensions, whereas non-numeric text commonly denotes component categories, axis labels, room types, and so on. The text direction feature indicates whether the text aligns with the line horizontally. The purpose of encoding text direction is that text describing a line typically has the same orientation as the line. Furthermore, both vertical and extended text directions are encoded to comprehensively capture the topological relationships with neighboring texts. In each direction, the text features include five elements, representing the presence of text, the proportion of numeric to non-numeric text, and the ratio of parallel numeric and non-numeric text lines.

Table 1. Nodal features for encoding line element attributes

Feature type	Description	Feature dimension
Position	2D coordinate of line endpoints	4
Dimension	Length	1
Angle	Angle with horizontal direction in drawing	1
Relationship with texts	Proportion of different types and directions of annotation text within the vertical direction neighbor of line elements	5
	Proportion of different types and directions of annotation text within the extended direction neighbor of line elements	5

The proposed heterogeneous graph incorporates two different kind of edges that represent adjacent and parallel relationship between line elements separately. The adjacent edges are generated from the connected or intersecting line elements, capturing the combination relationship of contour lines or intersection between component contours and annotation lines. The parallel edges are generated from the nearby parallel lines within a specified threshold distance. Parallelism holds significant relevance in construction engineering drawings, evident in features like walls delineated by parallel lines and dimension lines indicating component sizes and scales. The introduction of parallel edges aims to depict the parallel relationships among line elements, enabling a more detailed representation of complex drawings. In summary, the graph structure can be defined as:

$$G(N, E_a, E_p), N \in \mathbb{R}^{n_N \times 16}, E_a \in \mathbb{Z}^{n_{E_a} \times 2}, E_p \in \mathbb{Z}^{n_{E_p} \times 2} \quad (1)$$

where N represents the nodal features of the constructed graph, with n_N denoting the number of nodes which is also the number of line elements in drawing. E_a denotes the edge indices of undirected adjacent edges, with n_{E_a} indicating the quantity of adjacent edges. E_p represents the edge indices of undirected parallel edges, with n_{E_p} representing the number of parallel edges. The line element category, as well as the nodal label is defined as $Y \in \mathbb{Z}^{n_N}$. In this study, the line element category is separated as wall lines, annotation lines, axis lines and other lines. The method for constructing ground truth labels is partially adapted from (W. Zhang et al., 2023). As shown in Figure 3, in original vectorized format drawings, each category is specified with a unique color. The extracted line elements are redrawn to the original drawings (as depicted by black pixels in Figure 3), then the ground truth label is determined by counting the number of pixels of corresponding color in redrawn range.

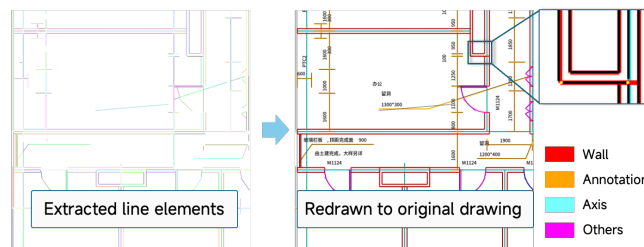


Figure 3. Extracted line elements ground truth label obtaining process

3.3 GraphSAGE-based line classification

Based on the constructed graph structure, the line classification task transformed into a nodal classification task. This study introduces a GraphSAGE-based model to tackle nodal classification. GraphSAGE is a superior nodal classification model, which allows the training of an embedding function to aggregate nodal features from local neighborhood. The embedding function can be generalized to unseen graphs (such as drawings in the

validation set) without requiring additional training, thereby enhancing the applicability and adaptability of the line classification method. Figure 4 (a) shows the structure of the proposed GraphSAGE-based model, the node embeddings of the input graph are generated through two heterogeneous GraphSAGE layers. These node embeddings are then transformed into one-dimension vector through linear layers to represent classification result. The structure of each hidden layer is depicted in Figure 4 (b), showcasing a hetero-GraphSAGE layer structure is proposed to aggregate information from adjacent neighbors and parallel neighbors of central node. For each type of edge, the hetero-GraphSAGE layer samples neighbor nodes attribute features to generate messages, which are then aggregated with the central node's attribute features to produce the node representation. The node representations stemming from the two types of edge connections are combined via a normalized sum and passed through the ReLU activation function to derive the final node embedding. This study specifies the number of hidden layers as 2, indicating that the model predicts the line category based on aggregated information from its 1- and 2-hop neighbors.

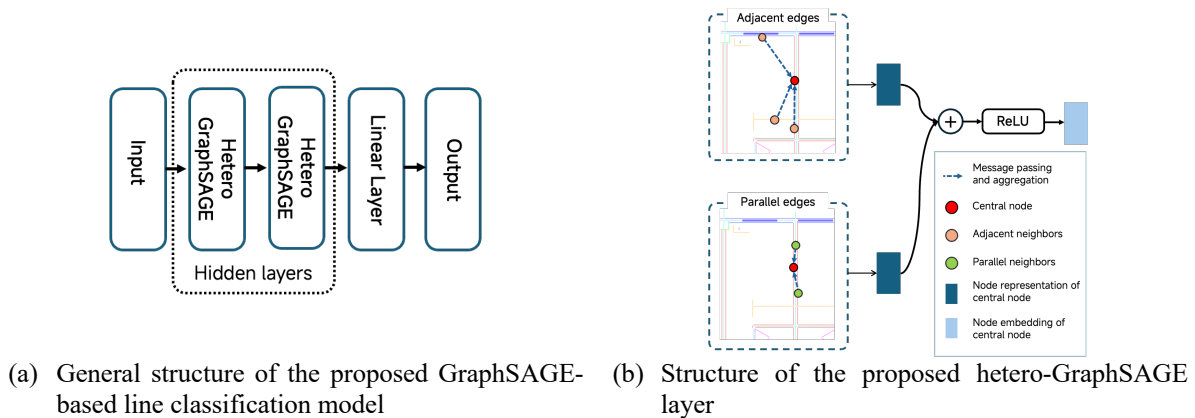


Figure 4. Structure of the proposed GraphSAGE-based line classification model

4. Experiment

To verify the feasibility of the proposed method, actual engineering drawings from the FloorPlanCAD (Fan et al., 2021) dataset were employed for the experimentation. The subsequent sections detail the dataset processing, experimental setup and results.

4.1 Experiment setup

This section elaborates on dataset processing method, model parameters, experimental environment and evaluation metrics.

(1) Dataset processing method

The FloorPlanCAD dataset stores actual engineering floor drawings in vectorized format, encompassing over 30 component categories within the drawings. Initially, certain components were excluded based on the study's focus. The object of this study is paper drawings of building floor plans, which should include basic building structural components and annotation elements. Thus, this study separated the line categories as wall contour lines (Wall), annotation lines (Annotation), axis lines (Axis) and other lines (Others). The purpose of choosing these categories is that walls are structural backbone of buildings, with other elements like doors and windows typically attached to walls. Axis lines offer component coordinates concerning the building's coordinate system, while annotation lines furnish precise dimensional parameters for components. Furthermore, this study assumed that furniture and décor with specific range and countable instances could be removed by adding masks during preprocessing stage (such as instance segmentation-based pre-processing). Therefore, this study removed the furniture and decoration category components, as well as anomalous samples such as blank drawings during dataset processing.

The filtered vectorized drawings were exported into raster drawings to simulate scanned versions of ageing building's paper drawings. Subsequently, employing the proposed vectorization and graph construction method described in Section 3, the line elements in each drawing are extracted and used to construct the graph structure. Finally, this study constructs a dataset including 2046 graphs. The statistics of the node number of each category are shown in Table 2. The training and test sets are divided from these graphs at a ratio of 4:1.

Table 2. The quantity of each category of nodes

Category	Wall	Axis	Annotation	Others
Number of nodes	132753	56696	46492	70942
Proportion of node (%)	43.26	18.47	15.15	23.12

(2) Model parameters

Using the GraphSAGE-based model structure described in Section 3.3, this study trains a line classification model (Hetero GraphSAGE) to predict categories of line elements on drawing. In addition to the Hetero GraphSAGE model, a baseline model (Baseline) adapting to homogeneous graph is introduced in the experiment. The dataset for Baseline model only encompasses adjacent edges, aligning with previous studies (W. Zhang et al., 2023). For the nodal features of the baseline model, relevant research primarily focuses on the geometric attributes of line elements when constructing nodal features (W. Zhang et al., 2023). Since there is no relevant research using exactly the same vectorization method as this study, the nodal features proposed by these studies cannot be directly applied to baseline model. Consequently, the nodal features of the baseline model are confined to Position, Dimension, and Angle attributes as specified in Table 1. To maintain a fair comparison, the two models are designed to have similar structure and depth, which detailed by:

Hetero GraphSAGE: 2 Hetero GraphSAGE convolutional layers for hidden layers+1 linear layer+1 output layer, each hidden layer has 256 nodes (each kind of edge has 128 nodes). $2 \times \text{ReLU} + 1 \times \text{Logsoftmax}$ as nonlinear activation functions.

Baseline: 2 GraphSAGE convolutional layers for hidden layers+1 linear layer+1 output layer, each hidden layer has 256 nodes. $2 \times \text{ReLU} + 1 \times \text{Logsoftmax}$ as nonlinear activation functions.

(3) Experimental environment

The experimental environment is Ubuntu 22.04, an i7-13700kf processor, an RTX 4090 graphics card, 32 GB of ram, Python 3.8, PyTorch 2.0.1 and PyTorch Geometric 2.5.2. In this study, a strategy of progressively decreasing the learning rate over multiple training iterations is employed. The optimizer utilized is Adaptive Moment Estimation (Adam). Across three training phases, the model is trained for [200, 3000, 3000] epochs, with the learning rate for each phase set at [0.01, 0.001, 0.0005]. Additionally, weight decay is incorporated with values of [5e-4, 5e-5, 5e-6] for the respective training rounds.

(4) Evaluation metrics

The recall (R) and precision (P) is used to evaluate the model performance, which are defined as:

$$P = \frac{TP}{TP + FP} \quad (2)$$

$$R = \frac{TP}{TP + FN} \quad (3)$$

Moreover, the F1 score (F_1) is used to measure the overall performance of line classification model, which is defined as:

$$F_1 = \frac{2 \times P \times R}{P + R} \quad (4)$$

4.2 Experimental results

This section presents the line classification results of the proposed Hetero GraphSAGE model and comparison with the Baseline model.

(1) Line classification results

The quantitative results of the Hetero GraphSAGE model are shown in Table 3. The results show that the F_1 indicators of wall, annotation and others exceed 0.75, with good performance in the line classification task. Notably, the r indicator of wall stands at 0.84, indicating the model's proficiency in predicting most wall contour lines from paper drawings, providing an accurate reference for other components in BIM reconstruction. In terms of p indicators, the model exhibits the highest precision in predicting the annotation category, facilitating the precise restoration of geometric data. However, the model did not perform as well in predicting axis lines. This can be attributed to the fact that axis tend to coexist with walls in the drawings, as well as the lack of distinctive features of the axis in the absence of axis tags. Additionally, the imbalance in the dataset affects the prediction results for the axis category due to the relatively small number of axes in the dataset. The model demonstrates relatively strong predictive performance in the others category, aiding in filtering out extraneous components (such as doors and windows) prior to extracting wall information. The classification outcomes for others lines can be effectively integrated with object detection or instance segmentation methodologies to extract architectural component information efficiently.

Table 3. Quantitative results on the proposed dataset of Hetero GraphSAGE model

Indicators	Wall	Axis	Annotation	Others	F_1
r	0.842	0.625	0.743	0.736	-
p	0.759	0.731	0.785	0.774	-
F_1	0.799	0.674	0.763	0.755	0.748

The qualitative results of the Hetero GraphSAGE model are illustrated in Figure 5. The model effectively classifies the wall lines. While some of the short edge contours of the wall may be missing, the accurately categorized lines are a good representation of the contours and extent of the wall. The result of annotation line

classification in the left-bottom of Figure 5 (c) shows the model performs well even in a relatively line dense drawing. However, errors in classification become more prevalent when the distance between wall lines and annotation lines is minimal, thereby impacting the continuity of the local wall contour. Errors in axis line classification commonly arise when axis lines intersect walls. Notably, when axis lines are encircled by numerical texts, they are prone to misclassification as annotation lines, as exemplified in Figure 5 (c).

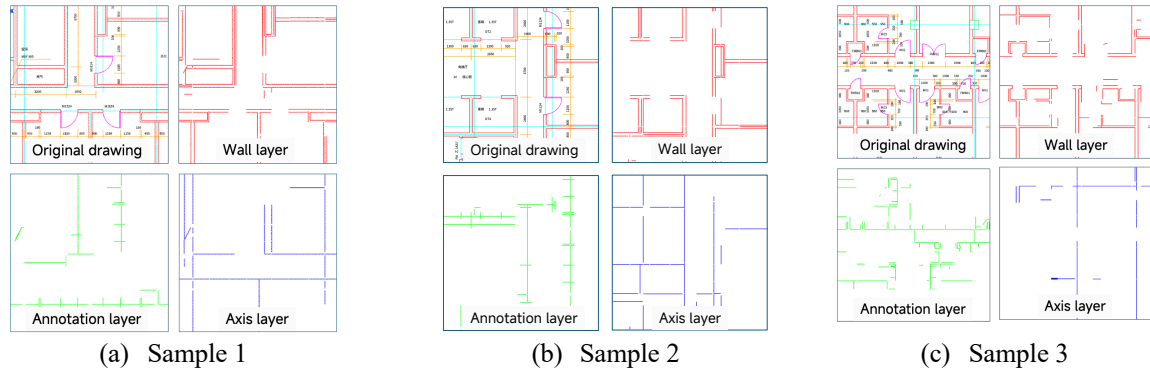


Figure 5. Qualitative results of the Hetero GraphSAGE model

(2) Baseline comparison

In order to demonstrate the significance of the introduced nodal features and heterogeneous graph structure in enhancing line classification accuracy, an experiment involving the baseline model is conducted. The F_1 indicators are used to assess model performance, and the comparative results between the baseline model and the proposed Hetero GraphSAGE model are detailed in Table 4. The results indicate that Hetero GraphSAGE model outperforms Baseline model across all line classification categories. The overall performance in categorization tasks shows an enhancement of 0.102. The F_1 scores for wall, annotation and others line classification all exceeded 0.75. Particularly in the annotation category, the Hetero GraphSAGE model F_1 shows a 0.193 improvement over the Baseline model, which indicates the significance of integrating text-related nodal features and a heterogeneous graph structure in achieving superior classification accuracy.

Table 4. Comparison results for line classification performance of two models.

Categories	Wall	Axis	Annotation	Others	F_1
Baseline	0.736	0.597	0.570	0.682	0.646
Hetero GraphSAGE	0.799	0.674	0.763	0.755	0.748

5. Discussion

This study proposes a high accuracy automatic line classification method for the paper drawing layering process in BIM reconstruction. As mentioned in the previous section, most line elements in drawings can be accurately classified. This section will analyze the effect of the proposed nodal feature list and heterogeneous graph structure on the line classification performance, as well as the shortcomings of current research.

(1) Effect of nodal feature list on line classification task

The proposed nodal feature list plays an important role in encoding line element geometric attributes and topological relationship with nearby texts. To illustrate the effectiveness of these text-relevant features, a comparative experiment was conducted to assess the performance of the Hetero GraphSAGE model using different nodal features. The two models used for comparison have the same structure, the "WithText" model, trained on the proposed nodal features including text-relevant attributes, and the "WithoutText" model, trained on a dataset devoid of text-relevant nodal features as detailed in Table 1 (encompassing only Position, Dimension, and Angle features). The results are shown in Table 5, the integration of text-relevant features leads to a slight improvement in the classification performance of wall, axis, and other categories of lines. Notably, the performance in annotating line classification sees a substantial enhancement of nearly 0.14 with the inclusion of text-relevant features. Furthermore, the overall performance witnesses an improvement of approximately 0.15 upon the introduction of these text-relevant features, which proves the effectiveness of the proposed nodal feature list.

Table 5. Comparison results for F_1 of the models trained with different nodal features

Categories	Wall	Axis	Annotation	Others	F_1
WithText	0.799	0.674	0.763	0.755	0.748
WithoutText	0.781	0.652	0.624	0.727	0.696

(2) Effect of the heterogeneous graph structure on line classification task

The incorporation of a heterogeneous graph structure serves as an additional strategy to enhance line classification performance. In order to evaluate the impact of this heterogeneous graph structure on line classification accuracy, both the Hetero GraphSAGE model (Hetero) and the Baseline model (Homo) were trained on identical datasets, the results are shown in Table 6. The results highlight that the influence of the graph structure on annotation line classification is relatively minor compared to other categories. The enhancement in predicting annotation lines primarily stems from the inclusion of text-relevant features. However, for wall, axis, and other line classifications, there is an improvement ranging from approximately 0.05 to 0.07. Overall, the heterogeneous model demonstrates a 0.05 enhancement in performance, shows the effectiveness of the proposed heterogeneous graph structure.

Table 6. Comparison results for F_1 of models on different graph structure

Categories	Wall	Axis	Annotation	Others	F1
Hetero	0.799	0.674	0.763	0.755	0.748
Homo	0.750	0.600	0.748	0.693	0.698

(3) Limitations and future work

Even though the experiments have demonstrated that the proposed method can improve the line classification accuracy thereby supporting the paper drawing layering pre-processing in ageing building BIM reconstruction, the current study still has some limitations.

Firstly, although the proposed method can correctly classify most of the wall lines, the margin represented by short lines are often missing, which will affect the subsequent contour detection. The primary reason is that current drawing vectorization method is designed to filter out line segments that are too short to reduce the noise caused by text contour, however, current strategy also causes some information loss. The line missing problem also causes losing of adjacent connection relationship, potentially affect the classification accuracy. Future study could focus on refining the drawing vectorization method to address text interference more effectively. Secondly, the study employs a uniform nodal feature list for all line categories, overlooking distinctive features like axis tags and annotation line arrows that provide clear hints for line categorization. Future endeavors could explore integrating object detection methods to extract these features beforehand, thus enhancing the classification performance of relevant lines. Finally, current text-relevant nodal features only distinguish text types as numeric and nonnumeric. The semantic type of the text, such as dimension text, category text requires be further analyzed and encoded to the nodal feature list. Moreover, the improvement of the proposed line classification method on the downstream information extraction accuracy, such as building components object detection task, will be quantitatively evaluated in future work.

6. Conclusions

BIM reconstruction plays an important role in enabling smart O&M practices for aging buildings. However, the low efficiency of paper drawing layering pre-processing affects the practical application of BIM reconstruction. To address this challenge, this study proposes a heterogeneous GNN-based paper drawing line classification method to improve the drawing layering efficiency. Compared to existing studies, the proposed method has the following innovations. Firstly, the nodal feature list incorporates not only the geometric attributes of the line itself but also its relationship with nearby text elements, thereby improving the ability of node features to represent the properties of different categories of lines. Secondly, the proposed heterogeneous graph structure captures the complex topological relationships among lines by representing the adjacent and parallel relationship using different types of edge. Finally, the Hetero GraphSAGE model structure is proposed for conducting the line classification task on heterogeneous graph. Experimental results reveal that the model achieves a F_1 score of 0.748 in the line classification task and yields superior performance compared to the baseline model. In conclusion, the proposed method improves the efficiency and accuracy of paper drawing layering pre-processing. It will also provide a reliable information extraction tool to aid in the BIM reconstruction of ageing buildings.

ACKNOWLEDGMENTS

This work was supported by the Innovation and Technology Fund Seed Research Project (ITS/070/22), the Innovation and Technology Fund Platform Research Projects (ITP/002/22LP), and the Guangdong Basic and Applied Basic Research Foundation (2022B1515130006).

REFERENCES

- de Wilde, P. (2019). Ten questions concerning building performance analysis. *Building and Environment*, 153, 110–117. <https://doi.org/10.1016/j.buildenv.2019.02.019>

- Fan, Z., Zhu, L., Li, H., Chen, X., Zhu, S., & Tan, P. (2021). FloorPlanCAD: A Large-Scale CAD Drawing Dataset for Panoptic Symbol Spotting. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 10108–10117. <https://doi.org/10.1109/ICCV48922.2021.00997>
- Feng, Y., Wang, J., Fan, H., & Hu, Y. (2021). A BIM-Based Coordination Support System for Emergency Response. *IEEE Access*, 9, 68814–68825. IEEE Access. <https://doi.org/10.1109/ACCESS.2021.3077237>
- Grompone von Gioi, R., Jakubowicz, J., Morel, J.-M., & Randall, G. (2010). LSD: A Fast Line Segment Detector with a False Detection Control. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(4), 722–732. IEEE Transactions on Pattern Analysis and Machine Intelligence. <https://doi.org/10.1109/TPAMI.2008.300>
- Hamilton, W. L., Ying, R., & Leskovec, J. (2017). Inductive representation learning on large graphs. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 1025–1035.
- Hu, Z.-Z., Leng, S., Lin, J.-R., Li, S.-W., & Xiao, Y.-Q. (2022). Knowledge Extraction and Discovery Based on BIM: A Critical Review and Future Directions. *Archives of Computational Methods in Engineering*, 29(1), 335–356. <https://doi.org/10.1007/s11831-021-09576-9>
- Hu, Z.-Z., Zhang, J.-P., Yu, F.-Q., Tian, P.-L., & Xiang, X.-S. (2016). Construction and facility management of large MEP projects using a multi-Scale building information model. *Advances in Engineering Software*, 100, 215–230. <https://doi.org/10.1016/j.advengsoft.2016.07.006>
- Kamel, E., & Memari, A. M. (2019). Review of BIM's application in energy simulation: Tools, issues, and solutions. *Automation in Construction*, 97, 164–180. <https://doi.org/10.1016/j.autcon.2018.11.008>
- Kang, S.-O., Lee, E.-B., & Baek, H.-K. (2019). A Digitization and Conversion Tool for Imaged Drawings to Intelligent Piping and Instrumentation Diagrams (P&ID). *Energies*, 12(13), Article 13. <https://doi.org/10.3390/en12132593>
- Kim, H., Lee, W., Kim, M., Moon, Y., Lee, T., Cho, M., & Mun, D. (2021). Deep-learning-based recognition of symbols and texts at an industrially applicable level from images of high-density piping and instrumentation diagrams. *Expert Systems with Applications*, 183, 115337. <https://doi.org/10.1016/j.eswa.2021.115337>
- Lin J., Zhou Y., Zheng Z., & Lu X. (2023). Research and application of intelligent design review. *Engineering Mechanics*, 40(7), 25–38. <https://doi.org/10.6052/j.issn.1000-4750.2021.11.0908>
- Moon, Y., Han, S.-T., Lee, J., & Mun, D. (2023). Extraction of line objects from piping and instrumentation diagrams using an improved continuous line detection algorithm. *Journal of Mechanical Science and Technology*, 37(4), 1959–1972. <https://doi.org/10.1007/s12206-023-0333-9>
- Moon, Y., Lee, J., Mun, D., & Lim, S. (2021). Deep Learning-Based Method to Recognize Line Objects and Flow Arrows from Image-Format Piping and Instrumentation Diagrams for Digitization. *Applied Sciences*, 11(21), Article 21. <https://doi.org/10.3390/app112110054>
- Moreno-García, C. F., Elyan, E., & Jayne, C. (2017). Heuristics-Based Detection to Improve Text/Graphics Segmentation in Complex Engineering Drawings. In G. Boracchi, L. Iliadis, C. Jayne, & A. Likas (Eds.), *Engineering Applications of Neural Networks* (pp. 87–98). Springer International Publishing. https://doi.org/10.1007/978-3-319-65172-9_8
- Pan, Z., Yu, Y., Xiao, F., & Zhang, J. (2023). Recovering building information model from 2D drawings for mechanical, electrical and plumbing systems of ageing buildings. *Automation in Construction*, 152, 104914. <https://doi.org/10.1016/j.autcon.2023.104914>
- Xie, L., Lu, Y., Furuhashi, T., Yamakawa, S., Zhang, W., Regmi, A., Kara, L., & Shimada, K. (2022). Graph neural network-enabled manufacturing method classification from engineering drawings. *Computers in Industry*, 142, 103697. <https://doi.org/10.1016/j.compind.2022.103697>
- Xu, Z., Jha, N., Mehadi, S., & Mandal, M. (2024). Multiscale object detection on complex architectural floor plans. *Automation in Construction*, 165, 105486. <https://doi.org/10.1016/j.autcon.2024.105486>
- Yin, M., Tang, L., Zhou, T., Wen, Y., Xu, R., & Deng, W. (2020). Automatic layer classification method-based elevation recognition in architectural drawings for reconstruction of 3D BIM models. *Automation in Construction*, 113, 103082. <https://doi.org/10.1016/j.autcon.2020.103082>
- Zhang, T. Y., & Suen, C. Y. (1984). A fast parallel algorithm for thinning digital patterns. *Communications of the ACM*, 27(3), 236–239. <https://doi.org/10.1145/357994.358023>
- Zhang, W., Joseph, J., Yin, Y., Xie, L., Furuhashi, T., Yamakawa, S., Shimada, K., & Kara, L. B. (2023). Component segmentation of engineering drawings using Graph Convolutional Networks. *Computers in Industry*, 147, 103885. <https://doi.org/10.1016/j.compind.2023.103885>
- Zhao, Y., Deng, X., & Lai, H. (2021). Reconstructing BIM from 2D structural drawings for existing buildings. *Automation in Construction*, 128, 103750. <https://doi.org/10.1016/j.autcon.2021.103750>

Investigating Solutions to the Shortage of Construction Site Managers in Japan: A Research Study on the Introduction of Young BIM Engineers from China

Rui Sucheng¹, Koji Makanae², Liu Jun³, Makoto Fujii⁴ and Yuma Morisaki⁵

- 1) Ph.D. Candidate, Graduate School of Natural Science and Technology, Kanazawa University, Kanazawa, Japan.
Email: ruisucheng@stu.kanazawa-u.ac.jp
- 2) Ph.D., Prof., School of Project Design, Miyagi University, Japan. Email: makanae@myu.ac.jp
- 3) Ph.D., Assoc. Prof., School of Environment and Architecture, University of Shanghai for Science and Technology, Shanghai, China. Email: liujun@usst.edu.cn
- 4) Ph.D., Assoc. Prof., Institute of Transdisciplinary Sciences for Innovation, Kanazawa University, Kanazawa, Japan.
Email: fujii@se.kanazawa-u.ac.jp
- 5) Ph.D., Assistant Professor., Institute of Transdisciplinary Sciences for Innovation, Kanazawa University, Kanazawa, Japan.
Email: morisaki@staff.kanazawa-u.ac.jp

Abstract: Japanese construction industry is facing a critical labor shortage due to its aging population. To address this issue, utilizing Building Information Modeling (BIM)/Construction Information Modeling (CIM) to enhance productivity and employing foreign workers are considered effective solutions. However, research on the integration of foreign technical talents in Japanese construction sector remains limited. This paper aims to explore the potential of introducing young Chinese technicians trained in BIM to alleviate the shortage of site management personnel in Japan. First, a comparative study of BIM education in universities in China and Japan is conducted to assess the current state of BIM training in both countries. Following this, the willingness of 187 young Chinese technicians to take on management roles in Japan is investigated and analyzed, along with the satisfaction levels of 10 Chinese individuals currently managing sites in Japan. Finally, by synthesizing the survey results, the paper identifies the challenges in practical implementation and confirms the feasibility of employing BIM-educated young Chinese technicians to address Japanese labor shortage in site management.

Keywords: BIM, civil engineers, construction management, education, survey

1. INTRODUCTION

Japan is currently facing a labor shortage, which is particularly acute in the construction industry. According to a survey by the Japan Federation of Construction Contractors, while construction investment has gradually increased since 2010, the number of construction workers has continued to decline from its peak in 1997 of 6.85 million, falling to 4.83 million in 2023, which is only 70.5% of the peak figure. In the same year, approximately 36% of the workforce was aged 55 and above, and about 12% were under the age of 29, indicating a significant aging trend compared to other industries. Data published by the Ministry of Health, Labour and Welfare show that while the effective job openings-to-applicants ratio for general occupations is 1.31, it is 5.57 for construction and civil engineering management-related occupations, highlighting the severe shortage of site management personnel in Japan. Universities are a crucial source of young engineers, and according to the Basic Survey of Schools by the Ministry of Education, Culture, Sports, Science and Technology, the number of students in university programs directly related to construction, such as civil and architectural engineering, has decreased by about 2,900, and technical school students by about 1,700 over the decade from 2011 to 2020. This decline is particularly notable in engineering-related departments.

Global literature has detailed the advantages of Building Information Modeling (BIM) technology, which offers comprehensive lifecycle solutions from design to construction and maintenance, allowing for efficient labor savings and enhanced productivity. This could effectively address the issues facing Japanese construction industry. Consequently, the Ministry of Land, Infrastructure, Transport, and Tourism (MLIT) has intensified its promotion of BIM/Construction Information Modeling (referred to as BIM hereafter) in both architecture and civil engineering. The introduction of foreign talent is also seen as an effective way to quickly alleviate the shortage of construction engineers. Japan has entered the global competition for talent, and the government is designing attractive policies and plans to recruit and retain foreign engineers and professionals with advanced education and skills.

China has the world's largest Architecture, Engineering, and Construction (AEC) industry, reaching approximately \$4.5 trillion in 2023, with about 100,000 civil engineering and architecture university students graduating annually. The advantages of BIM technology are of great importance to Chinese AEC industry, and since 2014, The National Master Specification has been tracking the status of BIM higher education worldwide. According to the latest reports, China provides BIM higher education to students and professionals in the AEC field, with many universities and professional schools offering BIM and smart construction-related courses by 2022, indicating that China will produce many BIM professionals in the future. With the transformation of Chinese construction industry and the expansion of international exchanges, young engineers with architectural expertise are increasingly seeking employment overseas. According to a 2022 survey conducted by MLIT, Chinese

individuals engaged in construction management in Japan account for 45% of foreign employees, significantly surpassing those from other countries, demonstrating the pivotal role Chinese engineers play in Japanese construction sector.

Currently, there is scant research on the introduction of site management personnel in the construction industry, particularly concerning the introduction of foreign talent trained in BIM technology. This study focuses on the introduction of young Chinese talent trained in BIM education to address the shortage of site management personnel in Japan. The objective is to elucidate these challenges through survey analysis. Figure 1 illustrates the flow of the research. The remainder of this paper is structured as follows: Chapter 2 investigates the content and awareness of BIM education in universities in China and Japan. Chapter 3 explores the intentions of young Chinese engineers to engage in management roles in Japan. Chapter 4 surveys the satisfaction of Chinese individuals working in construction management in Japan. Each chapter describes the purpose, method, and results of each survey. Chapter 5 explains and discusses the research findings, and Chapter 6 concludes the paper.

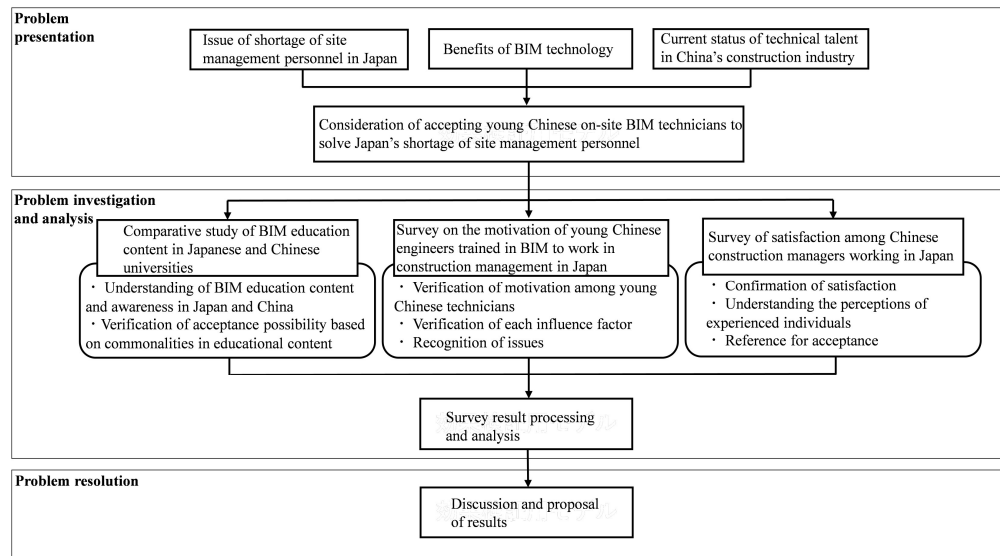


Figure 1. Research flowchart

2. Comparative study of BIM education content in universities of Japan and China

To determine whether Chinese universities BIM education content is applicable to Japanese construction industry, it is necessary to compare the educational content and awareness in both countries. University students are the future bearers and leaders of the construction industry. As an exemplary case of BIM education in Japan, Shibaura Institute of Technology has been offering BIM education since 2008. At this university, undergraduate students receive training in the operation of BIM software, and exercises using BIM for cost estimation are conducted. Similarly, in China, comparable education and exercises are carried out in relevant majors. According to a survey targeting universities in Shanghai, Fujian, and Shandong, like in Japan, BIM education in Chinese universities primarily focuses on the generation of 3D models, with applications in construction management such as site layout, scaffolding design, and material cost estimation also being developed (Figure 2).

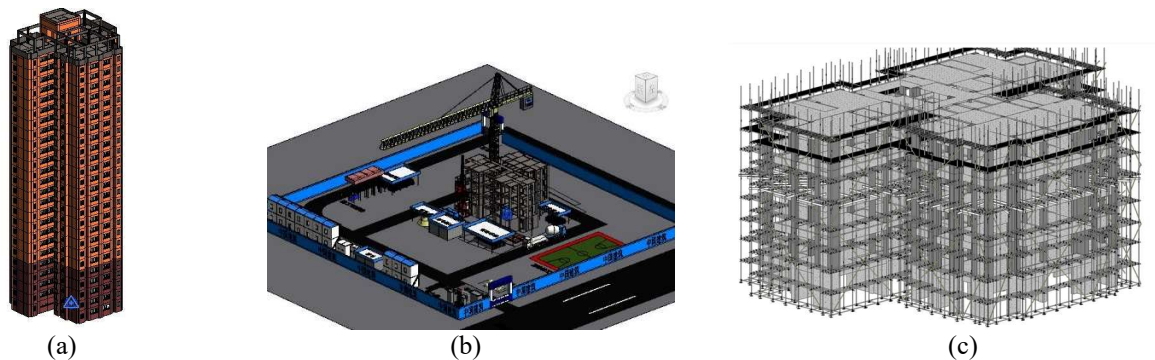


Figure 2. example (a)Example of 3D model (b)Example of site layout drawing (c)Example of temporary construction model

Through a survey of four educators from Japan and China who are involved in BIM education (Table 1), the comparison of BIM education conditions in both countries is shown in Table 2. It was found that universities in both Japan and China provide education related to design and 3D modeling. However, Chinese universities place greater emphasis on applications in construction management, aligning with practical on-site construction work. This indicates that BIM education in Chinese universities is comprehensive, sharing commonalities with Japanese educational content. This provides a basis for the technical feasibility of introducing site management personnel. Both Japan and China recognize the importance of BIM education, and Chinese universities, in particular, receive strong government support, which is expected to further enhance the quality of their education.

Table 1. Respondent characteristics in BIM education survey

country	Position	Years of teaching experience	BIM teaching experience years
Japan	Professor	40years	26years
	Professor	24years	5years
	Professor	21years	6years
	Associate Professor	10years	Unknown
China	Professor	18years	8years
	Professor	16years	10years
	Associate Professor	8years	8years
	Associate Professor	5years	5years

Table 2. Survey results on BIM education

Item	Selected content	Japan	China
BIM-related course content	Design and 3D Modeling	4	4
	Structural Analysis		1
	Construction Management Related	1	4
	Cost Estimation		3
	Sustainable Design	1	1
	Others	1	
The importance of BIM education in university education	Very important	4	1
	Important		3
The university provides sufficient resources and support for BIM education	Think so		2
	Ordinary		2
	Don't think so	2	
	Don't think so at all	2	
How will the role of BIM education in university education change over the next five years	Unchanged from the current situation	1	
	A slight change has no overall impact	3	1
	Changes have an overall impact		3

3. A survey on the willingness of young Chinese engineers who studied BIM to engage in construction management in Japan

3.1 Purpose and Method of the Survey

In the context of rapid global environmental changes, the intention of young Chinese technicians, who have learned BIM-related technologies, to utilize their expertise in construction site management jobs in Japan constitutes a significant topic of discussion. For this purpose, a survey was conducted targeting young students and graduates from universities and vocational schools in Shanghai, Fujian, and Shandong that offer BIM courses. The survey aimed to investigate these young technicians' intentions to engage in site management in Japan and the primary factors influencing such intentions. The objectives of this survey include providing insights into how the Japanese construction industry should approach Chinese technicians, what kind of work environment and benefits they should offer to attract them.

The questionnaire was designed with 12 related questions, four of which concerned basic attributes (gender, educational background, major, and on-site experience). The remaining eight items covered the following areas: motivation to work in site management in Japan, interest in construction site management, current professional skills, understanding of Japanese culture and work environments, expected salary for site management roles in Japan, tolerance for overtime, and concerns about working in construction site management in Japan. The survey was conducted from March to July 2024.

The data collected were statistically analyzed using SPSS and Excel. Correlation and regression analyses were specifically performed to clarify the relationship between respondents' attributes and their motivations. Furthermore, factor analysis was employed to identify key factors that could influence their intentions to work in Japan.

3.2 Survey Results

The basic attributes of this survey are shown in Table 3. There were 187 valid responses, with 140 men (74.87%) and 47 women (25.13%), reflecting the gender ratio typically seen in the construction industry. Regarding educational background, there were 92 graduates from technical colleges (49.20%), 85 university graduates (45.45%), and 10 graduate students (5.35%). In terms of majors, 97 were in architecture (43.32%), 81 in civil engineering (51.87%), and 9 in other related fields such as facilities and cost engineering. Among engineers under 30, those without any site experience accounted for half of the responses, with many having only short-term site experience.

According to the survey results (Figure 3a), 45% of respondents showed interest in construction management work in Japan. Considering the number of young Chinese engineers, the market for those wishing to try working in Japan is substantial. In contrast to the decreasing number of young people under 29 in Japan, results from Figure 3b indicate that 78% of young Chinese engineers are interested in site management roles. Results from Figure 3c show that, in addition to BIM technology, respondents have acquired skills directly related to site management, such as drawing and surveying capabilities. Graduates from universities and professional schools possess expertise in mathematics and mechanics, providing a solid foundation for future certification and career advancement. According to the Ministry of Health, Labour and Welfare's Basic Survey of Wage Structure, the

Table 3. Basic attributes (N=187)

Item		Headcount	Proportion
Gender	Male	140	74.87%
	Female	47	25.13%
Educational background	Technical College	92	49.20%
	Undergraduate	85	45.45%
	Graduate student	10	5.35%
Major field	Architecture	97	43.32%
	Civil Engineering	81	51.87%
	Others	9	5.81%
On-site experience status	Inexperienced	98	52.41%
	0 to 2 months	55	29.41%
	2 months to 1 year	24	12.83%
	More than 1 year	10	5.35%

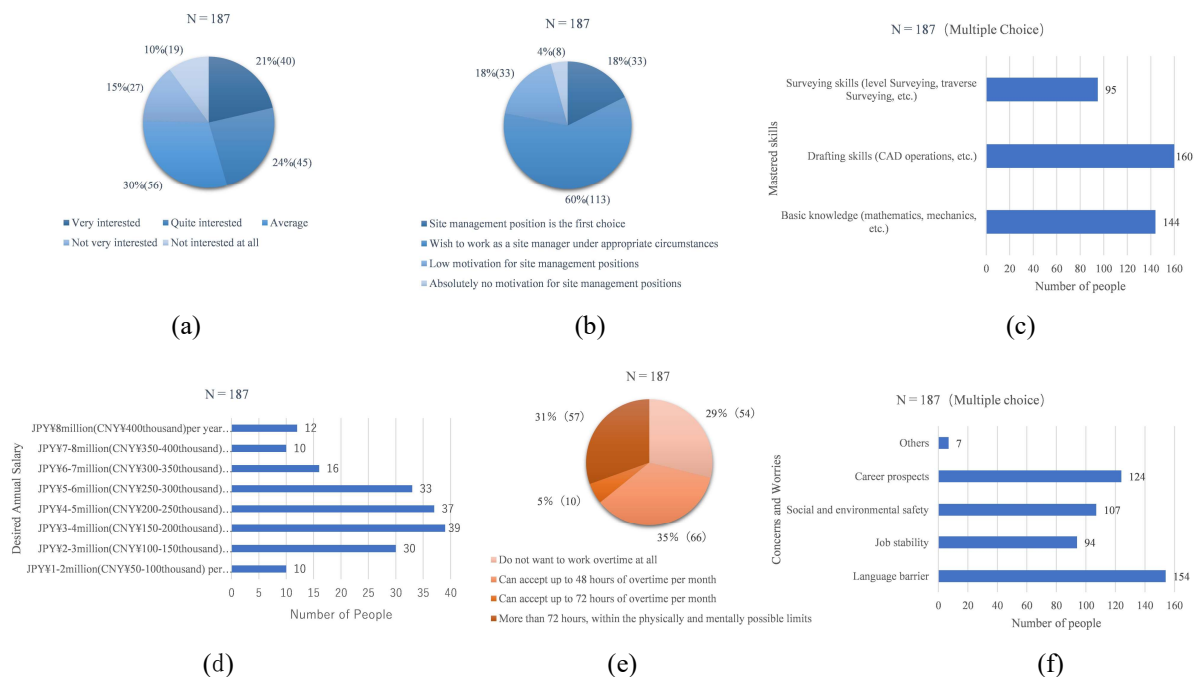


Figure 3. Survey results (a) Enthusiasm for field management in Japan (b) Enthusiasm for a position in field management (c) Skills acquired (aside from BIM technology) (d) Desired annual salary (e) Tolerance for Overtime (f) Points of concern

average annual salary for construction engineers, including site supervisors, is over 5 million yen. As shown in Figure 3d, the most desired minimum annual salary for construction site management positions in Japan ranges from 3 to 4 million yen, with a median of 4 to 5 million yen, which is within a feasible range for such positions in Japan. Regarding overtime, results from Figure 3e indicate that over 70% of young engineers are willing to accept some level of overtime. Considering Japanese labor reforms and the 45-hour monthly overtime cap, many engineers find this amount of overtime acceptable. Concerns (Figure 3f) show that the biggest worry among young engineers is language issues, with 154 out of 187 concerned about this, which is a significantly high proportion. Additionally, some are also concerned about the future prospects of their professions and the safety of society, which are issues worthy of discussion.

In this study, regression analysis was performed using SPSS, examining multiple factors such as gender, educational background, site management experience, understanding of Japanese culture and workplace environments, tolerance for overtime, and interest in construction site management roles in Japan. The analysis specifically investigated the impact of these factors on the intentions of young technicians to work in Japan.

The results (Table 4) revealed that an understanding of Japanese culture and workplace environments significantly positively influences the ability to accept employment in Japan (regression coefficient 0.84, greater than 0.5). This outcome suggests that a deep comprehension of Japanese culture and workplace environments is directly linked to the motivation of young technicians to engage actively in construction site management roles in Japan. Other independent variables (gender, educational background, site experience, and tolerance for overtime) did not show a statistically significant impact on the acceptance of jobs in Japan (regression coefficients less than 0.5). This indicates that these factors have minimal direct influence on the motivation of young technicians toward site management positions in Japan. The goodness of fit for the regression model was satisfactory ($R^2 = 0.289$), explaining 29% of the variance in the dependent variable. The F-value ($F(5,181) = 14.716, p < 0.001$) confirmed that the regression equation overall was statistically significant.

From this analysis, it is understood that the understanding of Japanese culture and workplace environments plays a crucial role in shaping the willingness of young technicians to undertake site management work in Japan. This finding underscores the importance of measures to promote cultural adaptation and understanding in educational programs and workplace design. It has been found that age, educational background, and site management experience in China have minimal impact on the motivation to engage in site management work. Additionally, the fact that tolerance for overtime does not influence their motivation suggests that providing more flexible working conditions could also attract their interest. This provides valuable insights for the Japanese construction industry in devising strategies to attract and retain international talent.

Table 4. Results of regression analysis (N=187)

Item	regression coefficient	t value	p value	VIF value
Gender	0.20	1.08	0.28	1.18
Educational background	-0.12	-0.89	0.38	1.07
On-site experience status	-0.13	-2.06	0.04	1.19
Understanding of Japanese culture and work environment	0.84	7.03	0.00	1.06
Tolerance for overtime	-0.05	-0.81	0.417	1.06
Number of observations	187			
R ² -value	0.289			
F -value	$F(5,181) = 14.716, P=0.000$			

4. A survey on the satisfaction of Chinese individuals engaged in construction management in Japan

4.1 Purpose and Method of the Survey

Following the investigation into the intentions of young Chinese engineers, a survey was conducted among 10 Chinese individuals who have worked (or are working) in construction management roles in Japan to understand their satisfaction levels and explore the challenges related to the integration of Chinese talent. The survey included 11 tailored questions: five collected basic personal information (gender, age, educational background, field of study, and on-site experience), while three assessed satisfaction with the work environment in Japanese site management, salary, and overall satisfaction with construction management duties. Additionally, three questions addressed personal views on language barriers, intention to reside in Japan, and whether the professional environment in construction management in Japan is better than in their home country. The survey was conducted entirely within the month of April 2024. Through this survey, the aim was to gain significant insights into the actual satisfaction levels, linguistic challenges, and the willingness to work in Japan from diverse perspectives, thereby evaluating the potential for integrating Chinese talent in Japanese construction management roles.

4.2 Survey Results

Table 5. Basic attributes (N=10)

Item		Headcount
Gender	Male	7(70%)
	Female	3(30%)
Age	20s	5(50%)
	30s	4(40%)
	40s	1(10%)
Educational background	Technical College	1(10%)
	Undergraduate	3(30%)
	Graduate student	6(60%)
Major field	Architecture	6(60%)
	Civil Engineering	4(40%)
Experience	0 to 2 years	3(30%)
	More than 5 year	5(50%)
	More than 5 year	2(20%)

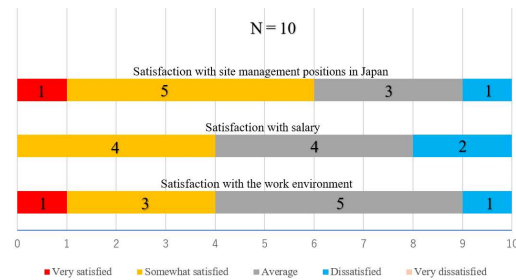


Figure 4. Customer satisfaction survey results

In consideration of the scope and representativeness of the survey, the basic attributes of the selected respondents are shown in Table 5. Among the 10 participants, there are 7 men and 3 women, roughly reflecting the gender ratio typical in the industry. Educational backgrounds include 1 technical college graduate, 3 university graduates, and 6 graduate degree holders, all of whom are engineers trained in architecture or civil engineering. Regarding fields of study, there are 6 in architecture and 4 in civil engineering. Experience on-site ranges from 0-2 years for 3 individuals, 2-5 years for 5, and over 5 years for 2 individuals.

Figure 4 displays the satisfaction levels of Chinese nationals engaged in construction management in Japan. Six respondents are somewhat satisfied with their site management work in Japan, three expressed neutral opinions, and one indicated some dissatisfaction. Breaking it down by specific areas, four people are satisfied with their salaries, four are neutral, and two are somewhat dissatisfied. Regarding the working environment, including labor conditions and workplace atmosphere, four are satisfied, five are neutral, and one is somewhat dissatisfied. This indicates a relatively high level of satisfaction among Chinese engaged in site management.

Survey results concerning unavoidable language barriers for foreign employees are shown in Figure 5a. No one feels that language is a very significant barrier or no barrier at all; one person feels it is a relatively significant barrier, four feel it is normal, and five, the majority, feel it is a minor barrier that does not impact their work. This suggests that Chinese nationals have a sufficient grasp of Japanese to prevent language from becoming a major obstacle.

The results concerning intentions to reside in Japan (Figure 5b) show that one individual has a very high intention to reside, five have relatively high intentions, and four maintain a neutral stance, with none expressing low or no intention. This implies that the favorable living conditions in Japan contribute to a high willingness among Chinese nationals to reside and work there, which in turn supports job sustainability.

Regarding whether construction management positions in Japan offer a better career than those in China (Figure 5c), seven people believe to varying degrees that they do, two are neutral, and one does not think so. Considering salary, job content, employment pressure, and work environment collectively, the majority rate Japanese site management roles more highly.

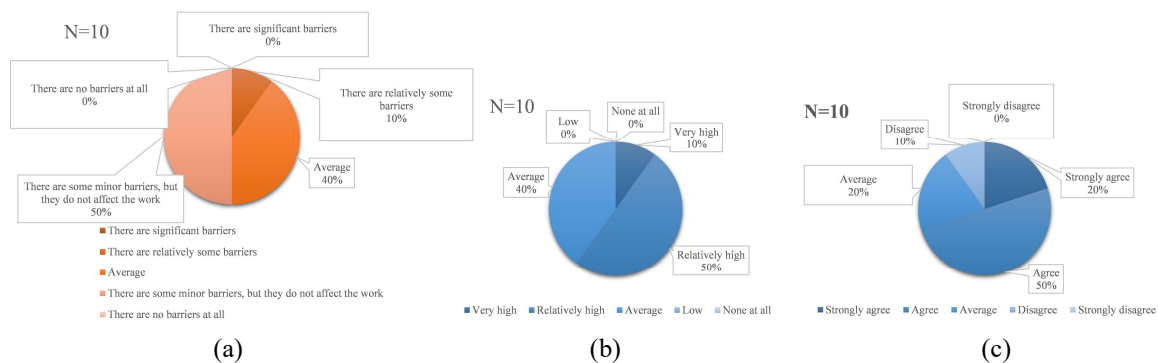


Figure 5. Survey results (a) Language barrier issue (b) Intention to reside in Japan (c) Provision of a better career

5. DISCUSSION

The survey research has highlighted the commonalities between Chinese and Japanese BIM-related education, notably in the generation and visualization of three-dimensional models. Amidst the transformation of the architectural industry and propelled by national initiatives, Chinese universities are at the forefront of research in BIM-related education and smart construction. The rapid construction growth nationally underscores the necessity of integrating BIM technology and site management applications into university curriculums. This strategic integration aims to prepare engineers capable of managing construction sites in Japan and solving complex problems with BIM technology.

Furthermore, surveys conducted among young engineers in Shanghai, Fujian, and Shandong reveal a robust eagerness to engage in site management roles in Japan, influenced minimally by gender, educational background, or local experience. Instead, familiarity with Japanese culture and workplace norms emerges as a crucial component. This insight stresses the need for tailored promotional efforts and training programs to facilitate the smooth transition of talent into the Japanese market. Japan's ongoing labor reforms, such as reducing overtime and enhancing work conditions for female employees, anticipate the integration of skilled Chinese professionals, which is vital for addressing Japan's labor shortages.

The satisfaction expressed by Chinese engineers working in Japan regarding their work conditions, salaries, and career opportunities, despite language and cultural differences, underscores the minimal barriers these differences pose. The structural similarities between the Chinese and Japanese languages offer unique advantages for Chinese engineers, who are expected to excel in roles requiring both technical skills and communication prowess. Many engineers also show a long-term commitment to working in Japan, attracted by superior career development opportunities, signaling positive prospects for sustained talent recruitment.

To further enhance the integration of Chinese technical talent into the Japanese workforce, three recommendations are proposed:

Strengthen Sino-Japanese Educational Cooperation: Boost exchanges and collaborations between universities in both nations in the realms of BIM technology and construction management. Regular updates to educational content are essential to align with ongoing industry developments.

Provide Language and Professional Training Through Third Parties: Implement language training and specialized professional development programs through third-party organizations to ensure these professionals can swiftly adapt to the Japanese societal and workplace environments.

Enhance On-the-Job Training and Career Development: Urge Japanese construction firms to offer continuous on-the-job training and career development opportunities, aiming to enhance job satisfaction and workforce integration for foreign technical talent.

6. CONCLUSIONS

This study confirmed the potential contribution of young Chinese BIM technicians to addressing the shortage of construction site managers in Japan, based on comparative surveys of BIM education content at universities in China and Japan, surveys on the willingness of young Chinese technicians trained in BIM to engage in construction management in Japan, and satisfaction surveys of Chinese individuals working in construction management in Japan. It identified key factors influencing the applicability of young Chinese BIM talent, their expected salaries, and their work attitudes, demonstrating the feasibility of addressing these challenges.

However, this study has not yet surveyed the perceptions of companies in Japan that actually employ site managers, which is a direction for future research. Furthermore, by conducting and comparing surveys on BIM education and the employment intentions of young technicians in universities across Southeast Asian countries, differences in integrating technicians from different countries can be clarified. This approach will explore better ways to resolve Japanese construction industry labor shortage.

ACKNOWLEDGMENTS

This work was supported by JST SPRING, Grant Number JPMJSP2135.

REFERENCES

- General Incorporated Association, Japan Federation of Construction Contractors. (2024). Current Status of the Construction Industry. Construction Industry Digital Handbook, Available at: <https://www.nikkenren.com/publication/handbook/chart6-4/index.html> (Accessed: Sept. 1, 2024).
- Ministry of Education, Culture, Sports, Science and Technology. (2024). Basic School Survey, Available at: https://www.mext.go.jp/b_menu/toukei/chousa01/kihon/1267995.htm (Accessed: Sept. 1, 2024).
- Meng, Q., Zhang, Y., Li, Z., Shi, W., Wang, J., Sun, Y., Xu, L., Wang, X. (2020). A Review of Integrated Applications of BIM and Related Technologies in Whole Building Life Cycle, *Engineering Construction and Architectural Management*, 27, 1647–1677.
- Jin, R., Zhong, B., Ma Hashemi, A., Ding, L. (2019). Integrating BIM with Building Performance Analysis in

- Project Life-Cycle, *Automation in Construction*, 106, 102861.
- Conrad, H., Meyer-Ohle, H. (2019). Overcoming the Ethnocentric Firm? Foreign Fresh University Graduate Employment in Japan as a New International Human Resource Development Method, *The International Journal of Human Resource Management*, 30(17), 2525–2543.
- Oishi, N. (2014). Redefining the ‘Highly Skilled’: The Points-Based System for Highly Skilled Foreign Professionals in Japan, *Asian and Pacific Migration Journal*, 23(4), 421–450.
- Banh, T. (2023). BIM Education-Global–2023 Update Report.
- Ministry of Land, Infrastructure, Transport and Tourism. (2022). Results of the Construction Industry Activity Survey, Available at: https://www.mlit.go.jp/report/press/joho04_hh_001185.html (Accessed: Sept. 1, 2024).
- Xu, J., Li, B. K., Luo, S. M. (2018). Practice and Exploration on Teaching Reform of Engineering Project Management Course in Universities Based on BIM Simulation Technology, *Eurasia Journal of Mathematics, Science and Technology Education*, 14(5), 1827-1835.
- Wang, L., Yan, X., Fan, B., Jin, R., Yang, T., Kapogiannis, G. (2020). Incorporating BIM in the Final Semester Undergraduate Project of Construction Management—A Case Study in Fuzhou University, *KSCE Journal of Civil Engineering*, 24(8), 2403-2418.

BIM to FEM model conversion of low-rise residential buildings for debris flow damage analysis

Xuhong HUANG¹ and Yantao YU²

1) Ph.D. Student, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: xhuanget@connect.ust.hk

2) Ph.D., Asst. Prof., Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: ceyantao@ust.hk

Abstract: In Hong Kong, low-rise residential buildings, which are mostly built next to mountains and occupy half of the total residential area, are often seriously damaged by debris flow. The extent of damage to each building depends not only on the intensity of the debris flow but also on the interaction of the neighboring components of the building and their collective performance. With the application of BIM (Building Information Modeling) technology in structural analysis, data conversion and information transfer between the BIM model and the FEM (Finite Element Method) model have become the main factors limiting its efficiency and quality. Unlike the automatic conversion method for frame structure building models, the conversion method for such models is a semi-automated method that relies on manual labor at the cost of losing a large amount of BIM information due to the problems of non-standardization and confusing information. Therefore, this study proposes an automatic standardized conversion method based on OpenBIM data (IFC) for developing refined FEM models of low-rise residential buildings. It consists of three parts: 1) guaranteeing model refinement and spatial robustness through model geometry transformation and calibration; 2) standardized conversion during the conversion process through material information database establishment and matching; and 3) integration of the output information to achieve data interoperability and provide a reference for debris flow damage assessment. Conversion tests of this method and the traditional conversion method were conducted in the same BIM model, and the results show that the FEM model generated by this method is significantly better than the traditional conversion method in terms of spatial closure and mesh quality (i.e., spatial robustness and model refinement). Future research will continue refining spatial robustness detection, standardized conversion, and data interoperability.

Keywords: building information modeling (BIM); finite element model; industry foundation classes (IFC); model conversion; debris flow

1. INTRODUCTION

Debris flows, as one of the most frequent geohazards in mountainous areas (Zhang et al., 2018), pose a significant threat to human lives, buildings, and critical infrastructure (Tang et al., 2011). Debris flow events occurring in populated mountainous areas are catastrophic, especially in areas with high populations and building densities (Prieto et al., 2018; Zhou et al., 2019). Facilities exposed to debris flow threats include buildings, roads, railroads, mines, and reservoirs. Residential buildings are more vulnerable to debris flow damage than other facilities (Hu et al., 2012).

Low-rise residential buildings, as one of the types of residential buildings, are commonly found in small houses (Hong Kong village houses) of indigenous villagers in Hong Kong (*New Territories Exempted Houses (NTEH) - Buildings Department*, 2024). According to the statistics of the Hong Kong Housing Bureau and the Hong Kong Planning Department in 2023, village houses take up 48% of the housing land in Hong Kong (*Planning Department - Land Utilization in Hong Kong*, 2024). As more than 70% of Hong Kong's land is mountainous, debris flows are not uncommon in Hong Kong (Chau & Lo, 2004). Meanwhile, village houses in Hong Kong are often clustered at high densities, and there are instances where rooms in village houses are separated to accommodate more residents (Aijmer, 1975; Nelson, 1969). This makes debris flows catastrophic for the areas where village houses are clustered in Hong Kong. Therefore, debris flow damage assessment of low-rise residential buildings in Hong Kong would be useful for more effective risk assessment, emergency management, and housing maintenance.

Debris flow damage assessment has been extensively studied. Existing studies consist mainly of regional assessments and assessments of individual buildings.

1.1 Regional Damage Assessment

Regional damage assessment calculates the overall damage level of a house by generalizing buildings into categories (e.g., building age, building structure type, etc.) and applying predefined curves (Amirebrahimi et al., 2015; de Moel et al., 2015). While the validity of these methods has been demonstrated in large-scale assessments (Hu et al., 2012; Zhang et al., 2018), they are not applicable to damage assessments such as those for Hong Kong village houses. Hong Kong village houses are exempted from Hong Kong's laws and regulations, which makes the information on materials, spatial relationship of building components, and building structure vary from one

village house to another (*New Territories Exempted Houses (NTEH)* - Buildings Department, 2024). This is mainly because regional assessment methods have limited input data and are often based on regional information to output a quantitative rating of the overall building physical damage or a single figure for the overall building economic loss (Ciurean et al., 2017; Hu et al., 2012; Kang & Kim, 2016; Mejia-Navarro et al., 1994). As a result, regional assessments cannot provide significant feedback on differences within different buildings (e.g., spatial relationships between house components, material properties, information in component attributes such as size and shape).

1.2 Individual Building Damage Assessment

Compared with regional assessments, the assessment of individual buildings can provide better feedback on the differences within different buildings. However, all existing individual building assessment methods often use simplified building models or typical building models for finite element analysis (Huan et al., 2024; P. Li et al., 2017; Liu et al., 2018; Luo et al., 2019, 2020; Milanese et al., 2018) rather than performing finite element analysis of different buildings (e.g., different buildings do not have the same floor height, occupation shape, and spatial relationship of components inside the building). Moreover, unlike the automated model conversion for frame-structured buildings, the conversion of BIM models to FEM models for such buildings is semi-automated and based on proprietary software (e.g. Revit). This is mainly due to limited computational resources and the need for increased computational efficiency. This is especially because of the financial and time constraints that would be imposed if detailed damage assessments were used on many buildings (Jeong & Elnashai, 2007; Kwon & Elnashai, 2006). Simplified models often downscale the 3D components of a building to one-dimensional units or two-dimensional units (e.g., one-dimensional beams, one-dimensional columns, and two-dimensional walls) (Feng et al., 2019; Milanese et al., 2018; Taucer et al., n.d.). Typical building models are based on the structural characteristics of the buildings in the study area to represent all buildings in the study area (Huan et al., 2024; Liu et al., 2018; Luo et al., 2019). However, the simplified models cannot correspond one-to-one with the structural geometry information of the buildings themselves, thus failing to identify building-specific damage components. The typical model cannot correspond with the actual buildings in the study area because it is a generalized model. As for low-rise residential buildings consisting of components with different structural information, their model conversion is often labor-intensive and not a fully automated process. Their model conversion is also difficult to apply with the above methods. Data conversion and information transfer between BIM models and FEM models have become major constraints to their efficiency and quality.

1.3 BIM-to-FEM under Damage Assessment

Individual building damage assessment of debris flows is more challenging compared to hurricanes, floods, and earthquakes. Unlike floods, where the primary considerations are immersion damage and lateral pressures on buildings, debris flows, which are similar to floods, primarily consider the response of building components under impact pressures (Ali Khan et al., 2024; Luo et al., 2020). The building damage assessment caused by debris flows often needs to consider the dynamic response analysis of the interaction between the debris flow and the building (Luo et al., 2020), i.e., the cumulative damage after the debris flow damages the building components at different moments of the debris flow from the beginning of the contact with the building to the end of the complete encirclement of the building. In contrast, the building damage assessments caused by hurricanes and earthquakes consider the mechanical analysis of the static or overall dynamic response (Charalambos et al., 2014; Nofal et al., 2022; Quinay et al., 2020). As for the flood damage assessment, its analysis of the BIM model of the building is only a cumulative analysis of static mechanical analysis with duration (Amirebrahimi et al., 2016; Aribisala et al., 2022). Other hazards require less refinement and robustness of the models converted by BIM (Charalambos et al., 2014; Nofal et al., 2022; Quinay et al., 2020). For example, the BIM model was transformed into a shell model with only the building facade in the hurricane case (Nofal et al., 2022). The BIM model was transformed into a simplified model with the row frame in the earthquake case (Quinay et al., 2020; Ren et al., 2019). Unlike the finite element analysis described above, the finite element analysis of the damage caused to the building by the debris flow is a coupled fluid-structure collision analysis. This requires that the model transformation has the robustness of automatically matching different material information for each component transformed and spatial relationships among the transformed components.

1.4 Research Gaps

It is observed that most of the model conversion methods, when faced with the model of Hong Kong village houses, which is not standardized and has confusing information, 1) need to rely on labor-intensive manual operations and are not able to standardize the conversion of the model automatically; 2) need to simplify the components of the model to one or two dimensions to reduce the computational usage, and are not able to reduce the model refinement level without excessively reducing the model refinement level; and 3) the loss of the BIM information in the process of the conversion causes the spatial relationship between components to be The loss of BIM information during the conversion process makes the spatial relationship between the components not touch

or overlap each other, and the spatial robustness of the model in the FEM cannot be guaranteed. As a result, it can not automatically standardize the model transformation and ensure the model's high accuracy and spatial robustness, which affects the effectiveness of the debris flow damage assessment of the village houses in Hong Kong.

1.5 Research Objectives and Questions

This study aims to propose a preprocessing method from OpenBIM data (IFC) to develop FEM models of low-rise residential buildings. The objectives of the study are (1) to transform a normalized transformation of the IFC model automatically, (2) to reduce the level of refinement based on the non-excessive reduction of the converted model of the IFC model, and (3) to guarantee the spatial robustness of the converted model. To achieve these objectives, the subsequent questions will be answered.

First, how to standardize the conversion of BIM models with problems such as irregular establishment and confusing information of different components? Although 'IfcMaterial' of 'IfcMaterialDefinition' and 'IfcProperties' of 'IfcPropertyDefinition' can extract model information and provide a way to standardize it. However, the 'IfcMaterial' and 'IfcProperties' of the above BIM model built at the early stage are not set, or their information sets are confusing. Moreover, the 'IfcMaterial' and 'IfcProperties' of each BIM model are set differently for each due to the model creators' different habits. These problems can lead to loss of information, confusing material information, or even garbled information in some parts of the components of the converted model.

Second, how to control the degree of simplification of the model during the conversion process? BIM contains digital parameters of a large number of building components. This large amount of information brought by BIM makes the debris flow damage assessment with numerical simulation, which will cause the problem of excessive computational volume and a long computational period. Excessive computational volume will lead to an increase in computational cost, which is why previous studies have commonly used the method of simplified modeling of typical buildings to represent all the buildings in the study area to reduce computational usage. However, those methods lose most of the building information. If the model is not simplified as in the above approach, the computational usage for analyzing the model will increase significantly. Figure 1(a) shows the model component after automatic conversion by previous methods without simplification (From Revit). The component is a window in the building, which is split into finer components under this method. These finer components lose their physical meaning as windows and cause a sudden increase in the mesh density in this area, significantly increasing computational usage.

Third, how to guarantee the spatial robustness of the converted model? Existing methods often lose more BIM information and cause the spatial relationships between components to not touch or overlap when converting the above BIM model. These errors in spatial relationships are not reported in the BIM, but they are explicitly reported in the FEM analysis and even implicitly affect the analysis results in the FEM. Figure 1(b) shows the problem of components overlapping in the converted model. The blue part is the wall and column of the model, which spatially overlap with each other making the meshing of the part collapse, and then the model cannot be analyzed by FEM. Figure 1(c) shows the properties of each component of the converted model in FEM. The model loses BIM information during the conversion process and leaves the components spatially out of contact, resulting in the inability to generate the correct lines and surfaces to form solids.



Figure 1. Example diagram of previous model conversion results

To answer these questions, an IFC-based model transformation method is proposed. The method consists of three parts: model geometry transformation and calibration, establishment and matching of the material information database, and output information integration. The paper is organized as follows: Section 2 provides a partial overview of the proposed methodology; Section 3 describes the experimental setup and demonstrates the method's current results; Section 4 shows future work; Section 5 presents the conclusions drawn from the study.

2. METHOD

The steps of the method proposed in this paper are shown in Figure 2. In the first part, the Python code is used to realize the geometric entity transformation of each model component and verify each component's closure and the spatial relationship's robustness. In this, the geometric entity transformation constructs line faces by finding the maximum outer contour envelope of the corresponding component to form the entity. This allows the model to reduce the computational usage of FEM analysis without unduly reducing the level of detail. The stitching relationship of each set of line faces then verifies the closure of each component. The second part establishes the interface for importing material information, and the database is automatically generated according to the corresponding component types. The material information matching is realized by Python code on the FEM platform side. Among them, the generated database can be used to standardize the model information by external input information when the model information is lost or confused. Then the standardized conversion of the model is realized through the information matching at the FEM platform. In the third part, the FEM results are merged with the corresponding component information to generate a database and provide data visualization. In particular, the final generated database enables the analysis results to be data-interoperable in dual platforms, which can help the mudslide damage assessment of village houses in Hong Kong.

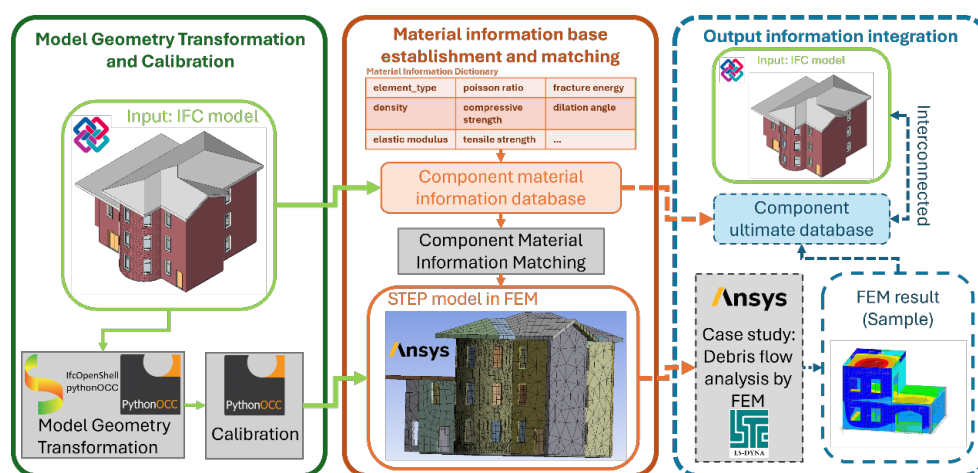


Figure 2. The overall framework of the methodology

2.1 Model Geometry Transformation and Calibration

To avoid being restricted by proprietary software, this study is consistent with the openBIM method and uses the IFC model as the input model. Some of the acquired public BIM models in RVT file format are exported to IFC files (IFC4) through the built-in IFC interpreter of Revit software (*Revit IFC Resources / Manual*, n.d.). The model conversion method between IFC and FEM is implemented through Python code, and an interface is provided for importing material information.

Model transformation requires extracting components within the model, and a BIM model usually contains the structural components, wiring components, furniture components, etc. (N. Li et al., 2020). Only structural components in the model are extracted to analyze the damage caused by debris flows on buildings. Components such as furniture components, which highly occupy the computational resources of FEM but do not play a role in analyzing the debris flow impact on the building structure, are not extracted. For structural components, the following components were extracted based on Giuffrè's work (Giuffrè, 1993): (1) walls, (2) beams, (3) columns, (4) floor slabs, (5) roofs, (6) doors, (7) windows, and (8) foundations. The following "IfcBuildingElement" are needed to represent the different objects mentioned earlier: "IfcWall" for walls, "IfcBeam" for primary and secondary beams, "IfcColumn" for columns, "IfcSlab" for floor slabs, "IfcRoof" for roofs, "IfcDoor" for doors, "IfcWindow" for windows, and "IfcFooting" for foundations. Each IFC object belongs to a specific "IfcBuildingElementType." The component types are shown in Figure 3.



Figure 3. Schematic diagram of the types of components extracted

In this paper, if the component material is masonry, reinforced concrete, etc., the geometry of the component is modeled as an equivalent homogeneous entity without the need to geometrically model the wall bricks, mortar, reinforcement bars, etc. Similarly, when the component type is door and window, the largest contour in the component is found as the corresponding equivalent homogeneous entity without geometrically modeling the battens, glazing, frames, etc. If components such as walls, doors, windows, etc. are modeled as smaller solids such as wall tiles, retaining strips, glass, etc., these smaller solids would not be able to represent the components such as walls, doors, windows, etc., as having their physical meaning. This approach reduces the complexity of the converted model while preserving the component information and making the component still physically meaningful. For the IFC model view format imported into the conversion method, this paper selects the Design Transfer View instead of the Reference View. This is mainly because the Design Transfer View is used to share component information, such as spatial relationships and geometry of components. “The Reference View is mainly used for collision detection between components.

The model geometry conversion algorithm converts the IFC file to a STEP file, which enables the FEM platform to read the model file and realize the debris flow analysis. Although IFC to STEP conversion can be achieved directly and automatically by proprietary software such as Revit, its conversion function cannot control the component fineness of the resulting model, as shown in Figure 4. In this case, the frames, glazing, and retaining strips between the glazing of the door and window were converted by this function into entities that are independent of each other. This caused the door and window, which should have been treated as a whole in the FEM platform, to lose their physical meaning. At the same time, the subcomponents located in the window and door sections are small in size, which leads to a drastic increase in mesh density when meshing in FEM, thereby increasing the computational cost.

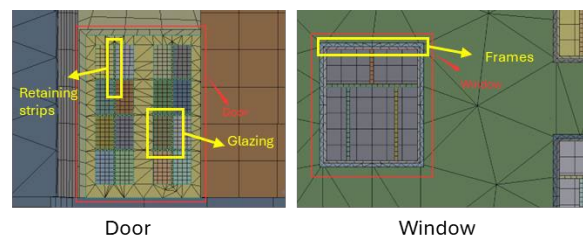


Figure 4. Drastically increasing mesh density and components that lose physical meaning

The details of the model geometry transformation algorithm are as follows:

Iterate over all the IfcElements collected in the input model and classify each component by a “for” loop relying on the “element.is_a()” function. At the same time, the “corresponding type + name” is stored as its component name in the FEM model and the primary key in the material information database. The geometry of each IFC component is extracted in B-Rep format using “PythonOCC.” The surfaces are extracted from the shape composite (TopoDs composite) of the IFC component when it is converted to a STEP component. At this point, these surfaces are recognized by “PythonOCC” as sub-levels of the component. These surfaces are collaged into a true 3D solid by their envelopes using the sewing function. When encountering an IFC component that has subcomponents, the component's entity in the STEP file is created by finding the largest contour in the component

and creating the smallest cube that completely wraps that contour. Finally, the sewing function is used again for each converted entity to check whether each face is properly connected according to the envelope. The flowchart of the algorithm is shown in Figure 5.

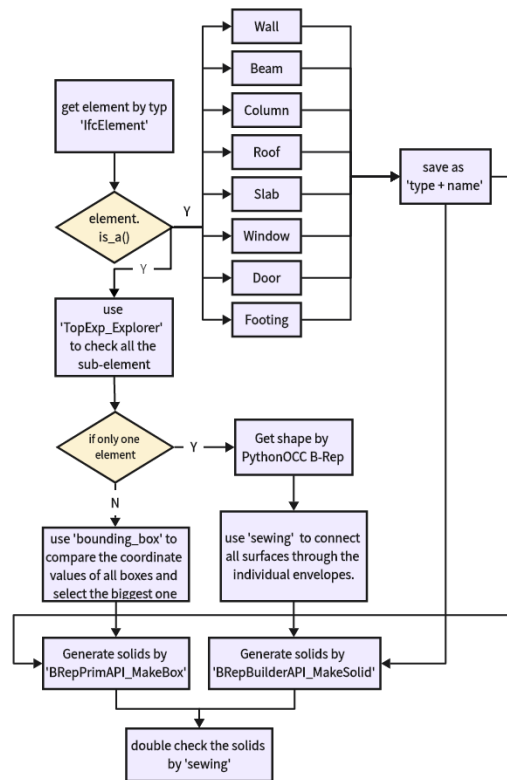


Figure 5. Flowchart of the model geometry transformation algorithm

2.2 Material Information Database Establishment and Matching

In real projects, the original BIM model may have problems such as non-standardized establishment, non-harmonized standards, and the confusing information of different components. This makes it difficult to utilize the 'IfcMaterial' of the 'IfcMaterialDefinition' and the 'IfcPropertyDefinition' of the 'IfcPropertyDefinition' in the IFC architecture. 'IfcMaterial' and 'IfcProperties' will have problems when modeling the component materials, as shown in Figure 6. The material information extracted from (a) and (b) in Figure 6 comes from different IFC models, and due to the above problem, the material information in (a) is confusing, while in (b), the material information is even missing. For the IFC models that may have the above problems, it is necessary to supplement the material information from the outside world during the model conversion and correspond the material information to each component in the IFC model.

'333322',,BAM-3, #133581-Ifcbeam('00MgawVf	Clad - White
'NormIgn-Viga rectangular:Viga .15 x 35 cm',	Wood - Stained
#18,'NormIgn-Viga rectangular:Viga .15 x 35 :	None
'333323',,BAM-3, #133587-Ifcbeam('00MgawVf	None
'NormIgn-Viga rectangular:Viga .15 x 35 cm',	Clad - White
#18,'NormIgn-Viga rectangular:Viga .15 x 35 :	Wood - Stained
'333363',,BAM-3, #127899-Ifcbeam('1WJaurZ	None
cm:349138',S,'NormIgn-Viga rectangular:Pergo	None
('1WJaurZv88737L9uJ5',#18,'NormIgn-Viga :	None
rectangular:Pergolas 10 x 20 cm',#128093,#128	Clad - White
'NormIgn-Viga rectangular:Pergolas 10 x 20 cm	Wood - Stained
#128237,'349138',,J0157,3, #128299-Ifcbeam('1	None
cm:349131',S,'NormIgn-Viga rectangular:Viga	None
('1WJaurZv88737L9uJ5',#18,'NormIgn-Viga :	Aluminium
rectangular:Pergolas 10 x 20 cm',#128503,#128	Aluminium
'NormIgn-Viga rectangular:Pergolas 10 x 20 cm	Aluminium
#128605,'349138',,J0157,3, #128724-Ifcbeam('1	Aluminium
20 cm:349138',S,'NormIgn-Viga rectangular:Pe	Aluminium
('1WJaurZv88737L9uJ5',#18,'NormIgn-Viga :	Aluminium
rectangular:Pergolas 10 x 20 cm',#128925,#128	Aluminium
'NormIgn-Viga rectangular:Pergolas 10 x 20 cm	Aluminium
#129094,'349138',,J0157,3, #129122-Ifcbeam('1	Aluminium
20 cm:349138',S,'NormIgn-Viga rectangular:Pe	Aluminium
('129094ar3J0u2UyWboc',#18,'NormIgn-Viga :	Aluminium
rectangular:Pergolas 10 x 20 cm',#129317,#129	Aluminium

(a) Confusing material information

(b) Missing material information

Figure 6. Example diagram of material information in the IFC model

The details of the model material information conversion algorithm are as follows:

Iterate over all the IfcElements collected in the input model and classify each component by a “for” loop relying on the “element.is_a()” function. At the same time, the “corresponding type + name” is stored as its component name in the FEM model and the primary key in the material information database.

Iterate over all the IfcElements collected in the input model, and through a “for” loop, rely on the ‘element.is_a()’ function to categorize each component and store the “corresponding type + name” as its primary key in the material information database. Then, through a “for” loop, the ‘RelatingMaterial’ function is used to extract the related material name. At the same time, read the externally defined material information, match the traversed components with the material names in the material information, and add the corresponding material information to the material information database after successful matching. If the match is unsuccessful, the default material information of the component type is added to the material information database and printed to correspond to the primary key value. Similarly, create a “for” loop of the same level as the “for” loop that relies on the ‘RelatingMaterial’ function. This loop relies on the ‘RelatingType’ function to perform the same operations as its sibling “for” loop. Finally, a CSV file is generated so that the material information of each component can be automatically added to the model utilizing the APDL command flow in the FEM platform. The flowchart of the algorithm is shown in Figure 7.

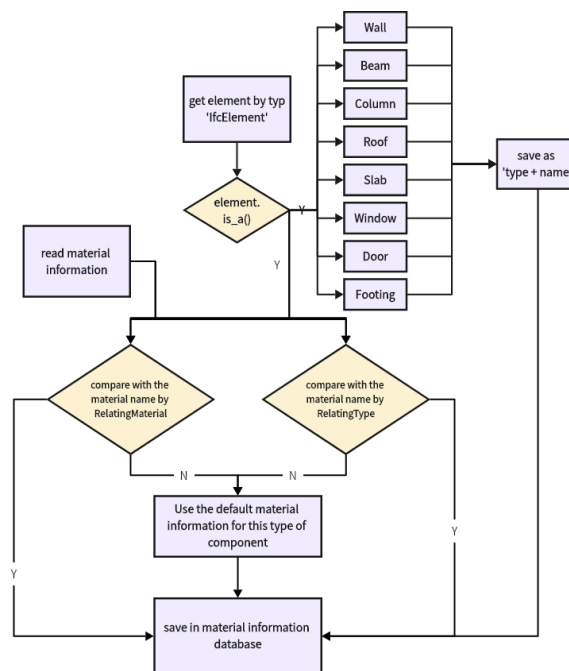


Figure 7. Flowchart of material information database establishment and matching algorithm

3. EXPERIMENT AND RESULT

3.1 Experimental Setup

To verify the feasibility and reliability of the proposed methodology and to fit the actual situation of Hong Kong village houses, the experiments were conducted using public BIM models downloaded from online sources. Most of the downloaded models face problems such as the version being too old (component information is lost in the version upgrade) and the component confusing information. In this experiment, randomly selected one model from the downloaded BIM models is used for preliminary validation. Among them, the model base information is shown in Table 1, the model overview diagram in Figure 8, and the material information in Table 2.

Table 1. Basic information about the selected BIM model

	Model
Revit version	2011
Number of stories	3
Housing structure	brick hybrid structure
Material information completeness	incompleteness

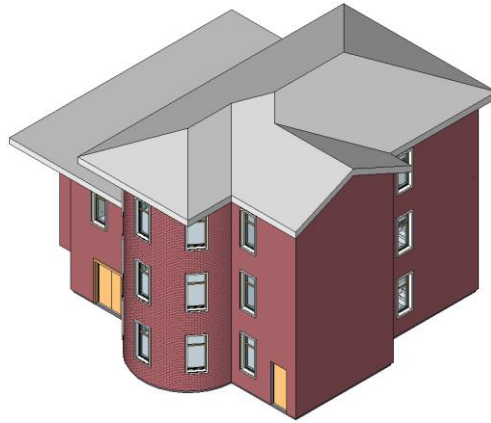


Figure 8. Overview diagram of the selected BIM model

 Table 2. Material information source
 (Kržan et al., 2015), (P. Li et al., 2017), (Milanesi et al., 2018), (Luo et al., 2020)

element type	Material in IFC	density (kg/m ³)	elastic modulus (MPa)	Poisson ratio	compressive strength (MPa)	tensile strength (MPa)	fracture energy (KJ/m ²)	dilation angle (degree)
IfcWall	default	1800	3000	0.2	3	1.7	0.055	20
IfcWall	StructuralMasonry6	2200	3000	0.2	3	1.7	0.055	20
IfcWall	reinforced concrete	2400	25000	0.2	30	3	0.1	35
IfcRoof	default	1800	3000	0.2	3	0.03	0.055	20
IfcSlab	default	2400	25000	0.2	30	3	0.1	35
IfcFooting	default	2400	25000	0.2	30	3	0.1	35
IfcWindow	default	2500	70000	0.2	120	35	0.015	10
IfcDoor	default	700	11000	0.3	40	60	0.035	15

The above building was modeled in Autodesk Revit and exported to IFC4 Design Transfer View via Revit's built-in IFC interpreter. The structural model was subsequently loaded into the code folder. After running the code, the corresponding STEP file and the corresponding material information database were automatically generated, and the STEP file was imported into the Ansys FEM platform used in this experiment. Finally, the quality of the model transformation at the current stage is compared using Ansys meshing.

3.2 Result

The feasibility and reliability of the proposed method can be initially assessed by visualizing the generated STEP file and viewing the component properties in Ansys. In the proprietary software approach, the above BIM model was directly exported as a STEP file through Revit. In the method of this paper, the above BIM model is exported to an IFC file through Revit and converted to a STEP file through the proposed method. Then, the models are imported into Ansys and meshed using the built-in mesh generator with the same default configuration. Finally, the model quality is assessed by preliminary comparison and comparison of mesh quality criteria.

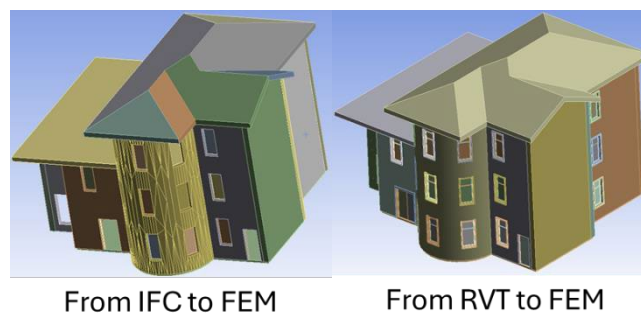


Figure 9. Overview comparison of converted models

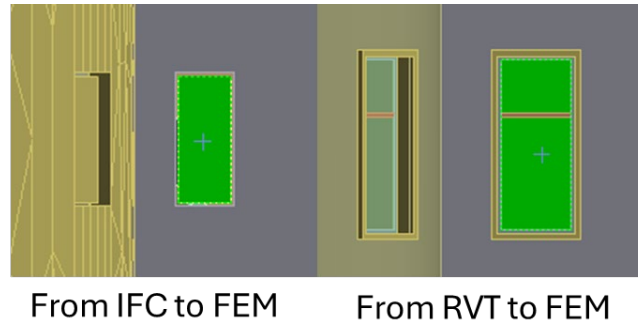


Figure 10. Detailed comparison of the converted components

Table 3. Conversion model component statistics

	From IFC to FEM	From RVT to FEM
Total	119	444
Solid	119	331
Surface	0	113

A preliminary comparison reveals that the proposed method can transform the component as a whole without continuing to split it when targeting components with sub-levels, such as doors and windows, as shown in Figure 10. This keeps the component from losing its physical meaning. Table 3 illustrates that the proposed method ensures that the converted components are closed and also reflects that the method reduces the number of split components. Therefore, it is tentatively concluded that the proposed method is more advantageous than the conversion function of the proprietary software.

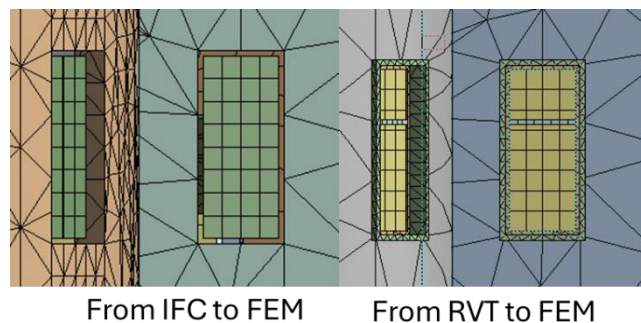


Figure 11. Mesh comparison of converted models

Table 4. Statistical table of grid quality standards

	From IFC to FEM	From RVT to FEM
Element Quality (Avg.)	0.71	0.43
Aspect Ratio (Avg.)	4.92	7.07
Jacobian Ratio (Avg.)	0.986	0.875

For the above two models generated by different conversion methods, the meshing is done with default control cells, and the cell size, resolution, and other parameters are consistent. For FEM, components such as doors and windows should not occupy a large amount of computational resources in the structural analysis. Therefore, the number of meshes in this region should be minimized, and the mesh density should be reduced when meshing. Figure 11 illustrates that the method significantly reduces the mesh density in this region. In terms of data, the mesh quality standard can well reflect the mesh quality of the whole model (Sorgente et al., 2023). As shown in Table 4, Element Quality (Avg.) is a generalized mesh checking criterion, and the larger the value, the better. Moreover, Element Quality (Avg.) should be greater than 0.7; otherwise, the error is significant. The Aspect Ratio (Avg.) should be less than 7 as much as possible to obtain a better displacement solution. The Jacobian Ratio (Avg.) is in the range of 0-1, and the value of 1 indicates a perfect mesh. The lower value indicates a poor mesh. Negative values indicate the presence of a negative volume mesh, which cannot be accepted by the solver. In the default configuration without human intervention, the quality of the models generated by this method after automatic meshing meets the mesh quality criteria. The quality of the model generated by the dedicated software did not

meet the mesh quality criteria. The model generated by this method is more advantageous in terms of closure and mesh quality compared to the conversion methods of specialized software.

4. FUTURE WORKS

In terms of spatial relationships of components, the spatial relationships between the transformed components do not touch or overlap each other, making them lack robustness. The finite element analysis is needed to eliminate this adverse effect. Therefore, future work may extract more component information to solve this problem. This is because more component information will help determine whether the spatial relationships between components are contained or overlapping conflicts, and more coordinate information will be recorded to help recover from coordinate misalignment after transformation. Regarding material matching, only the creation of a material information database has been completed. Adding material information to each component to match it in the FEM platform has not yet been completed. Future work will be done to interoperate the data by writing APDL command flow scripts. For the third component, future work may be done to interoperate the data by creating a database that is accessible to both BIM and FEM.

5. CONCLUSIONS

The establishment of a BIM to FEM low-rise residential building model conversion method for debris flow analysis is helpful for debris flow damage assessment of village houses in Hong Kong. Then, it is difficult to automatically convert BIM models into FEM models and achieve model robustness due to the non-standardization and confusing information of each BIM model. The aim of this study is to convert the non-standardized BIM models into FEM models for debris flow FEM analysis and to achieve model robustness. For this purpose, a preprocessing method for automatic conversion of BIM models to FEM models was developed. The method has the following innovations compared to previous studies.

The goal of the Model Geometry Conversion and Calibration section is to automate the conversion of BIM models into FEM models. Previous studies have used simplified and typical models to represent the target model or generated FEM models through labor-intensive manual conversion. That is because the existing BIM models are unstandardized and have confusing component information. As a result, the robustness of the converted models and the accuracy of component information are susceptible to conflicting spatial relationships, confusing material information, and difficulty in closing components into solids. In contrast, the first part is based on 'IfcOpenShell' and 'pythonOCC' to design the closure detection algorithm and component splitting degree detection algorithm. The experimental results show that the method exceeds the conversion effect of proprietary software in terms of closure and model quality.

The material information database creation and matching section is designed to extract the component information from the BIM and add the missing or confusing information through the matching process. Previous studies have used 'IfcMaterial' and 'IfcProperties' from IFC to match the material information, but this method fails when faced with non-standardized and confusing BIM models. In contrast, Part II not only supplements the missing or confusing information but also supports the data integration in Part III.

Finally, future goals of this work include (1) detection and repair of spatial relationships between components, (2) matching and assignment of material information, and (3) design of dual-platform data interoperability after data integration.

ACKNOWLEDGMENTS

The authors would like to acknowledge the Theme-based Research Scheme (T22-60623R) supported by University Grants Committee of The Government of the Hong Kong Special Administrative Region, Innovation Technological Fund Platform Research Projects (ITP/002/22LP) supported by Innovation Technological Center of The Government of the Hong Kong Special Administrative Region, Guangdong Basic and Applied Basic Research Foundation (SZC0301) funded by Department of Science and Technology of Guangdong Province.

REFERENCES

- Aijmer, G. (1975). An Enquiry Into Chinese Settlement Patterns: The Rural Squatters of Hong Kong. *Man*, 10(4), 559–570. <https://doi.org/10.2307/2800133>
- Ali Khan, M., Mustaffa, Z., Sati Hamonangan Harahap, I., & Ben Seghier, M. E. A. (2024). A comprehensive approach for understanding debris flow interaction with pipelines through dynamic impact pressure modeling. *Engineering Failure Analysis*, 162, 108383. <https://doi.org/10.1016/j.engfailanal.2024.108383>
- Amirebrahimi, S., Rajabifard, A., Mendis, P., & Ngo, T. (2015). A data model for integrating GIS and BIM for assessment and 3D visualisation of flood damage to building. *CEUR Workshop Proceedings*, 1323, 78–89.
- Amirebrahimi, S., Rajabifard, A., Mendis, P., & Ngo, T. (2016). A framework for a microscale flood damage assessment and visualization for a building using BIM–GIS integration. *International Journal of Digital*

- Earth*, 9(4), 363–386. <https://doi.org/10.1080/17538947.2015.1034201>
- Aribisala, O. D., Yum, S.-G., Adhikari, M. D., & Song, M.-S. (2022). Flood Damage Assessment: A Review of Microscale Methodologies for Residential Buildings. *Sustainability*, 14(21), Article 21. <https://doi.org/10.3390/su142113817>
- Charalambos, G., Dimitrios, V., & Symeon, C. (2014). *Damage Assessment, Cost Estimating, and Scheduling for Post-Earthquake Building Rehabilitation Using BIM*. 398–405. <https://doi.org/10.1061/9780784413616.050>
- Chau, K. T., & Lo, K. H. (2004). Hazard assessment of debris flows for Leung King Estate of Hong Kong by incorporating GIS with numerical simulations. *Natural Hazards and Earth System Sciences*, 4(1), 103–116. <https://doi.org/10.5194/nhess-4-103-2004>
- Ciurean, R. L., Hussin, H., van Westen, C. J., Jaboyedoff, M., Nicolet, P., Chen, L., Frigerio, S., & Glade, T. (2017). Multi-scale debris flow vulnerability assessment and direct loss estimation of buildings in the Eastern Italian Alps. *Natural Hazards*, 85(2), 929–957. <https://doi.org/10.1007/s11069-016-2612-6>
- de Moel, H., Jongman, B., Kreibich, H., Merz, B., Penning-Rowsell, E., & Ward, P. J. (2015). Flood risk assessments at different spatial scales. *Mitigation and Adaptation Strategies for Global Change*, 20(6), 865–890. <https://doi.org/10.1007/s11027-015-9654-z>
- Feng, S.-J., Gao, H.-Y., Gao, L., Zhang, L. M., & Chen, H.-X. (2019). Numerical modeling of interactions between a flow slide and buildings considering the destruction process. *Landslides*, 16(10), 1903–1919. <https://doi.org/10.1007/s10346-019-01220-9>
- Giuffr , A. (1993). Sicurezza e conservazione dei centri storici. *Il Caso Ortigia*, 279.
- Hu, K. H., Cui, P., & Zhang, J. Q. (2012). Characteristics of damage to buildings by debris flows on 7 August 2010 in Zhouqu, Western China. *Natural Hazards and Earth System Sciences*, 12(7), 2209–2217. <https://doi.org/10.5194/nhess-12-2209-2012>
- Huan, L. I. U., Xiaoyi, F. a. N., Shujun, T., & Xin, D. (2024). Dynamic response of buildings under debris flow impact. *Journal of Mountain Science*, 21(5), 1581–1597. <https://doi.org/10.1007/s11629-023-8572-x>
- Jeong, S.-H., & Elnashai, A. S. (2007). Probabilistic fragility analysis parameterized by fundamental response quantities. *Engineering Structures*, 29(6), 1238–1251. <https://doi.org/10.1016/j.engstruct.2006.06.026>
- Kang, H., & Kim, Y. (2016). The physical vulnerability of different types of building structure to debris flow events. *Natural Hazards*, 80(3), 1475–1493. <https://doi.org/10.1007/s11069-015-2032-z>
- Kr zan, M., Gosti , S., Cattari, S., & Bosiljkov, V. (2015). Acquiring reference parameters of masonry for the structural performance analysis of historical buildings. *Bulletin of Earthquake Engineering*, 13(1), 203–236. <https://doi.org/10.1007/s10518-014-9686-x>
- Kwon, O.-S., & Elnashai, A. (2006). The effect of material and ground motion uncertainty on the seismic vulnerability curves of RC structure. *Engineering Structures*, 28(2), 289–303. <https://doi.org/10.1016/j.engstruct.2005.07.010>
- Li, N., Li, Q., Liu, Y.-S., Lu, W., & Wang, W. (2020). BIMSeek++: Retrieving BIM components using similarity measurement of attributes. *Computers in Industry*, 116, 103186. <https://doi.org/10.1016/j.compind.2020.103186>
- Li, P., Li, T., Lu, Z., & Li, J. (2017). Study on Dynamic Response of Novel Masonry Structures Impacted by Debris Flow. *Sustainability*, 9(7), Article 7. <https://doi.org/10.3390/su9071122>
- Liu, W., Yan, S., & He, S. (2018). Landslide damage incurred to buildings: A case study of Shenzhen landslide. *Engineering Geology*, 247, 69–83. <https://doi.org/10.1016/j.enggeo.2018.10.025>
- Luo, H. Y., Fan, R. L., Wang, H. J., & Zhang, L. M. (2020). Physics of building vulnerability to debris flows, floods and earth flows. *Engineering Geology*, 271, 105611. <https://doi.org/10.1016/j.enggeo.2020.105611>
- Luo, H. Y., Zhang, L. L., & Zhang, L. M. (2019). Progressive failure of buildings under landslide impact. *Landslides*, 16(7), 1327–1340. <https://doi.org/10.1007/s10346-019-01164-0>
- Mej a-Navarro, M., Wohl, E. E., & Oaks, S. D. (1994). Geological hazards, vulnerability, and risk assessment using GIS: Model for Glenwood Springs, Colorado. In M. Morisawa (Ed.), *Geomorphology and Natural Hazards* (pp. 331–354). Elsevier. <https://doi.org/10.1016/B978-0-444-82012-9.50025-6>
- Milanesi, L., Pilotti, M., Belleri, A., Marini, A., & Fuchs, S. (2018). Vulnerability to Flash Floods: A Simplified Structural Model for Masonry Buildings. *Water Resources Research*, 54(10), 7177–7197. <https://doi.org/10.1029/2018WR022577>
- Nelson, H. G. H. (1969). The Chinese Descent System and the Occupancy Level of Village Houses. *Journal of the Hong Kong Branch of the Royal Asiatic Society*, 9, 113–123.
- New Territories Exempted Houses (NTEH)—Buildings Department. (2024, October 2). https://www.bd.gov.hk/en/safety-inspection/ubw/UBW-in-new-territories-exempted-houses/index_ubw_nteh_intro.html
- Nofal, O. M., van de Lind, J. W., & Zakzouk, A. (2022). BIM-GIS integration approach for high-fidelity wind hazard modeling at the community-level. *Frontiers in Built Environment*, 8.

- <https://doi.org/10.3389/fbuil.2022.915209>
Planning Department—Land Utilization in Hong Kong. (2024).
https://www.pland.gov.hk/pland_tc/info_serv/open_data/landu/
- Prieto, J. A., Journeay, M., Acevedo, A. B., Arbelaez, J. D., & Ulmi, M. (2018). Development of structural debris flow fragility curves (debris flow buildings resistance) using momentum flux rate as a hazard parameter. *Engineering Geology*, 239, 144–157. <https://doi.org/10.1016/j.enggeo.2018.03.014>
- Quinay, P. E. B., Soliman, J. M. M., & Fader, A. R. F. (2020). Development of Simulation-Based Approach Using Frame Models Generated From GIS Features and BIM Data for Application to City Seismic Response Analysis of Low- to Mid-Rise RC Structures in Metro Manila. *Journal of Earthquake and Tsunami*, 14(06), 2050021. <https://doi.org/10.1142/S1793431120500219>
- Ren, X., Fan, W., Li, J., & Chen, J. (2019). Building Information Model-based finite element analysis of high-rise building community subjected to extreme earthquakes. *Advances in Structural Engineering*, 22(4), 971–981. <https://doi.org/10.1177/1369433218780484>
- Revit IFC resources / manual. (n.d.). Retrieved October 28, 2024, from <https://www.autodesk.com/support/technical/article/caas/sfdcarticles/sfdcarticles/Title-Revit-IFC-resources-manual.html>
- Tang, C., Rengers, N., van Asch, T. W. J., Yang, Y. H., & Wang, G. F. (2011). Triggering conditions and depositional characteristics of a disastrous debris flow event in Zhouqu city, Gansu Province, northwestern China. *Natural Hazards and Earth System Sciences*, 11(11), 2903–2912. <https://doi.org/10.5194/nhess-11-2903-2011>
- Taucer, F. F., Spacone, E., & Filippou, F. C. (n.d.). *A FIBER BEAM-COLUMN ELEMENT FOR SEISMIC RESPONSE ANALYSIS OF REINFORCED CONCRETE STRUCTURES*.
- Zhang, S., Zhang, L., Li, X., & Xu, Q. (2018). Physical vulnerability models for assessing building damage by debris flows. *Engineering Geology*, 247, 145–158. <https://doi.org/10.1016/j.enggeo.2018.10.017>
- Zhou, S. Y., Gao, L., & Zhang, L. M. (2019). Predicting debris-flow clusters under extreme rainstorms: A case study on Hong Kong Island. *Bulletin of Engineering Geology and the Environment*, 78(8), 5775–5794. <https://doi.org/10.1007/s10064-019-01504-3>

BIM ADOPTION AND IMPLEMENTATION CHALLENGES IN MONGOLIA

Khuvilai Erdene¹, Sung Hoon Kim², and Sang-Ho Lee³

1) Ph.D. Candidate, Department of Civil and Environmental Engineering, Yonsei University, Seoul, South Korea. Email: khuvilai.e@yonsei.ac.kr

2) President, TAESUNG SNI, Seoul, South Korea. Email: shkim@tssni.com

3) Ph.D., Prof., Department of Civil and Environmental Engineering, Yonsei University, Seoul, South Korea. Email: lee@yonsei.ac.kr

Abstract: The primary aim of this research is to identify and address the key barriers to Building Information Modeling (BIM) adoption and implementation within the Mongolian construction industry, serving as a case study for developing economies. BIM offers significant benefits and represents a key technological innovation shaping the future of the Architecture, Engineering, Construction (AEC), and Infrastructure sector towards Digital Transformation. However, there are substantial challenges in BIM adoption and implementation in the construction industry. To identify the barriers to BIM adoption in Mongolia's construction sector, this study conducts a comprehensive analysis of the industry's current state and the status of BIM adoption, supported by a review of relevant literature and case-specific data. The comprehensive analysis utilizes SWOT (Strengths, Weaknesses, Opportunities, and Threats) and PESTLE (Political, Economic, Social, Technological, Legal, and Environmental) analysis approach to identify the key challenges, including technical and policy, socio-economy, human resource related factors. Then, framework to address those challenges is developed based on the analysis results and data. Also, the framework draws insight from the review and comparison study of global BIM implementation best practices. This study contributes to filling the knowledge gap regarding BIM adoption in developing economies and offers practical insights that can aid in developing BIM adoption models for countries with similar economic and industry conditions.

Keywords: BIM, BIM adoption challenges, Mongolian construction industry, SWOT, PESTLE analysis, BIM implementation strategy.

1. INTRODUCTION

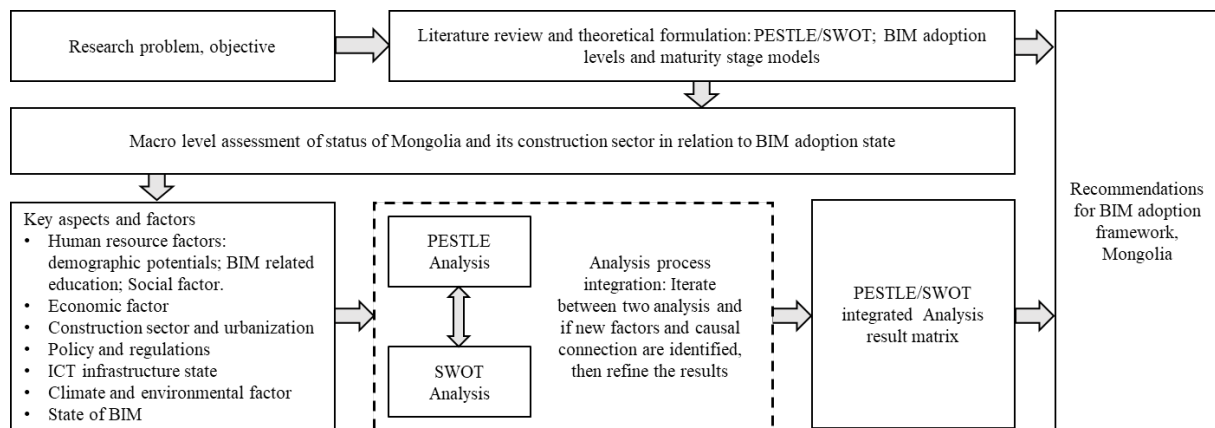
Digital transformation in the construction industry involves adopting digital technologies and processes to modernize and enhance various aspects of construction projects. This transformation integrates data, automation, and digital tools to improve project management, design, construction, and operation. Technologies such as Building Information Modeling (BIM), project management software, IoT devices, and drones streamline processes, reduce errors, and enhance collaboration. Developed countries are increasingly focusing on digital transformation to boost efficiency and productivity, reduce costs, and promote sustainability. By leveraging digital tools, they can better track and manage environmental impacts, enhance safety through real-time monitoring, and remain competitive in the global market. BIM, in particular, is a cornerstone of this transformation, offering significant benefits in terms of efficiency, cost reduction, collaboration, and sustainable practices. The advantages of BIM extend beyond the initial delivery of constructed assets to their ongoing management and operations. Numerous studies have highlighted the benefits of embracing BIM during the initial delivery phase of an asset, providing different estimates of the return on investment (ROI) and overall value derived from its implementation. Despite current limitations in digital transformation, forward-looking studies project significant cost savings and productivity gains through comprehensive digital adoption. For instance, vertical construction could achieve cost savings of 10 to 25%, while infrastructure projects could see even higher reductions (Gerbert et al., 2016). Additionally, McKinsey estimates that the industry has the potential to increase overall productivity by 50 to 60% through design-to-manufacturing processes supported by digital information (Barbosa et al., 2017). Countries with comprehensive BIM strategies, like the UK, have already reported notable decreases in public construction costs, achieving a 15 to 20% reduction through a BIM mandate on all public projects (Cabinet Office, 2011). Furthermore, case studies on open BIM adoption highlight benefits such as cost reductions, improved quality, and shorter project delivery times. Meanwhile, developing economies fall behind in the adoption of BIM, mostly because of the particular difficulties they encounter. For example, a study by Al-Sarafi et al. (2022) reveals that BIM, despite its potential to revolutionize construction practices, faces significant adoption barriers in developing countries. Their work identifies five key obstacles hindering wider BIM implementation: resistance to changing established industry practices, the high cost of implementation, insufficient client demand, inadequate collaboration among stakeholders, and a general lack of awareness regarding BIM's potential benefits. These points are further supported by developing country cases. Fitriani et al. (2019) identified the high initial cost of software and hardware as the most significant barrier against the BIM adoption in Indonesia. Studies on the Malaysian construction industry, such as the one by Enegbuna et al. (2015), have identified a need for greater awareness and understanding regarding BIM adoption. Despite the growing body of knowledge on BIM adoption barriers and benefits, further studies are essential to address the specific contexts of a broader range of developing countries.

As such, Mongolia, a developing country with a rapidly growing construction sector, requires further investigation to understand its unique context regarding BIM adoption. Given the limited availability of data and research on BIM adoption in Mongolia, this study aims to initiate a discussion to advocate the need for data availability, that will create the baseline for accurately assessing BIM's current adoption state. By employing an integrated SWOT and PESTLE analysis approach, this research identifies the core internal and external factors influencing BIM adoption in Mongolia's construction sector. The insights gathered from this analysis will inform a targeted recommendation framework, designed to guide future studies and strategic actions to support and measure BIM adoption more effectively.

2. RESEARCH METHODOLOGY

This study follows a structured approach to assess BIM adoption challenges and opportunities in Mongolia's construction sector, beginning with the comprehensive macro-level assessment of Mongolia's construction sector, examining key factors that could influence BIM adoption. This phase involves surveying critical aspects and evaluating factors that will populate the PESTLE (Political, Economic, Social, Technological, Legal, and Environmental) and SWOT (Strengths, Weaknesses, Opportunities, and Threats) analyses. The assessment is specifically tailored to Mongolia's unique economic, regulatory, and technological landscape to ensure relevance and accuracy in capturing the country's BIM adoption potential. The core of the analysis lies in the integration of PESTLE and SWOT, creating an iterative process where insights from one framework inform and refine the other. This iterative integration involves continuously revisiting each framework; if new factors or causal relationships emerge, the findings are updated to capture a more nuanced and interconnected view of Mongolia's BIM adoption environment. This process results in an integrated PESTLE/SWOT matrix, which synthesizes both macro and micro factors into a cohesive analysis.

Finally, using insights from the PESTLE/SWOT analysis and guided by BIM adoption and maturity stage models, the study proposes a recommendation framework tailored to Mongolia. This framework addresses specific stages and levels necessary for advancing BIM adoption, providing actionable steps aligned with the maturity needs and sectoral context of Mongolia's construction industry.



Integrated PESTLE/SWOT matrix result

PESTLE/SWOT	Internal/External	
	Strengths (Positive)	Weaknesses (Negative)
Policy	P.S.	P.W.
Economic	Ec.S.	Ec.W.
Social	S.S.	S.W.
Technological	T.S.	T.W.
Legislative	L.S.	L.W.
Environmental	En.S.	En.W.
PESTLE/SWOT	Internal/External	
	Opportunities (Positive)	Threats (Negative)
Policy	P.O.	P.T.
Economic	Ec.O.	Ec.T.
Social	S.O.	S.T.
Technological	T.O.	T.T.
Legislative	L.O.	L.T.
Environmental	En.O.	En.T.

Figure 1. Research design based on the integration of the SWOT, PESTLE analysis

SWOT and PESTLE tools was selected for this study because these are proven tools for strategic planning, that provides a structured framework. Despite some criticisms, such as being labeled "atheoretic" by Grant (2008), SWOT remains a widely used and effective strategic tool. Coman and Ronen (2009) highlight its ability to distill strengths and weaknesses into core competences and core problems, which can then be linked into a plan of action. Evans and Wright (2009) also emphasize its usefulness but suggest combining it with other tools like Porter's 5-Forces analysis for a more comprehensive strategic planning process. For our study we combine with PESTLE.

3. MACRO LEVEL ASSESSMENT OF STATUS OF MONGOLIA AND ITS CONSTRUCTION SECTOR IN RELATION TO BIM ADOPTION STATE

3.1 Demographic Profile of Mongolia

Mongolia's population of 3,457,500 presents a unique demographic profile, characterized by a prominent youth bulge (National Statistics Office of Mongolia, 2022). As of 2022, nearly 43.59% of the population falls within the 20-49 age group (see Figure 2). This age group plays a significant role in the country's workforce, consumer market, and overall social dynamics. In many countries younger generations are quick to adapt and learn to utilize new technologies, thus this status is potential benefit for Mongolia when it comes to BIM adoption. Because younger generations tend to be more comfortable with technology and have a greater aptitude for learning new software and digital tools. Also they are more likely to be aware of and interested in emerging technologies and trends shaping the future of the construction industry.

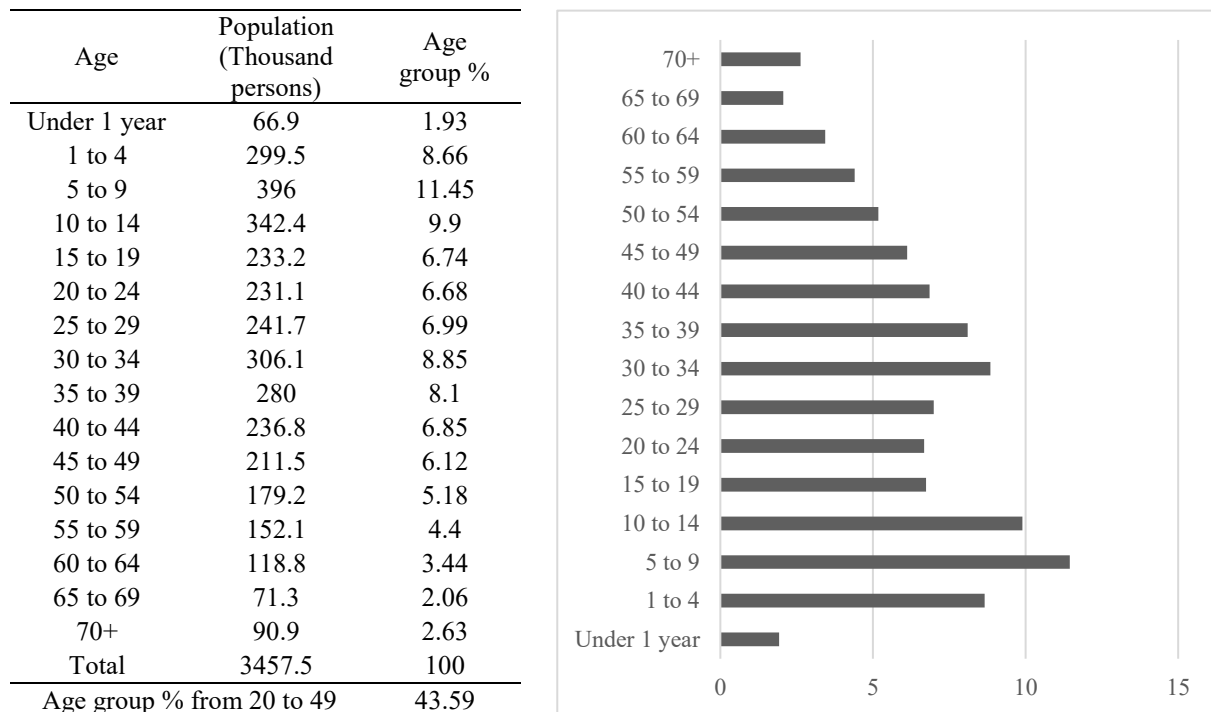


Figure 2. Population of Mongolia by age group (2022)

3.2 Economic Profile of Mongolia

In 2022, Mongolia's economy experienced significant challenges, largely due to the global impacts of the COVID-19 pandemic and other global events that occurred in recent years. As the year progressed, conditions started to improve, leading to an increase in mineral exports. As shown in the Table 3, the growth rate for 2022 was 4.8%, a substantial improvement from 1.6% in 2021 and a contraction of 4.6% in 2020 (World Bank, 2023; National Statistics Office of Mongolia, 2022). Despite the ongoing recovery, Mongolia's economy remains susceptible to external impacts. However, it is gradually transitioning towards economic growth in the post-pandemic era. BIM adoption in Mongolia's construction sector can significantly enhance cost effectiveness, which is crucial given the sector's substantial role in the national economy and its impact on controlling foreign debt and inflation. By providing precise cost estimates and streamlining budget management throughout a building's lifecycle, BIM minimizes financial risks associated with construction cost overruns, often a source of escalating debt. Furthermore, BIM's accurate modeling enables efficient resource allocation and reduces material waste, ensuring that projects stay within budget and timelines, which collectively fosters economic stability and mitigates inflationary pressures.

Table 3. Primary Economic Evaluation of Mongolia 2012-2022

Indicators	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Growth Rate (%)	12.3	11.6	7.9	2.4	1.5	5.6	7.7	5.6	-4.6	1.6	4.8
Inflation Rate (%)											
Consumer prices	14	10.5	12.3	5.7	0.7	4.3	6.8	7.3	3.8	7.4	15.1
GDP (Billion. \$)	12.2	12.5	12.2	11.7	11.1	11.4	13.18	14.21	13.31	15.29	16.81
Gross Foreign Debts (Billion. \$)	13.65	15.05	16.42	22.71	24.62	27.41	28.7	30.7	32.2	33.2	33.6

3.3 Climate Profile of Mongolia

Mongolia is a country with a high altitude, cold temperatures, and dry climate. It has an extreme continental climate with long, cold winters and short summers, during which most precipitation falls (Climate Change Knowledge Portal, 2024). The country experiences 257 sunny days a year, and it is typically located at the center of a region of high atmospheric pressure. Due to its short warm season and longer cold seasons, about 6 months of the year is suitable for the construction work (see Figure 3) activities without additional heating cost. Number of ways BIM adoption can improve Mongolia's construction industry in relation to the harsh climate:

1) BIM's virtual environment and precise simulations allow architects and engineers to test designs for extreme weather conditions like blizzards and scorching heat, optimizing for thermal performance and structural integrity before breaking ground. This can save time and money during construction, as well as prevent costly problems down the line.

2) By facilitating prefabrication and modular construction techniques, BIM streamlines the building process in areas with short construction seasons due to harsh winters. This reduces exposure to the elements and ensures projects are completed quickly and efficiently.

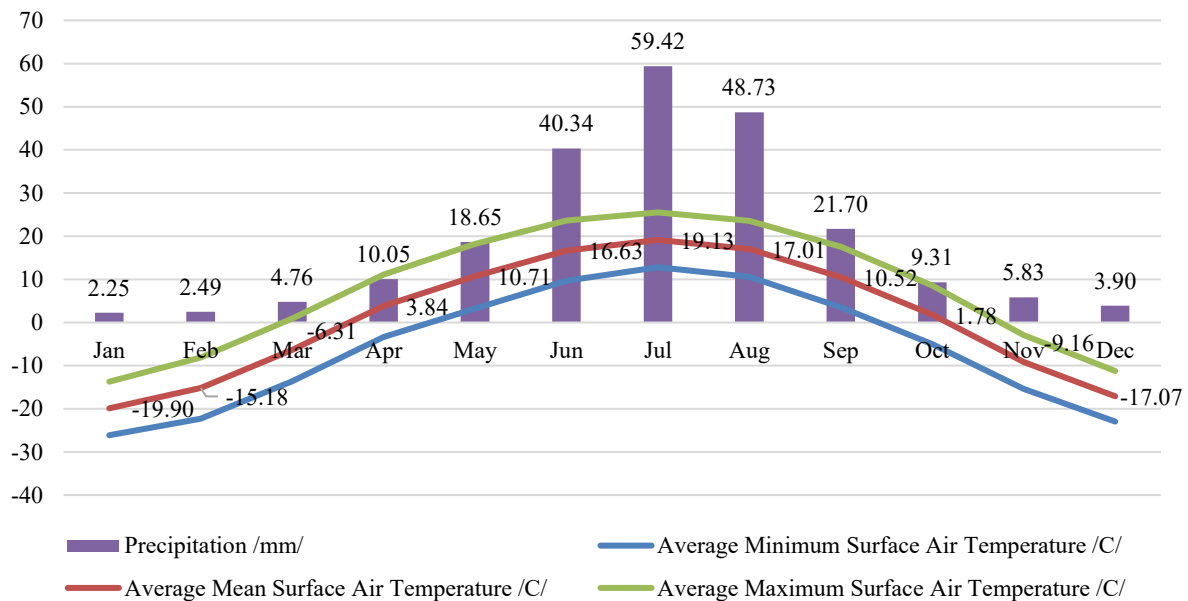


Figure 3. Monthly Average Min, Max, Mean Surface Air Temperature and Precipitation 1991-2020, Mongolia

3.4 Mongolia's Construction Sector Market

National Statistics Office of Mongolia sizes the Construction sector market in regards of Construction, Capital Repairs and Maintenances budgets of the sector. Economic operations aimed at creating, renovating, repairing, or extending fixed assets in the form of buildings, engineering-style land improvements, and other similar engineering works like roads, bridges, dams, and so forth are referred to as construction and capital repairs. The trend of new construction is still taking high percentage in the Construction Industry Market of Mongolia. This trend has been constant for the past 12 years as seen in Figure 4. This is related to the urbanization trend of Mongolia. As previously mentioned, Mongolia's economy grew by 4.8% in 2022 compared to 2021. The industry and construction sector contributed 0.5 percentage points to this growth. While the output of construction, capital repairs and maintenance at price of year 2022 increased by 37.4%. When we see from the table of last 5 years' stats of construction, capital repairs and maintenance of Mongolia, there can be seen a steady rise in this sector. Even considering the minor dip in the stats between 2020 to 2021, when the global pandemic COVID19 affected global economy, the construction, capital repairs and maintenance sector of Mongolia still maintained its size.

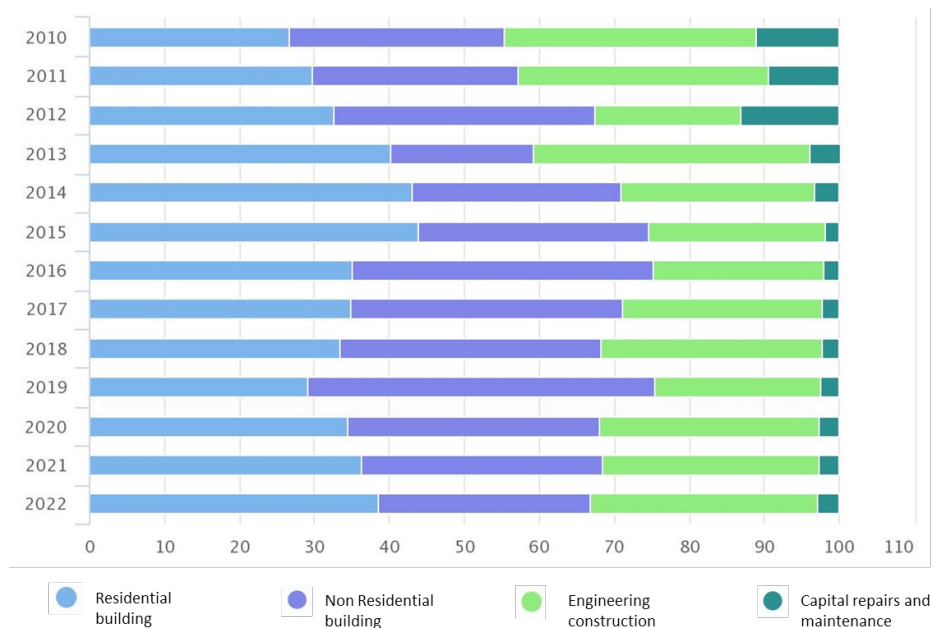


Figure 4. Construction, Capital repairs and maintenance, by type of construction and year in percentage

3.5 Mongolia's Construction Sector Regulations

To access the regulatory status of construction sector of Mongolia, 12 legislations that are relevant to the sector was surveyed. And 27 construction work related permits was also surveyed for this research. In the result, current regulatory environment is handling 2D CAD process but not ready to deal with information model in the scope of all aspects of construction related works. There are 27 types of construction work-related permits regulated by the Urban Development Agency of Capital City Administration of Mongolia. These permits require 2D-based design drawings for various construction work permissions, which later act as records in the urban database. The laws listed in above and other relevant policies serve as the basis for these permits. As Mongolia aims to become a digital nation according to the VISION 2050 policy document, BIM adoption will be a key digitization factor in the construction sector and the urban development sector. Thus, efforts to update necessary regulations with digitization-related factors should be expected to be done in the long run. While construction permits will have to focus on including Information Modelling in conjunction with traditional 2D-based CAD design drawings for recording and archiving purposes.

3.6 Status of BIM Utilization, Education and Understanding in Mongolia

According to our survey findings regarding the BIM education in Mongolia, so far two organizations publicly demonstrated activities to develop education curriculum and environment that supports BIM authoring tools training. The major public university that educates the Mongolian AEC industry professionals, the Mongolian University of Science and Technology, have started paying attention to BIM education from 2014. In 2015, cooperation agreement was made for opening the Autodesk training center branch at the school of Civil Engineering and Architecture, Mongolian University of Science and Technology. Then the Autodesk training center was opened in 2016, with focus of training Autodesk Revit in the beginning. By 2022, the training center prepared estimated number of 302 participants, from which 178 were certified in Autodesk Certification, 56 took training for advancing the professional status of Engineering, and 62 participants with Autodesk Training center issued certificate (BIM-Forum Mongolia, 2022).

A private sector Design Firm in Mongolia launched their own BIM authoring tool training center with name "Model Development Training Center". From their public engagement in the scope of BIM-Forum participation information, their education training is focused on utilizing Revit. The efforts are not minor, but the effects of those efforts are yet to be seen in the future. Judging from the training nature, these efforts appear to be focusing only on training the participants in certain BIM authoring tools. Hence, there is still a need for a training center that focuses on preparing BIM managers and an education curriculum at universities related to the introduction of BIM and advanced-level BIM understanding.

Since 2016, Mongolian Ministry of Construction and Urban Development has been pursuing on preparation for BIM adoption for Mongolia. While private sector is increasing their interest in understanding BIM in order to maintain their competitiveness on the market. But the understanding and utilization still remains at the level of BIM authoring tool utilization (BIM-Forum Mongolia, 2022-2023).

3.7 ICT Infrastructure Status of Mongolia

According to the Global Competitiveness Index 4.0, Mongolia ranks 14th out of 21 countries in terms of ICT adoption in Asia, with a score of 96.38 out of 100. In terms of companies embracing disruptive ideas, Mongolia ranks 17th out of 21 countries in the region of Asia, with a score of 41.43 out of 100. Other ICT infrastructure related statistics are shown in the Table 4.

Table 4. Mongolia's Information and Communications Technology (ICT) sector's Infrastructure and Capacity details and scoring

Name	Mongolia	World Rank	Country Ranked 1st in the World		Data year
Mobile Connectivity Index: Infrastructure: Network coverage: 2G coverage, score (0-100, higher is better)	100	1st/168	100	Mongolia	2022
Mobile Connectivity Index: Infrastructure: Network performance: Mobile download speeds, score (0-100, higher is better)	11.47	126th/168	100	Australia	2022
Global Competitiveness Index 4.0: Electricity infrastructure, score (0-100, higher is better)	91.49	93rd/139	100	Bahrain	2019
Mobile Connectivity Index: Infrastructure: Network coverage: 4G coverage, score (0-100, higher is better)	85.46	118th/168	100	Bahrain	2022
Mobile Connectivity Index: Infrastructure: Network performance: Mobile upload speeds, score (0-100, higher is better)	27.38	116th/168	100	China	2022
Mobile Connectivity Index: Infrastructure: Network performance, score (0-100, higher is better)	39.02	130th/168	97.65	Qatar	2022
Mobile Connectivity Index: Infrastructure, score (0-100, higher is better)	63.12	77th/168	97.94	Denmark	2022
Mobile Connectivity Index: Infrastructure: Network coverage: 3G coverage, score (0-100, higher is better)	95	103rd/168	100	Bahrain	2022
Secure Internet servers (per 1 million people)	1.7k	63rd/184	280k	Denmark	2020

4. INTEGRATED PESTLE AND SWOT ANALYSIS FOR MONGOLIAN CONSTRUCTION INDUSTRY IN RELATION TO BIM ADOPTION

This chapter presents a detailed analysis of the factors affecting BIM adoption in Mongolia's construction sector. By iteratively applying PESTLE and SWOT frameworks, we uncover the complex interplay between external and internal factors. This integrated approach enables us to identify key challenges and opportunities, as well as potential causal relationships between different factors. The resulting PESTLE/SWOT matrix provides a solid foundation for developing actionable recommendations (see Table 5).

Table 5. Integrated PESTLE and SWOT Analysis output matrix

PESTLE/SWOT	Internal factors	
	Strengths (Positive)	Weaknesses (Negative)
Political	Government support for technological advancement	Limited government support for BIM-specific initiatives Lack of clear and standardized policies for BIM adoption.
Economic	Growing economy Increasing demand for infrastructure. Potential for foreign investment	Financial constraints for smaller firms Limited access to funding for BIM adoption.
Social	Young and tech-savvy population Growing urbanization Positive public perception about BIM	Lack of awareness about BIM benefits Resistance to change Cultural barriers
Technological	Potential for technological advancement	Limited ICT infrastructure Skills gap: Limited educational programs and training opportunities specific to BIM. Lack of technological expertise

Table 5. Integrated PESTLE and SWOT Analysis output matrix (continued)

PESTLE/SWOT	Internal factors	
	Strengths (Positive)	Weaknesses (Negative)
Legal		Lack of clear regulations and standards for BIM Potential data security concerns Existing regulatory environment not ready for regulating new disruptive technological changes
Environmental	Potential for sustainable construction practices	Lack of clear environmental regulations specific to BIM
SWOT/PESTLE	External factors	
	Opportunities (Positive)	Threats (Negative)
Political	Government interest for technological advancement Potential for public-private partnerships Growing economy is driving opportunities for BIM adoption by increasing the demand for construction projects and attracting foreign investment.	Political instability
Economic	Infrastructure development Foreign investment: Foreign investment in Mongolia can bring in expertise, technology, and funding for BIM adoption. The increasing demand for construction projects in urban areas presents opportunities for BIM adoption to address the challenges of urban development.	Economic downturns can reduce construction activity and limit investment in new technologies like BIM. Fluctuations in market demand Limited International Collaboration
Social	Changing lifestyles: Changes in lifestyles and preferences can drive the demand for efficient and sustainable construction practices, which BIM can support.	Cultural barriers: Cultural attitudes or beliefs within Mongolian construction sector that influences the acceptance or rejection of BIM. Resistance to change: Resistance to adopting new technologies due to ingrained traditional practices, a lack of awareness about BIM's benefits, or a reluctance to change established workflows.
Technological	Technological advancements: Advancements in BIM technology and its integration with other technologies can enhance its capabilities and attractiveness for adoption. Through integration with other technologies, promotion of other new technologies	Inadequate internet connectivity, lack of access to advanced technology, and insufficient IT resources. Cybersecurity risks Technological obsolescence: BIM technology may become outdated over time, requiring regular updates and upgrades to ensure its effectiveness.
Legal	Potential for favorable regulations Clear contractual considerations	Changes in regulations Unclear liability issues
Environmental	Demand for sustainable construction practices. Potential for green building certifications	Environmental regulations: Strict environmental regulations might pose challenges for BIM adoption, but they can also drive innovation in sustainable construction practices. Natural disasters, climate change

5. BIM IMPLEMENTATION ROADMAP STRATEGIC FRAMEWORK PROPOSAL

Given the government's ability to influence and regulate technology diffusion through strategies such as knowledge building, knowledge deployment, subsidies, mobilization, standard setting, and innovation directives (King et al., 1994), it is evident that public authorities will play a leading role in the adoption of Building Information Modeling (BIM) strategies. Thus, our proposed framework emphasizes the government's leading role in the adoption and diffusion of Building Information Modeling (BIM) technologies. And the proposed framework as the recommendations are derived from a careful analysis of the interconnected factors revealed by the PESTLE/SWOT matrix. By understanding how these factors influence each other, we can identify specific actions to address the key challenges and opportunities. For instance, if the matrix highlights the need for stronger legal frameworks to support BIM's regulatory consistency, this will be a key component of the proposed framework.

5.1 Maturity Stage Model Based Recommendation for Mongolia

Figure 5 provides a visual representation of the timeline for the targeted implementation of Building Information Modeling (BIM) stages within the construction sector. It demonstrates the phased approach followed by many developed countries when mandating BIM adoption. These stages often align with the legal system's progression, ensuring a systematic and structured implementation process.

The first stage involves conducting pilot projects to test and assess the feasibility of BIM in real-world scenarios. This will allow for the identification of potential challenges and the development of best practices and guidelines. Following the pilot phase, guidelines are established to provide a framework for BIM implementation, addressing key aspects such as data standards, interoperability, and collaboration protocols. Once the guidelines are in place, the mandatory implementation phase begins. The public building sector is often at the forefront of this implementation, adopting BIM practices approximately five years earlier than the infrastructure facility sector, which includes projects related to roads, railways, and other public infrastructure.

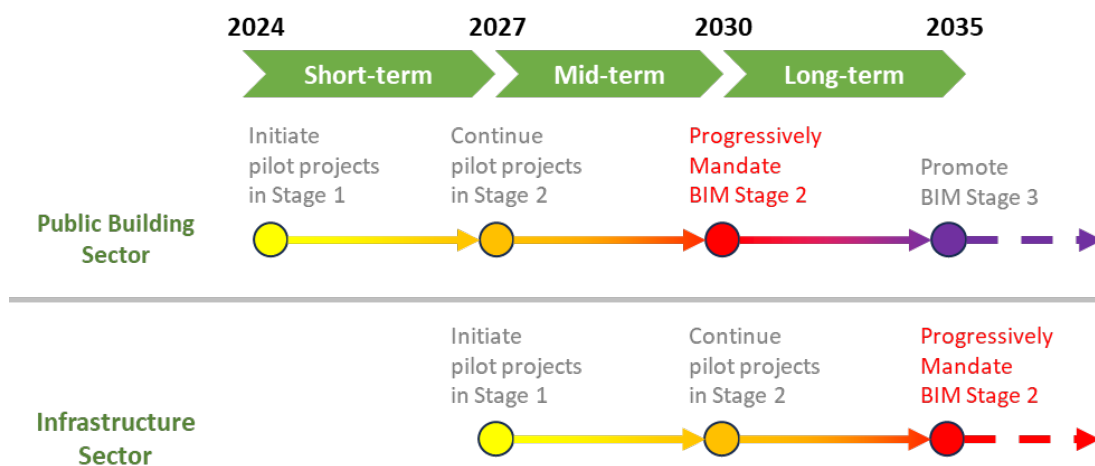


Figure 5. Target maturity stage by phase and construction sector

5.2 Recommended Strategic Goals for Mongolia by Phase and Action Framework

The strategic goals to achieve for each of the maturity stage described in Section 5.1 can be set as shown in Figure 6 and brief of each strategic goals are as followings:

Short-term goal: Establish fundamentals and begin 3D coordination in limited project domains by 2027. The goal of the short-term goal is to initiate the implementation of BIM in the construction industry, beginning with a simple use case - 3D model-based coordination. This serves as an entry point for the industry to understand the benefits of BIM and begin integrating it into their current workflows. To promote BIM adoption during this phase, it is essential to establish the necessary fundamentals, such as developing standardized guidelines and legal documents. In addition to creating guidelines and legal documents, it is also important to improve the legal system to ensure the availability of qualified BIM engineers. The legal system plays a critical role in enforcing regulations and guidelines, and promoting the adoption of BIM requires the support of a qualified workforce.

Mid-term goal: Mongolia should aim to be ready for implementation of BIM Stage 2 by 2030

The objective of the mid-term goal is to elevate the construction industry in Mongolia to an international level by achieving Stage 2 maturity in BIM implementation. This stage focuses on expanding the use and capabilities of BIM beyond simple 3D model-based coordination, laying the foundation for more advanced BIM practices. To meet this goal, it is crucial that government authorities responsible for construction projects are prepared to use BIM even before it becomes mandatory. This proactive approach ensures that the industry is ready

to embrace advanced BIM practices as mandated by international standards. During this phase, technical infrastructures such as Common Data Environment (CDE) operation (ISO 19650), internet-based communication, hardware, and software will be widely utilized. These technological tools facilitate information sharing, coordination, and collaboration among various stakeholders in the construction industry.

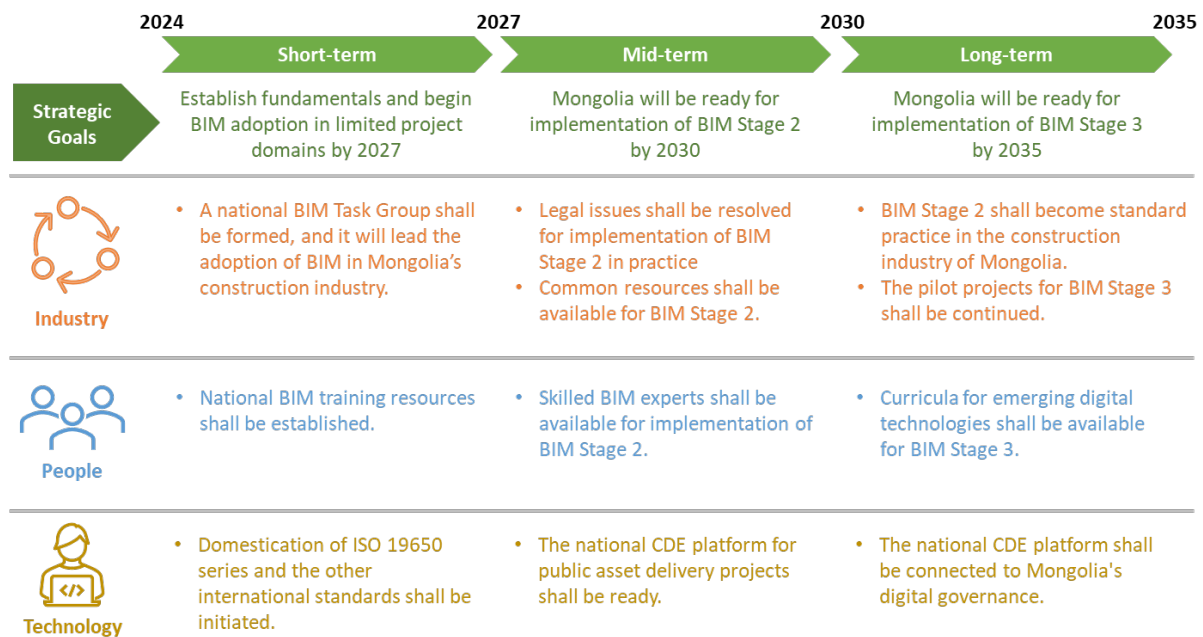


Figure 6. Strategic goals by phase and action framework

Close collaboration with relevant government departments is crucial to the success of this phase. This collaboration ensures that government policies and initiatives support and align with the adoption of BIM practices. It also helps in streamlining processes and removing any barriers or challenges that may hinder the successful implementation of BIM in the construction industry.

Long-term goal: Mongolia should be ready for implementation of BIM Stage 3 by 2035

The objective of the long-term goal is to finalize the digital transformation of the construction industry in Mongolia, marking a significant milestone in the adoption and integration of advanced technologies. This phase is in alignment with the strategies of advanced countries in expanding the applications of model-based construction information to reach the highest level of the BIM maturity model. With BIM technology firmly established in the construction industry, this phase aims to shift the focus towards leveraging model-based information to increase productivity and ensure safety throughout the entire lifecycle of built environments. By harnessing the power of digital information, stakeholders can make informed decisions, streamline processes, and optimize resource utilization from design to construction to facility management. Research and development efforts will play a crucial role during this phase as stakeholders seek to adopt and implement new technologies. Researchers will be at the forefront of driving innovation by exploring emerging technologies, conducting pilot projects, and validating the potential impact of these technologies on the construction industry in Mongolia.

Finally, the framework is separated into three aspects.

- 1) The first aspect is industry, which covers the requirements and plans for promoting BIM implementation in the construction industry of Mongolia.
- 2) The second aspect is people, which deals with policies and plans for upskilling human resources as enabling components for BIM implementation in the industry.
- 3) The third aspect is technology, which describes the technical infrastructures necessary for implementing BIM.

6. CONCLUSIONS

This study presents a structured approach to advancing BIM adoption in Mongolia's construction sector through an integrated SWOT and PESTLE analysis, forming the basis of short-, mid-, and long-term strategic goals aligned with BIM maturity stages.

- The short-term goal aims to establish foundational BIM practices, focusing on 3D coordination by 2027, supported by standardized guidelines, legal frameworks, and a qualified BIM workforce.
- The mid-term goal targets BIM Stage 2 maturity by 2030, with expanded applications and digital

infrastructure, like Common Data Environments, enabling more advanced BIM practices. This phase emphasizes collaboration with government to ensure alignment of policies and infrastructure with industry needs.

- The long-term goal envisions full digital transformation and readiness for BIM Stage 3 by 2035, integrating BIM across the lifecycle of construction projects to maximize productivity and safety. Research and development will be essential in this phase, promoting ongoing innovation and alignment with global standards.

The recommended framework addresses BIM adoption across industry, people, and technology, offering a balanced pathway for Mongolia's construction sector to progress through each maturity stage, align with international standards, and establish a foundation for sustainable digital transformation. While this study provides a comprehensive macro-level analysis of BIM adoption in Mongolia using SWOT and PESTLE frameworks, the limited availability of specific data on BIM implementation about Mongolia suggests that future research should focus on micro-level studies to validate these findings and offer more actionable insights.

ACKNOWLEDGMENTS

This work is supported by the Korea Agency for Infrastructure Technology Advancement(KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant RS-2021-KA163381).

REFERENCES

- Al-Sarafi, A. H., Alias, A. H., Jakarni, F. M., Shafri, H. Z. M., and Gamil, Y. (2022). Building Information Modelling: Challenges, Benefits, and Prospects for Adoption in Developing Countries, *International Conference on Information Systems and Intelligent Applications*, Cham: Springer International Publishing, pp. 551–566.
- Barbosa, F., Woetzel, J., Mischke, J., Ribeirinho, M. J., Sridhar, M., Parsons, M., Bertram, N., and Brown, S. (2017). *Reinventing construction: A route to higher productivity*. McKinsey and Company.
- BIM-Forum Mongolia (2022). *Presentation Proceedings*, Ulaanbaatar, Mongolia
- BIM-Forum Mongolia (2023). *Presentation Proceedings*, Ulaanbaatar, Mongolia
- Cabinet Office (2011). *Government Construction Strategy*, Cabinet Office.
- Coman, A. and Ronen, B. (2009). Focused SWOT: diagnosing critical strengths and weaknesses, *International Journal of Production Research*, 47 (20), 5677-5689.
- Climate Change Knowledge Portal. (2024). *Mongolia: Climatology*. Retrieved from website: <https://climateknowledgeportal.worldbank.org/country/mongolia/climate-data-historical#:~:text=Average%20temperatures%20range%20between%20around,varies%20dramatically%20throughout%20the%20year>.
- Enegbuma, W. I., Aliagha, G. U., and Ali, K. N. (2015). Effects of perceptions on BIM adoption in Malaysian construction industry, *Jurnal Teknologi*, 77 (15).
- Evans, C. and Wright, A. (2009). How to conduct a SWOT analysis, *The British Journal of Administrative Management*, 24 (65), 10-34.
- Fitriani, H., Budiarto, A., Saheed, A. and Idris, Y. (2019). Implementing BIM in architecture, engineering and construction companies: Perceived benefits and barriers among local contractors in Palembang, Indonesia. *International Journal of Construction Supply Chain Management*, 9 (1), 20-34.
- Gerbert, P., Castagnino, S., Rothballer, C., Renz, A., and Filitz, R. (2016), *Digital in Engineering and Construction: The Transformative Power of Building Information*. The Boston Consulting Group.
- Government of Mongolia. (2016). *Vision-2050 Introduction to Mongolia's Long-term Development Policy Document*.
- Grant, R.M. (2008). Why strategy teaching should be theory based, *Journal of Management Inquiry*, 17 (4), 276-91.
- King, J. L., Gurbaxani, V., Kraemer, K. L., McFarlan, F. W., Raman, K. S., and Yap, C. S. (1994). Institutional Factors in Information Technology Innovation, *Information Systems Research*, 5 (2), 139-169.
- National Statistics Office of Mongolia. (2018; 2019; 2020; 2021; 2022), *Mongolian Statistical Yearbook*, National Statistics office of Mongolia, Ulaanbaatar.
- UK BIM Framework. (2022). "ISO 19650 Guidance 1: Concepts," UK BIM Framework. Retrieved from website: <https://www.ukbimframework.org/>.
- USAID Data Services. *Country Dashboard: Mongolia, ICT*, IDEA. Retrieved from website: <https://idea.usaid.gov/cd/mongolia/information-and-communications-technology-ict>.
- World Bank. (2023). *Country Dashboard: "Mongolia"*, World Bank. Retrieved from website: <https://data.worldbank.org/country/mongolia?view=chart>

Intelligent Geometric Quality Inspection for Multiple Prefabricated Components in Large-Scale Storage Yard Based on BIM-LiDAR-UAV/UGV

Limei Chen¹, Zhigang Guo², Yi Tan³ and Penglu Chen^{4*}

1) Master, Research Assistant, Sino-Australia Joint Research Center in BIM and Smart Construction, Shenzhen University, Shenzhen, China. Email: chenlimei2021@email.szu.edu.cn

2) Ph.D., Engineer, Shenzhen Municipal Group Co.,Ltd, Shenzhen, China. Email: huily1984@163.com

3) Ph.D., Assoc. Prof., Sino-Australia Joint Research Center in BIM and Smart Construction; Shenzhen University, Shenzhen, China. Email: tanyi@szu.edu.cn

4) Master, Research Assistant, Sino-Australia Joint Research Center in BIM and Smart Construction, Shenzhen University, Shenzhen, China. Email: anchenyin@outlook.com

Abstract: Quality inspection in outdoor prefabricated storage yards is highly challenging due to the large volume, diversity, and complex management demands of components. Currently, these inspections are conducted manually, which is insufficient to fully meet industry needs. This study proposes an innovative approach to enable intelligent inspection for multiple prefabricated components in large-scale prefabricated storage yards by integrating Building Information Modeling (BIM), LiDAR Unmanned Aerial Vehicles (UAVs), and Unmanned Ground Vehicles (UGVs). First, an intelligent sensing environment and stepwise collaboration mechanism are established, where UAVs are used to reconstruct a 3D comprehensive environment of the prefabrication site, providing a map to plan the optimal scanning path for LiDAR-equipped UGVs. Next, a point cloud-driven integrated geometric quality inspection method is introduced, where UGVs autonomously collect, and process point cloud data to extract precise geometric features of large components within expansive spaces. To verify the effectiveness of the proposed method, an experiment at a large-prefabricated component factory that produces a variety of types of prefabricated components is conducted. By integrating point cloud processing results with BIM model design information, this research achieves high-precision, large-scale quality inspections of the non-structural performance of prefabricated components, significantly enhancing inspection efficiency and accuracy.

Keywords: Large-scale prefabricated storage yard, geometric quality inspection, prefabricated components, UAV/UGV integration, LiDAR.

1. INTRODUCTION

During the manufacturing process of prefabricated components, quality issues often arise, such as dimensional deviations and surface defects, which can significantly impact the quality of engineering projects and subsequent assembly. Currently, quality inspection in outdoor prefabricated storage yards faces challenges due to the large quantity, diversity, and management complexity of components. Traditional manual inspection methods are limited by time and labor costs (Tan et al., 2020), making them insufficient to meet the needs of large-scale storage yard inspections.

To improve the automation of geometric quality inspection for prefabricated components, many researchers have explored new algorithms and technologies, including computer vision (B. Wang et al., 2022), 3D laser scanning (M.-K. Kim et al., 2015; Qian Wang, Kim, Cheng, et al., 2016; Qian Wang, Kim, Sohn, et al., 2016; Li et al., 2020), and augmented reality (Chi et al., 2022). Also known as reality capture technology, 3D laser scanning offers unique advantages in efficiency and precision. Laser scanning can generate high-density, accurate point cloud data, capturing the geometry of components in fine detail, including subtle features and curved surfaces. Building Information Modeling (BIM) provides a digital model of the construction that includes geometric, spatial, and attribute information of components. Combining LiDAR-based 3D laser scanning with BIM in component inspection enables a more comprehensive, accurate, and real-time approach to detection and management (Tan et al., 2023). By comparing the acquired point cloud data with the design BIM model, geometric deviations in components can be effectively detected (Bosché & Guenet, 2014; Tan et al., 2024).

Although 3D laser scanning-based methods are fully automated, non-contact, and highly precise, current research primarily focuses on geometric quality inspection for single prefabricated components and typically requires manual intervention for point cloud data collection. This has clear limitations and does not meet the needs of diverse component detection in storage yards. With advancements in communication, machine intelligence control, and spatial positioning technologies, Unmanned Aerial Vehicles (UAVs) and Unmanned Ground Vehicles (UGVs) are increasingly used for environmental sensing and data collection in large-scale engineering contexts (Tan et al., 2021; Rachmawati & Kim, 2022). Therefore, combining UAVs with LiDAR-equipped UGVs could help address the current inefficiencies in point cloud data collection.

To address the above issues, this study aims to develop a framework for point cloud data collection and geometric quality inspection for multiple prefabricated components in large-scale prefabricated storage yards by integrating UAV, UGV, LiDAR, and BIM technologies. The innovations of this study include: (1) environmental sensing of the prefabricated storage yard and point cloud data collection method based on combining UAV and

UGV. (2) A target component extraction method based on a density projection algorithm by utilizing the Z-axis projection distribution of the point cloud for stacked components to locate the position of timbers and extract target component. (3) A two-step classification and recognition process for identifying component types and providing reliable data for subsequent quality inspection. Initially, the PointNet++ detection method is used to classify the component, followed by a Scan-vs-BIM comparison to determine the specific component type. The proposed technical framework has significant potential to advance the intelligent management of large-scale prefabricated storage yards and automate quality inspection of prefabricated components.

The organization of this paper is as follows: Section 2 presents background information, covering (1) research on data collection combining UAV and UGV, and (2) point cloud-based geometric quality inspection for prefabricated components. Section 3 details the proposed methodology for environmental sensing, point cloud data collection and geometric quality inspection of multiple prefabricated components in large-scale storage yards based on BIM-LiDAR-UAV/UGV. Section 4 describes the experimental validation of the proposed technique. Finally, Section 5 concludes the study and discusses directions for future research.

2. Literature review

2.1 Data collection based on UAV/UGV

To achieve integrated inspection of large-scale prefabricated components, it is crucial to rapidly and efficiently conduct comprehensive 3D sensing of the inspection environment. Traditional methods for manually reconstructing BIM models of building scenes are fraught with challenges, such as high costs and lengthy timelines, rendering them inadequate for continuously collecting dynamic project information throughout the lifecycle. Photogrammetry using UAV can quickly extract point cloud data to reconstruct 3D models of scenes or stitch images to create high-resolution panoramic views (S. B. H. K. H. Kim, 2017). However, the precision of this data often falls short of the requirements for geometric quality inspection of prefabricated components. Moreover, UGV control systems, as a core component of construction robotics, have attracted considerable attention from researchers both domestically and internationally. Scholars have utilized 3D laser point clouds (NamanPatel et al., 2019; Fei Yan, 2020), computer vision (Park et al., 2019), and the fusion of computer vision with 3D laser point clouds (LiuYisha et al., 2018; Pierzchała et al., 2018) to construct and localize sparse semantic maps, thereby providing route navigation for UGVs operating in complex environments. The integration of BIM technology with UGVs for autonomous path planning and obstacle avoidance facilitates intelligent inspection in challenging settings. Nevertheless, UGVs possess limited global sensing capabilities within their operational environments during inspections, which can diminish their overall efficiency. By considering the respective advantages of UAVs and UGVs, employing UAVs as a complementary tool for UGVs' environmental awareness can effectively address challenges related to trajectory planning (Hernandez et al., 2014; FernandoRoperro et al., 2019), localization and navigation (O.Sivaneri & N.Gross, 2017, 2018), and data collection (Heß et al., 2012; Asadi et al., 2020) in complex environments. Drawing on precedents established by scholars who have utilized heterogeneous collaborative control systems, this research intends to explore a collaborative control approach between UAVs and UGVs. The goal is to achieve comprehensive 3D sensing of the prefabricated component inspection environment, thereby enabling rapid and intelligent data collection for inspections.

2.1 Point cloud-based geometric quality inspection for prefabricated components

A substantial amount of research has proposed methods for geometric quality inspection of prefabricated components based on 3D point clouds. Bosché (Bosché, 2010) proposed a method using 3D CAD models to identify steel structures from laser scanning data and used them for dimension compliance control. Wang et al. (Q. Wang et al., 2017) used color laser scanning data and BIM to estimate the positions of rebars. The team also developed a mirror-aided laser scanning system for the geometric quality inspection of concrete elements (M. K. Kim et al., 2019). However, most of these algorithms primarily focus on the quality inspection of individual prefabricated components and do not adequately address the identification and measurement of multiple component types in complex scenarios. Deep learning methods have gained widespread attention in the Architecture, Engineering, and Construction (AEC) industry. Li et al. (Li et al., 2022) introduced a comprehensive indoor acceptance system that encompasses indoor semantic segmentation, component surface segmentation, as well as flatness and verticality quality assessments. Perez-Perez et al. (Perez-Perez et al., 2021) also presented Scan2BIM-NET to semantically segment building point clouds into structural, architectural, and mechanical subcomponents. Shu et al. (Shu et al., 2023) developed an automated recognition and measurement assessment method for prefabricated concrete components based on a Prefabricated Concrete Component Recognition Network (PCCR-Net) and 3D point clouds, succeeding in segmenting the synthetic point clouds dataset into rebars and concrete. In practical inspections of component storage yards, prefabricated components exhibit a variety of types and inspection criteria. Therefore, further research is needed to achieve classification of various components and identification of different inspection targets.

3. METHOD

This study proposes an integrated data collection and assessment framework using UAV(Unmanned Aerial Vehicle), UGV (Unmanned Ground Vehicle), LiDAR, and BIM (Building information modeling) to automate and efficiently inspect the quality of large-prefabricated components in the yard. Figure 1 illustrates the proposed framework, which consists of three main parts: (1) UAV-based environmental sensing; (2) UGV-based point cloud data collection; and (3) geometric quality inspection.

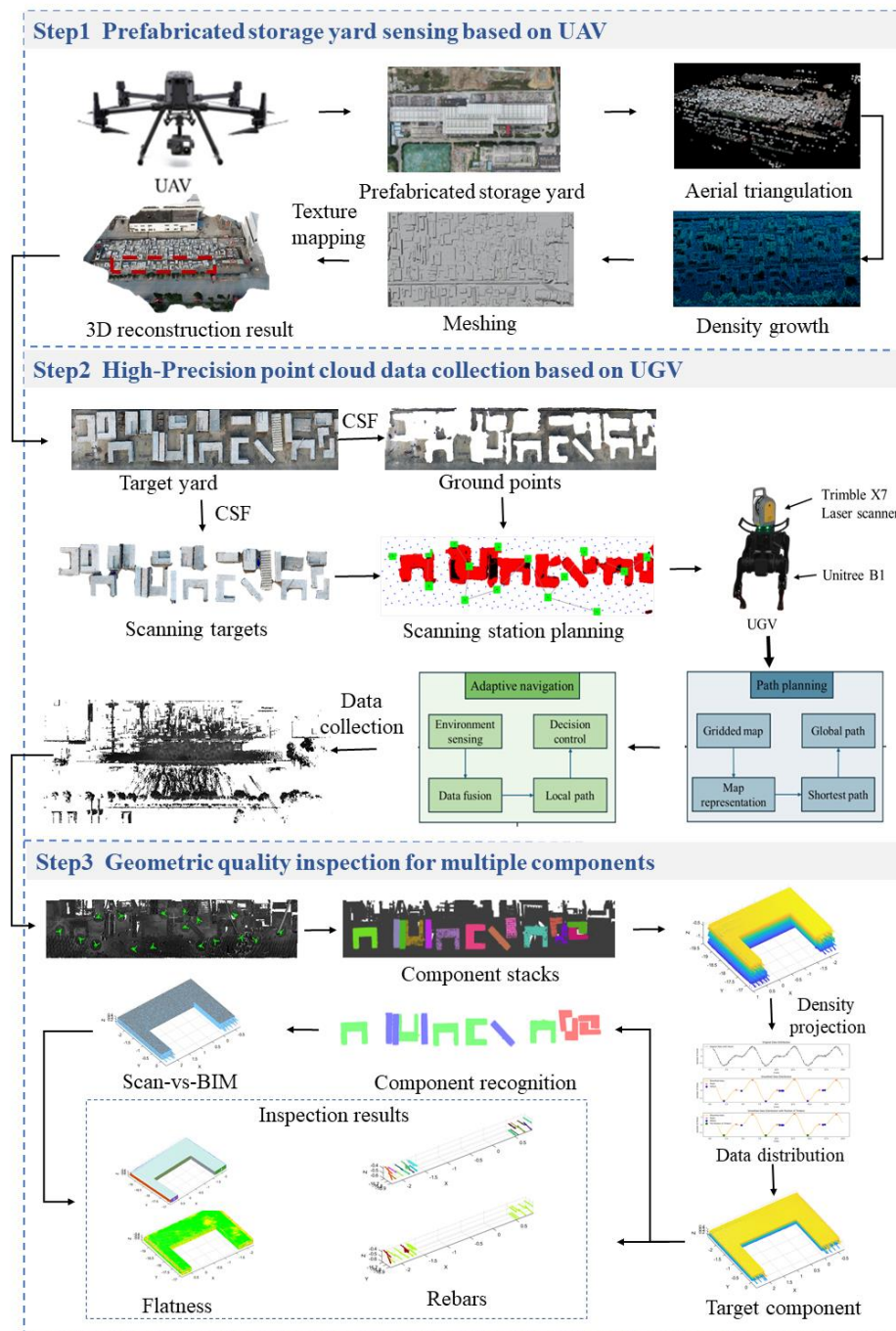


Figure 1. The framework of the proposed method.

3.1 Prefabricated storage yard sensing based on UAV

The UAV-based environmental sensing system consists of two primary components: image data acquisition and 3D reconstruction, as depicted in Figure 2. Initially, the UAV employs an HD camera to capture high-resolution images and utilizes advanced wireless technology to ensure real-time, seamless data transmission. Subsequently, this image data is processed to create a detailed and accurate 3D environmental model that effectively captures the intricacies and characteristics of the terrain. This system enhances the UGV's autonomous capabilities, enabling it to perform tasks like path planning and adaptive navigation with greater precision and efficiency.

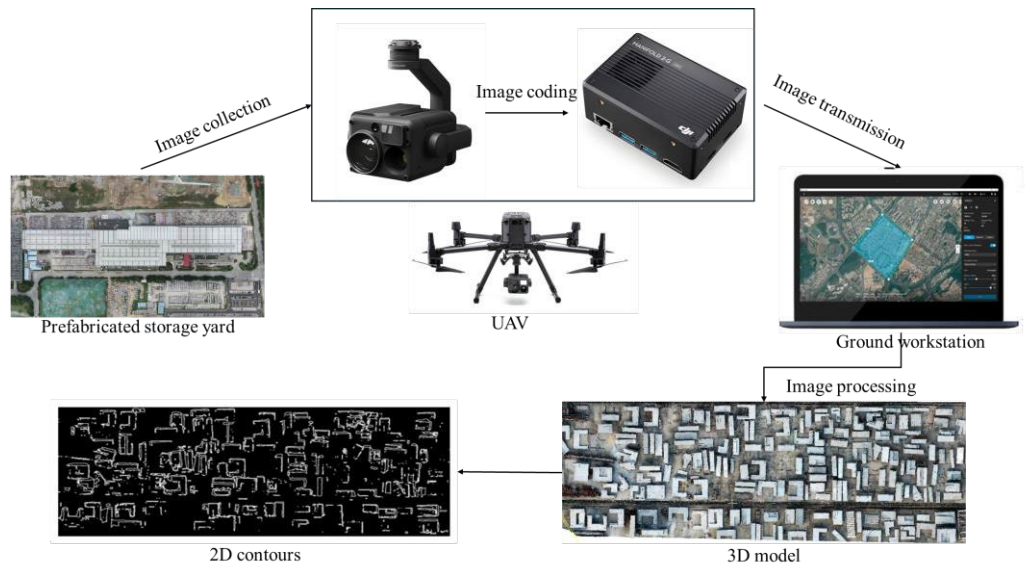


Figure 2. The process of 3D reconstruction of perceived environment.

(1) Image collection and transmission

Before image acquisition, operators only need to set the reconstruction range and image overlap rate, allowing the UAV to automatically plan its flight path and collection strategy based on these parameters. These UAVs are equipped with advanced navigation and positioning systems, enabling precise task execution and ensuring comprehensive coverage and high-quality image capture. With onboard high-definition cameras, UAVs can obtain high-resolution images in critical areas. The cameras feature high pixel counts and wide-angle lenses, adapting to various lighting conditions to capture clear and detailed visuals, providing solid data support for subsequent 3D reconstruction.

To enhance transmission efficiency and reliability, UAVs utilize mainstream video compression technologies such as H.264 and H.265, which effectively compress images to reduce storage and bandwidth requirements while preserving image quality. This optimization improves both data transmission speed and cost-effectiveness. By optimizing data streams through RTSP (Real-Time Streaming Protocol), UAVs achieve fast and stable transmission from the air to ground workstations: RTP ensures real-time packet transmission and minimizes latency, while RTSP provides flexible streaming control mechanisms, enhancing the manageability and reliability of data transfer. The combination of these protocols significantly boosts the data transmission efficiency and system robustness of UAVs, ensuring stable operation in complex environments.

(2) 3D reconstruction

During UAV data collection, changes in lighting and weather conditions significantly impact image quality. Factors like strong sunlight, shadows, clouds, and haze can reduce image clarity, cause color distortion, and increase noise, all of which affect 3D reconstruction quality. Additionally, limitations of imaging equipment and transmission interference can further degrade image quality. To address these issues, this study employs a range of image processing techniques to enhance image quality, as illustrated in Figure 3. (a). First, the BM3D algorithm is applied to remove random noise while preserving details, making images captured in low light or complex weather conditions clearer. Then, the CLAHE algorithm is used to enhance local contrast and emphasize texture details, improving the accuracy of feature extraction in 3D reconstruction. Furthermore, image data processing includes geometric correction and optical calibration to eliminate image distortion and ensure geographical accuracy. This study uses Ground Control Points (GCPs) and resampling techniques to correct geometric distortions and improves image clarity and realism through lens distortion, chromatic aberration, and radiometric calibration, allowing images to more accurately reflect scene details and colors.

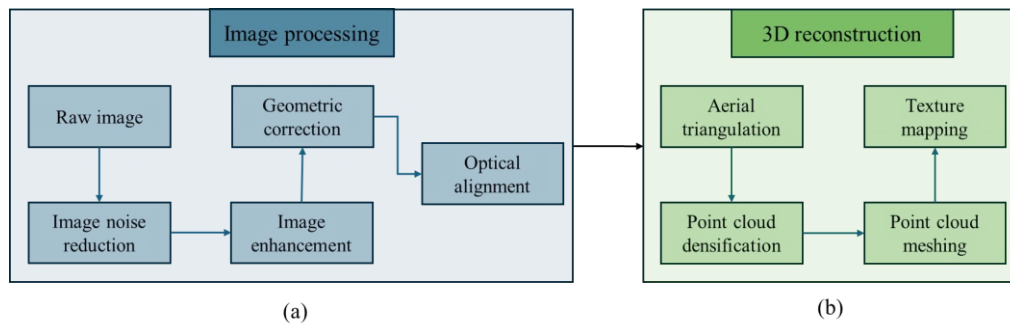


Figure 3. 3D reconstruction process.

3D reconstruction is the process of converting processed 2D image data into accurate 3D models, involving 4 key steps: aerial triangulation, point cloud densification, point cloud meshing, and texture mapping, as illustrated in Figure 3(b). The process begins with aerial triangulation, which utilizes overlapping images to match ground features, allowing for the deduction of geometric relationships that determine the geographic coordinates of objects. This method effectively combines GCPs with automatic feature extraction, employing adjustment algorithms to ensure precise positioning. Next, dense matching generates depth maps using techniques such as semi-global matching and multi-baseline stereo vision. These depth maps are then integrated to create a detailed 3D point cloud model. Following this, point cloud meshing transforms the discrete point cloud into a continuous surface through triangulation. Notably, TIN (Triangulated Irregular Networks) are more efficient than DEM (Digital Elevation Model) for modeling complex terrains, facilitating effective local updates. Finally, the process concludes with texture mapping, where image textures are overlaid onto the mesh surface to enhance realism. By aligning the images with the mesh, color information is accurately mapped, employing techniques like image-by-image mapping and multi-image blending to produce a textured model that showcases realistic colors and details.

3.2 High-Precision point cloud data collection based on UGV

After reconstructing the prefabricated component yard model, the UGV point cloud data collection is optimized for better efficiency and quality. First, optimal scanning station locations are determined to ensure comprehensive coverage and reduce data redundancy. Then, considering the construction site conditions and using these stations, precise path planning and navigation strategies are established for the UGV.

(1) Scanning station planning

Given the complex and obstructed environment of the prefabricated storage yard, a single scanning station is insufficient for capturing all target components effectively. Therefore, multiple scanning stations are necessary to obtain complete point cloud data. The process begins with the use of the CSF (Cloth Simulation Filter) ground filtering algorithm to distinguish between ground and non-ground points, where the ground points help identify suitable locations for scanning stations. Non-ground points are subsequently clustered using the Euclidean filtering algorithm to isolate the scanning targets. These targets are then transformed into voxel grids, and the voxel surfaces are analyzed to assess the visibility between the targets and potential scanning stations. For the scan planning, considerations include coverage, accuracy, detail, and overlap to ensure thorough coverage of all areas. A greedy algorithm is applied to solve the planning problem, iteratively choosing stations that offer the best coverage. With each selection, the additional coverage area is computed to minimize redundancy and optimize station placement and quantity. Lastly, the overlap between neighboring stations is evaluated to meet the requirements for global registration. Should the overlap be inadequate, the algorithm automatically adds new scanning stations to ensure effective global registration.

(2) Path planning and navigation

In Section 3.2.2, the optimal scanning station for UGV data collection was determined and point cloud data can be collected according to the process shown in Figure 4. To ensure the UGV completes the task safely and cost-effectively, path planning is divided into three stages: map representation, path planning, and global path optimization. First, grid-based mapping is used to create an accurate environment map, dividing the 3D reconstructed area into grids marked as passable or impassable based on obstacles. Next, the A* algorithm is employed to find the shortest obstacle-free path from the UGV's starting point to the target scanning station, combining heuristic search with a cost function. Finally, a genetic algorithm is used to

optimize the initial path, generating new path combinations through selection, crossover, and mutation, and evaluating the total cost to find the global optimum.

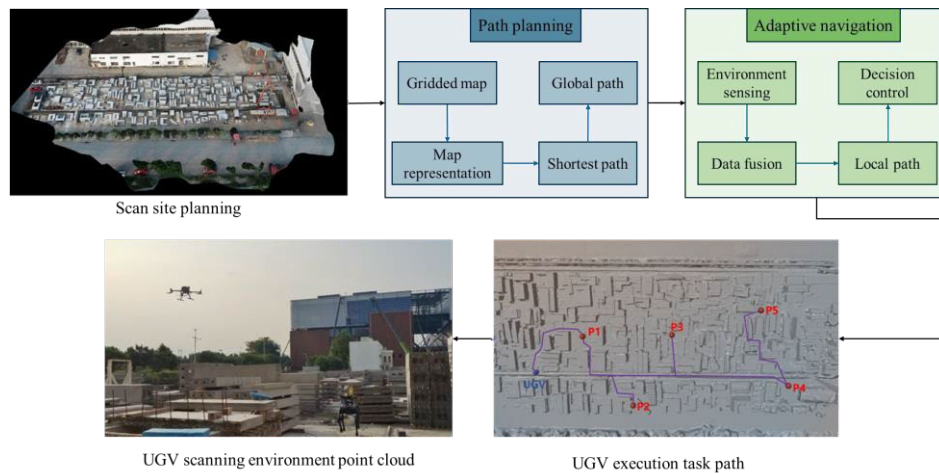


Figure 4. UGV data collection process.

However, the dynamic environment of the prefabrication plant requires adaptive navigation. This is achieved using SLAM (Simultaneous Localization and Mapping), which includes environmental sensing with LiDAR, cameras, and ultrasonic sensors to build real-time maps and identify obstacles; data fusion using Kalman and particle filtering to enhance accuracy; dynamic local path planning with the A* algorithm to ensure obstacle avoidance and efficient movement; and real-time decision-making and control to adjust the UGV's actions for safe navigation. Once the UGV reaches a designated scanning point, the LiDAR scanner captures high-precision 3D point cloud data of the surrounding environment.

3.3 Geometric quality inspection for multiple components

In the previous section, the high-precision point cloud data for the prefabricated component storage yard was obtained. This section introduces the processing of point cloud data and the methods for component recognition and inspection, including component stack extraction, target component extraction, component recognition, and quality inspection.

(1) Component stack extraction

On-site prefabricated components are typically organized by lifting sequence and types, with components of the same template and batch stacked together to save space and facilitate management. In Section 3.2.1, feasible areas and scanning targets (component stacks) were extracted. This step presents a sparse-dense component stack point clouds extraction method.

In this process, a global registration algorithm is first applied to achieve an initial alignment of the UAV and the dense UGV point clouds after voxel down-sampling. Following this, the Iterative Closest Point (ICP) algorithm is used to refine the alignment and improving the precision of the registration. Noting that the sparse UGV point cloud remains fixed during the registration process. Subsequently, the sparse point cloud collected by the UAV is clustered using a fast Euclidean clustering algorithm proposed by Cao et al.[3], generating bounding boxes for each component stack. These bounding boxes are then applied to the original high-precision point cloud data from the UGV. Since scan point clouds often contain mixed pixels, the DBSCAN algorithm is used to remove mixed pixels, resulting in isolated component stacks ready for inspection.

(2) Target component extraction

For small-prefabricated components with complex shapes, such as air-conditioning panels, balconies, and staircases, spacer blocks or beams are often added between stacked components for support. Since these prefabricated components are manufactured in a controlled factory environment following standardized processes, components of the same template and batch generally exhibit high consistency in dimensions and geometric features. As a result, a comprehensive inspection of a single component from the same template and batch can reasonably infer the quality and performance compliance of the other components.

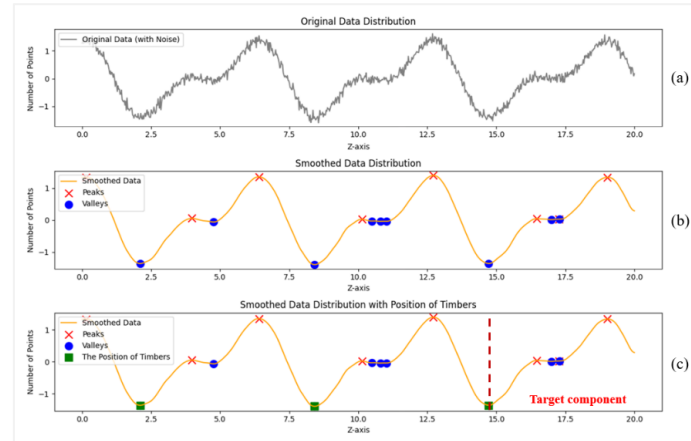


Figure 5. Target component extraction: (a) point cloud density projection on Z-axis; (b) Gaussian smoothing and valleys and peaks detection; (c) position of the timbers

In a component stack, the top component typically has the most comprehensive scan coverage, capturing both the top and four side surfaces, while other components only have their four side surfaces scanned. Therefore, it is recommended to select the top component in the stack as the representative of the component stack. To extract the representative component, a Z-axis-based point cloud density projection method is used. To minimize noise interference, Gaussian smoothing is first applied to the projected point cloud data. As shown in Figure 5. (a), the raw data with noise results in less distinct extremum points, while Figure 5. (b) shows the data curve after smoothing, where local extremum features become clearer. After smoothing, the distribution of point cloud data along the Z-axis is analyzed by calculating the first derivative to identify local maxima (peaks) and minima (valleys). Due to the lower number of scan points on spacer blocks between components, these blocks usually appear at the lowest valley in the Z-axis distribution. By identifying the lowest valley with a tolerance range, the location of the timber can be accurately determined, as illustrated in Figure 5. (c).

(3) Component recognition

Prefabricated components come in a wide variety, including floor slabs, beams, columns, walls, and stairs. These components vary significantly in shape, size, inspection criteria, and standards. Based on functional requirements, even components of the same type can differ in design and performance. For instance, floor slabs include both composite slabs and hollow-core slabs, with different specifications across projects. To identify these diverse components, this study developed a two-step classification approach: first, the PointNet++ classification network architecture determines the component category, followed by the Scan-vs-BIM method to identify specific types according to project requirements.

To support this classification, a point cloud dataset resembling ModelNet40 was constructed, covering slabs, beams, columns, walls, and stairs. This dataset combines point cloud data from onsite scans and simulated scans, with data augmentation applied through noise addition, random dropout, and scaling. After creating the dataset, training and testing were conducted. Ultimately, the PointNet++ classification network is used to classify point clouds of detected components and output the component category.

The component recognition method based on Scan-vs-BIM is applied by registering the point cloud under inspection to each BIM point cloud within the same category, with the overlap score used to evaluate the match. The component type corresponding to the BIM point cloud with the highest overlap score is assigned to the inspected point cloud. For any transformed scan point within the inspected point cloud, if a point from the BIM point cloud falls within the specified tolerance range, that point is considered overlapping. By Equation (1), the overlap rate (OR) is defined as the ratio of overlapping points to the total number of points, which is calculated on a scale from 0 to 1, where a score closer to 1 indicates a higher degree of overlap and better registration quality.

$$OR = N_{match} / N_{total} \times 100\% \quad (1)$$

- N_{match} : the number of overlapping points
- N_{total} : the total number of points

(4) Geometric quality inspection

Through the previous step, the BIM model corresponding to the scanned component is identified, and the information of the component can be matched. The method based on support vector machine

proposed by Wang et al. (Q. Wang et al., 2017) is used to segment concrete components and rebars. For concretes, horizontal and vertical points are extracted based on normal vectors. Subsequently, the RANSAC algorithm is employed to fit the planes of the nearest BIM model plane and the flatness is evaluated by the distance of each point to the corresponding BIM plane. For rebars, the center point of the bounding box of each rebar is calculated and compared with the center point of the adjacent BIM rebar to evaluate the position.

4. RESULTS

To validate the proposed method, a case study was conducted at a prefabricated component yard in Guangzhou, which produces a variety of prefabricated components, including beams, columns, slabs, and stairs, as shown in Figure 6. (a) - (b).

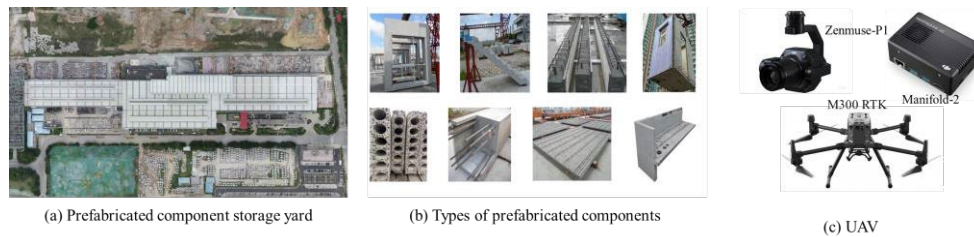


Figure 6. Experimental scenario and devices.

As shown in Figure 6. (c), a UAV (DJI M300 RTK) equipped with multifunctional HD camera (Zenmuse-P1) and edge computing module (Manifold-2) was used to capture environmental data for the prefabricated storage yard. First, during the UAV's waypoint mission, the HD camera captures real-time images of the site. The footage is then compressed into H.265 format in real-time using a compiled computing module and transmitted to the ground workstation via wireless network using the RTSP protocol. Finally, the images are processed to complete the 3D reconstruction, as shown in Figure 7.

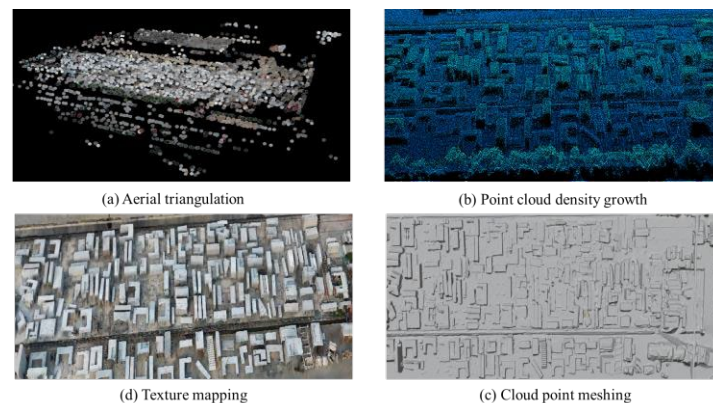


Figure 7. 3D reconstruction of the prefabricated storage yard.

Based on the 3D scene reconstructed by the UAV, feasible areas and the target component stacks are extracted to plan the scanning path for the UGV. Finally, the UGV was employed for high-precision point cloud data acquisition of the prefabricated yard equipped with a Trimble X7 laser scanner. In this case, a total of 19 stations were collected with 224,493,417 points.

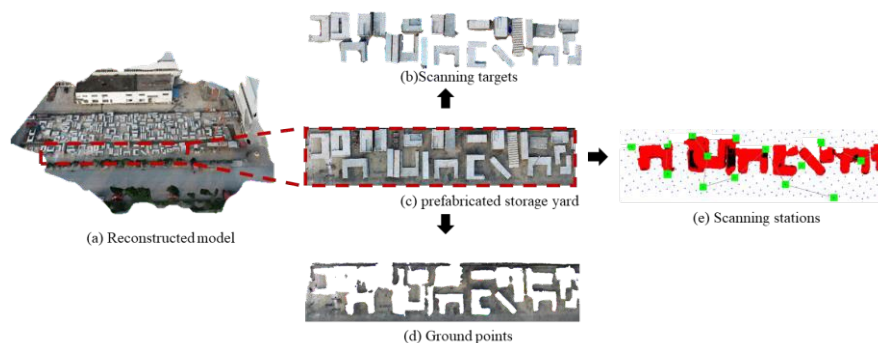


Figure 8. Prefabricated storage yard sensing and data collection.

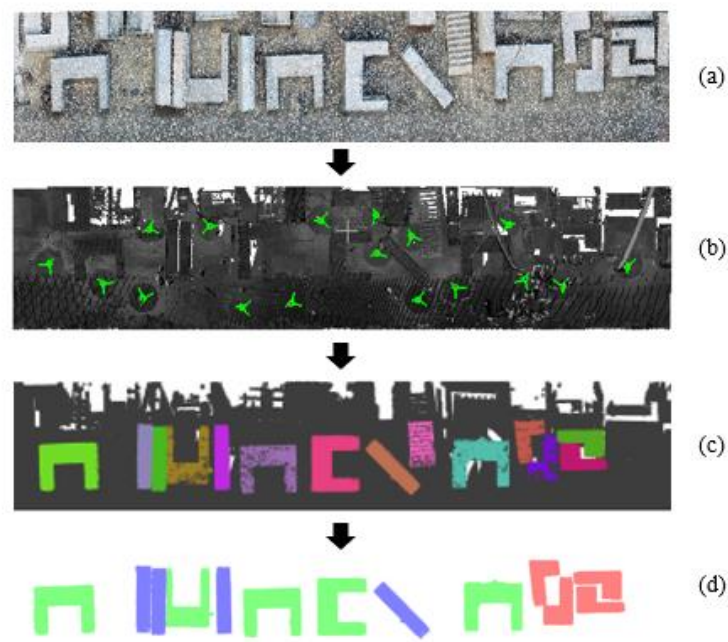


Figure 9. (a)sparse UAV point cloud; (b) dense UGV point cloud; (c) component stack extraction;(d) component recognition

After acquiring the high-precision point cloud data of the prefabricated component storage yard, individual component stacks were isolated using the proposed sparse-dense registration approach, bounding boxes, and DBSCAN for noise removal. Next, targeted components were extracted from the stacks based on point cloud density projection on Z-axis. As described in Figure 10. (b), the highest peak corresponds to the upper surface of the top component. Following this, the density curve sharply decreases and then levels off, representing the component's side edges. The lowest point in the density curve marks the spacer position between two components. Then, by extracting points with Z-axis values greater than the lowest valley, the target component point cloud is obtained (Figure 10. (c)).

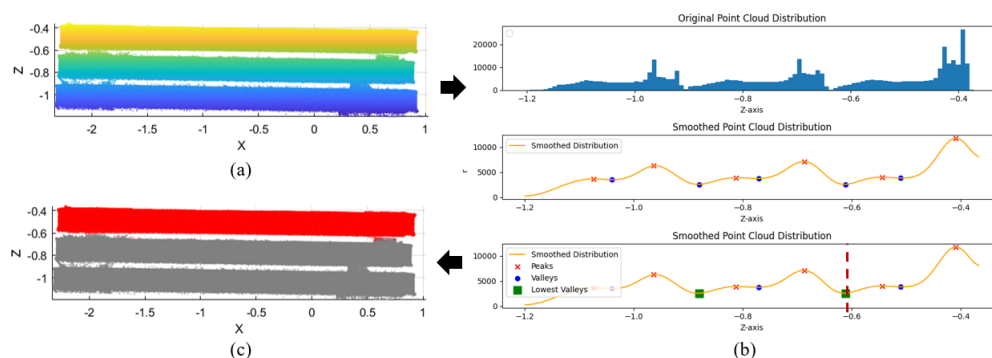


Figure 10. (a) a component stack; (b) the proposed method for target component extraction; (c) extraction of the target component

Following that, the target component point cloud was input into the PointNet++ network for classification to identify the component category, Figure 9. (d) shows the result of the recognition of each component stack. The classified point cloud is then matched individually with BIM component point clouds within the same category. The component with the highest overlap score is identified as the corresponding component (Figure 11. (b)).

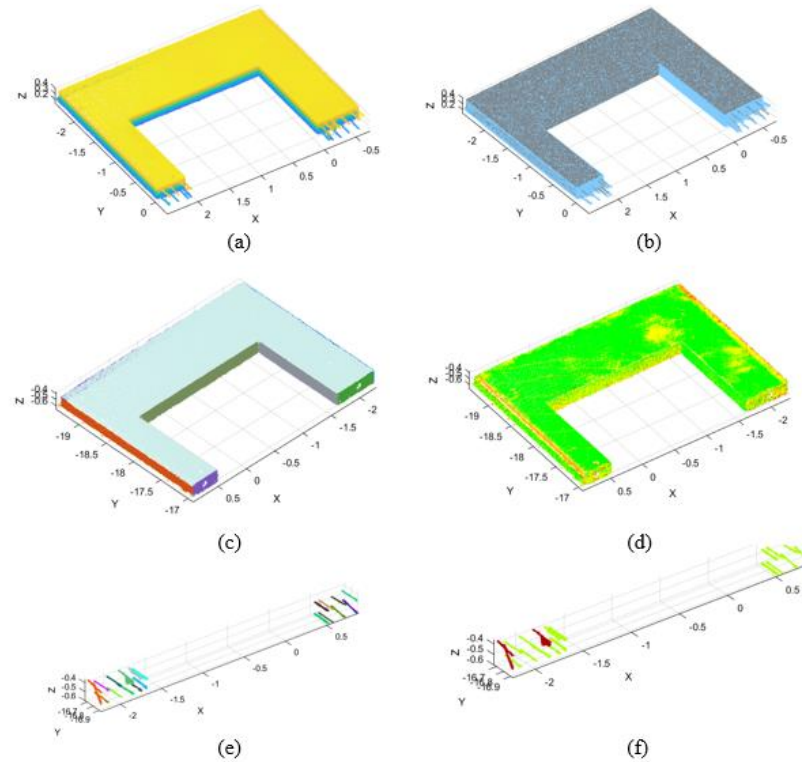


Figure 11. (a) target component; (b) target component with the corresponding BIM point cloud; (c) scanned concrete planes; (d) result for the flatness of the concrete planes; (e) scanned rebars; (f) result for the position of the rebars

After identifying the component type, the scanned point cloud was then segmented into concrete points and rebars, as shown in Figure 11. For concrete, the distance of each point to the nearest plane of the BIM point cloud is calculated to evaluate the flatness of accuracy. For rebars, the center point of the bounding boxes of each rebar is calculated and compared with the center point of the adjacent BIM rebar. Through computational validation, it was found that the proposed method can maintain an error within 5mm, with the inspection time for a single component being approximately 2 minutes, compared to 8 minutes for manual inspection. This represents a 75% increase in efficiency, which becomes even more significant as the number of components increases, which fully meets the geometric quality inspection requirements for large-prefabricated component factories, as shown in Table 1.

Table.1 Comparison of efficiency between manual and proposed method.

Method	Processing	Time	Total time for 14 components
Manual method	Flatness	5 mins	14*8=112 mins
	Rebars	3 mins	
Proposed method	Environmental sensing	40 mins	110 mins
	Data collection	42 mins	
	Geometric quality inspection	28 mins	

5. CONCLUSIONS

To automate and efficiently inspect the quality of large-prefabricated components in the yard, this study proposes a data collection and assessment framework that integrates UAV, UGV, LiDAR, and BIM. First, the UAV performs 3D reconstruction of the storage yard. Based on this 3D model, feasible regions and scan targets are identified, and the A* and genetic algorithms are used to plan the UGV's scanning path. The UGV then uses SLAM with onboard sensors to perceive real-time environmental information, enabling adaptive navigation and the collection of high-precision point cloud data across the yard. Using this high-precision data, a density projection method extracts target components, and a PointNet++ classification network combined with a Scan-vs-BIM approach identifies and inspects the geometry of components. This innovative method demonstrates high

efficiency and accuracy in data collection and component classification.

The proposed framework has been proposed and validated in large-scale prefabricated component yards, with the main contributions as follows: (1). A step-by-step collaborative mechanism for comprehensive intelligent environmental sensing based on UAV-UGV. (2). An intelligent matching and evaluation system for large-scale prefabricated components' 3D geometric features and inspection criteria based on point cloud processing has been developed.

However, several limitations remain: (1). The UGV's scanning height is limited, requiring component stacks to be below this height for full surface scanning. Future work could focus on enhancing the UGV's scanning capabilities or optimizing stacking arrangements. (2). The study primarily addresses geometric dimensions and rebar inspection, but other quality criteria, such as surface appearance (e.g., cracks, honeycombing) and embedded parts, need further consideration. Future research should aim for a more comprehensive inspection approach.

ACKNOWLEDGMENTS

This research is supported by the Young Scientists Fund of the National Natural Science Foundation of China [Grant No. 52308319] and the Natural Science Foundation of Guangdong Province, China [Grant No. 2023A1515011119].

REFERENCES

- Asadi, K., Suresh, A. K., Ender, A., Gotad, S., Maniyar, S., Anand, S., . . . Wu, T. F. (2020). An integrated UGV-UAV system for construction site data collection. *Automation in Construction*, 112. doi:10.1016/j.autcon.2019.103068
- Bosché, F. (2010). Automated recognition of 3D CAD model objects in laser scans and calculation of as-built dimensions for dimensional compliance control in construction. *Advanced Engineering Informatics*, 24(1), 107-118. doi:https://doi.org/10.1016/j.aei.2009.08.006
- Bosché, F., & Guenet, E. (2014). Automating surface flatness control using terrestrial laser scanning and building information models. *Automation in Construction*, 44, 212-226. doi:https://doi.org/10.1016/j.autcon.2014.03.028
- Chi, H. L., Kim, M. K., Liu, K. Z., Thedja, J. P. P., Seo, J., & Lee, D. E. (2022). Rebar inspection integrating augmented reality and laser scanning. *Automation in Construction*, 136, 104183. doi:https://doi.org/10.1016/j.autcon.2022.104183
- Fei Yan, J. W., Guojian He, Huan Chang, Yan Zhuang. (2020). Sparse semantic map building and relocalization for UGV using 3D point clouds in outdoor environments. *Neurocomputing*, 400, 333-342. doi:https://doi.org.ezproxy.lib.szu.edu.cn/10.1016/j.neucom.2020.02.103
- Fernando Ropero, Muñoz, P., & D.R.-Moreno, M. (2019). TERRA: A path planning algorithm for cooperative UGV-UAV exploration. *Engineering Applications of Artificial Intelligence*, 78, 260-272. doi:https://doi.org/10.1016/j.engappai.2018.11.008
- Hernandez, A., Cosmin Copot, Juan Cerquera, Harold Murcia, & Keyser, R. (2014). Formation Control of UGVs using an UAV as Remote Vision Sensor. *IFAC Proceedings Volumes*, 47, 11872-11877. doi:https://doi.org/10.3182/20140824-6-ZA-1003.01660
- Heß, R., Fritscher, M., Krauß, M., & Schilling, K. (2012). Setting up a surveillance system in the civil domain with cooperating UAVs and UGVs. *IFAC Proceedings Volumes*, 45, 19-24. doi:https://doi.org/10.3182/20121003-3-SF-4024.00003
- Kim, M.-K., Cheng, J. C. P., Sohn, H., & Chang, C.-C. (2015). A framework for dimensional and surface quality assessment of precast concrete elements using BIM and 3D laser scanning. *Automation in Construction*, 49, 225-238. doi:https://doi.org/10.1016/j.autcon.2014.07.010
- Kim, M. K., Wang, Q., Yoon, S., & Sohn, H. (2019). A mirror-aided laser scanning system for geometric quality inspection of side surfaces of precast concrete elements. *Measurement*, 141, 420-428. doi:10.1016/j.measurement.2019.04.060
- Kim, S. B. H. K. H. (2017). UAV-based automatic generation of high-resolution panorama at a construction site with a focus on preprocessing for image stitching. *Automation in Construction*, 84, 70-80. doi:https://doi.org/10.1016/j.autcon.2017.08.031
- Li, D., Liu, J., Feng, L., Zhou, Y., Liu, P., & Chen, Y. F. (2020). Terrestrial laser scanning assisted flatness quality assessment for two different types of concrete surfaces. *Measurement*, 154, 107436. doi:https://doi.org/10.1016/j.measurement.2019.107436
- Li, D., Liu, J., Hu, S., Cheng, G., Li, Y., Cao, Y., . . . Chen, Y. F. (2022). A deep learning-based indoor acceptance system for assessment on flatness and verticality quality of concrete surfaces. *Journal of Building Engineering*, 51, 104284. doi:https://doi.org/10.1016/j.job.2022.104284
- Liu Yisha, Xu Wenhao, M. Dobaie, A., & Zhuang Yanb. (2018). Autonomous road detection and modeling for UGVs using vision-laser data fusion. *Neurocomputing*, 275, 2752-2761.

- doi:<https://doi.org/10.1016/j.neucom.2017.11.042>
- NamanPatel, AnnaChoromanska, Krishnamurthy, P., & FarshadKhorrami. (2019). A deep learning gated architecture for UGV navigation robust to sensor failures. *Robotics and Autonomous Systems*, 116, 80-97. doi:<https://doi.org/10.1016/j.robot.2019.03.001>
- O.Sivaneri, V., & N.Gross, J. (2017). UGV-to-UAV cooperative ranging for robust navigation in GNSS-challenged environments. *Aerospace Science and Technology*, 71, 245-255. doi:<https://doi.org/10.1016/j.ast.2017.09.024>
- O.Sivaneri, V., & N.Gross, J. (2018). Flight-testing of a cooperative UGV-to-UAV strategy for improved positioning in challenging GNSS environments. *Aerospace Science and Technology*, 82-83, 575-582. doi:<https://doi.org/10.1016/j.ast.2018.09.035>
- Park, J., Kim, P., ho, Y. K. C., & Kangb, J. (2019). Framework for automated registration of UAV and UGV point clouds using local features in images. *Automation in Construction*, 98, 175-182. doi:<https://doi.org/10.1016/j.autcon.2018.11.024>
- Perez-Perez, Y., Golparvar-Fard, M., & El-Rayes, K. (2021). Scan2BIM-NET: Deep Learning Method for Segmentation of Point Clouds for Scan-to-BIM. *Journal of Construction Engineering and Management*, 147(9), 04021107. doi:[https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0002132](https://doi.org/10.1061/(ASCE)CO.1943-7862.0002132)
- Pierzchała, M., Giguère, a., & sAstrupa, R. (2018). Mapping forests using an unmanned ground vehicle with 3D LiDAR and graph-SLAM. *Computers and Electronics in Agriculture*, 145, 217-225. doi:<https://doi.org/10.1016/j.compag.2017.12.034>
- Rachmawati, T. S. N., & Kim, S. (2022). Unmanned Aerial Vehicles (UAV) Integration with Digital Technologies toward Construction 4.0: A Systematic Literature Review. *Sustainability*, 14(9). doi:10.3390/su14095708
- Shu, J., Li, W., Zhang, C., Gao, Y., Xiang, Y., & Ma, L. (2023). Point cloud-based dimensional quality assessment of precast concrete components using deep learning. *Journal of Building Engineering*, 106391. doi:<https://doi.org/10.1016/j.job.2023.106391>
- Tan, Y., Chen, L., Huang, M., Li, J., & Zhang, G. (2024). Automated geometric quality inspection for modular boxes using BIM and LiDAR. *Automation in Construction*, 164, 105474. doi:<https://doi.org/10.1016/j.autcon.2024.105474>
- Tan, Y., Chen, L., Wang, Q., Li, S., Deng, T., & Tang, D. (2023). Geometric Quality Assessment of Prefabricated Steel Box Girder Components Using 3D Laser Scanning and Building Information Model. *Remote Sensing*, 15(3), 556. Retrieved from <https://www.mdpi.com/2072-4292/15/3/556>
- Tan, Y., Li, S., Liu, H., Chen, P., & Zhou, Z. (2021). Automatic inspection data collection of building surface based on BIM and UAV. *Automation in Construction*, 131, 103881. doi:<https://doi.org/10.1016/j.autcon.2021.103881>
- Tan, Y., Li, S., & Wang, Q. (2020). Automated Geometric Quality Inspection of Prefabricated Housing Units Using BIM and LiDAR. *Remote Sensing*, 12(15), 2492. doi:<https://doi.org/10.3390/rs12152492>
- Wang, B., Wang, Q., Cheng, J. C., Song, C., & Yin, C. (2022). Vision-assisted BIM reconstruction from 3D LiDAR point clouds for MEP scenes. *Automation in Construction*, 133, 103997.
- Wang, Q., Cheng, J. C. P., & Sohn, H. (2017). Automated Estimation of Reinforced Precast Concrete Rebar Positions Using Colored Laser Scan Data. *Computer-Aided Civil and Infrastructure Engineering*, 32(9), 787-802. doi:<https://doi.org/10.1111/mice.12293>
- Wang, Q., Kim, M.-K., Cheng, J. C., & Sohn, H. (2016). Automated quality assessment of precast concrete elements with geometry irregularities using terrestrial laser scanning. *Automation in Construction*, 68, 170-182.
- Wang, Q., Kim, M.-K., Sohn, H., & Cheng, J. C. P. (2016). Surface flatness and distortion inspection of precast concrete elements using laser scanning technology. *Smart Structures and Systems*, 18(3), 601-623. doi:<https://doi.org/10.12989/SSS.2016.18.3.601>

BIM-BASED DISCRETE EVENT SIMULATION FOR EMBODIED CARBON ASSESSMENT

Yuqing Xu¹, Xingbo Gong², Helen H.L. Kwok³, and Jack C.P. Cheng⁴

1) Ph.D. Candidate, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: yxudv@connect.ust.hk

2) Ph.D. Student, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: xgongai@connect.ust.hk

3) Ph.D., Postdoc., Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: hlkwokab@connect.ust.hk

4) Ph.D., Professor, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: cejcheng@ust.hk

Abstract: The urgent need to reduce carbon emissions within the building and construction industry has underscored the importance of embodied carbon assessment. Building Information Modeling (BIM) and Discrete Event Simulation (DES) emerge as promising tools for enhancing this assessment process, offering detailed data extraction capabilities and dynamic simulation for energy consumption quantification. While previous research has explored the potential of BIM-DES integration, this paper addresses existing gaps by identifying and incorporating essential information requirements into BIM models for more effective DES-based embodied carbon assessment. This paper thus develops a BIM-based DES method for cradle-to-site embodied carbon assessment by (1) developing an integrated ontology to identify data requirements, (2) enriching BIM models with the necessary information for DES modeling based on the ontology, and (3) building the DES model based on the data extracted from the enriched BIM-related file and implementing scenario-based analysis. This integrated approach facilitates efficient and comprehensive analysis of cradle-to-site embodied carbon. The synergy between BIM and DES enables stakeholders to make informed decisions early in the project lifecycle, optimizing carbon reduction strategies through scenario-based analysis.

Keywords: Building information modeling (BIM), Embodied carbon assessment, Life cycle assessment (LCA), Discrete event simulation (DES)

1. INTRODUCTION

With the increase in carbon emissions from the building and construction industry, there is an urgent trend of carbon reduction in the whole industry (UNEP, 2024). As one of the major parts of the life cycle carbon emissions of buildings and construction projects, the assessment of embodied carbon plays a pivotal role in understanding the environmental impact of building materials and processes (Huang et al., 2018). Embodied carbon refers to the sum of greenhouse gas emissions associated with the production, transportation, and disposal of construction materials throughout their life cycle (UNEP, 2023). From the Life Cycle Assessment (LCA) perspective of the carbon emission management of a project, cradle-to-site embodied carbon accounts for the majority of total embodied carbon for a building infrastructure, which refers to the emissions associated with the production and procurement of building materials, from raw material extraction (A1 – A3 stages), transportation to site (A4 stage), to on-site construction activities (A5 stage) (BSI, 2011). As the global construction industry increasingly focuses on reducing carbon emissions and mitigating climate change, accurate assessment and management of cradle-to-site embodied carbon have become imperative.

However, obtaining accurate and comprehensive data can be challenging for the cradle-to-site embodied carbon assessment, especially when considering the multitude of materials and construction activities involved in cradle-to-site stages (Sharrard et al., 2008). The emergence of Building Information Modeling (BIM) as a digital representation of physical and functional characteristics of buildings has paved the way for more sophisticated and data-driven approaches to different application fields in construction management, including embodied carbon assessment (Bryde et al., 2013). It can be used to extract material type information, quantities, and other data directly as inputs for embodied carbon quantification, especially streamlining the data collection for the product (A1 – A3 stages) embodied carbon assessment. Discrete Event Simulation (DES) is another potential tool that can support data collection and analysis for cradle-to-site embodied carbon assessment. It is a method for modeling and analyzing system behavior across time, enabling the simulation of construction processes, material flows, and resource usage to measure fuel consumption during construction (Robinson, 2005). Given the challenges of managing extensive data on early-stage construction activities and the inaccuracies stemming from the impracticality of empirical data in certain construction contexts, DES serves as a valuable tool for gathering precise information through simulated outcomes for process-driven embodied carbon assessment, particularly for stages A4 and A5.

Previous research has investigated the integration of BIM and DES for cradle-to-site embodied carbon assessment. Usually, BIM data extraction offers essential quantity details for DES model elements, enabling the

simulation of energy consumption quantities crucial for cradle-to-site EC assessment (Wang et al., 2014; Dashti et al., 2021). For example, König et al. integrated BIM into the simulation input data creation process for construction scheduling management (König et al., 2012). This approach facilitated intelligent and efficient DES modeling by generating input data for construction simulation by using linked BIM data and reusable templates. However, the existing BIM-DES integration is not sufficient enough since common BIM models lack the necessary information required for DES modeling, such as detailed quantity-related data for different materials used in a single structural element, and on-site construction resource-related information. Therefore, this paper seeks to address the gap by exploring the integration of BIM with DES for embodied carbon assessment in construction projects. By identifying necessary information requirements for BIM-based DES for embodied carbon assessment, enriching the BIM model accordingly, and combining the rich information stored in BIM models with the dynamic simulation capabilities of DES, this approach offers a holistic and real-time analysis of cradle-to-site embodied carbon assessment. Through this integration, stakeholders can gain valuable insights into the environmental impact of design decisions, material selections, and construction processes based on scenario-based analysis, enabling them to optimize carbon reduction performance.

2. METHOD

Figure 1 shows the overview of the proposed method of BIM-based DES for embodied carbon assessment. There are three modules in this method. Module 1 presents the integrated ontology development process to identify data requirements and mapping relationships in BIM-based DES for embodied carbon assessment, which involves the integration of two knowledge fields, embodied carbon assessment parameters and BIM elements in the Industry Foundation Classes (IFC) format. This ontology integrating embodied carbon assessment and BIM elements is developed in a commonly used knowledge management software called Protégé. Module 2 is then designed for IFC extension based on the data requirements identified in the integrated ontology in Module, which helps semantically enrich and prepare the IFC file ready for DES-based modeling and embodied carbon assessment. In Module 3 of BIM-based DES for embodied carbon assessment, the DES model is built based on the data directly extracted from the extended IFC file. This model is also used for scenario-based analysis to identify the embodied carbon reduction strategies.

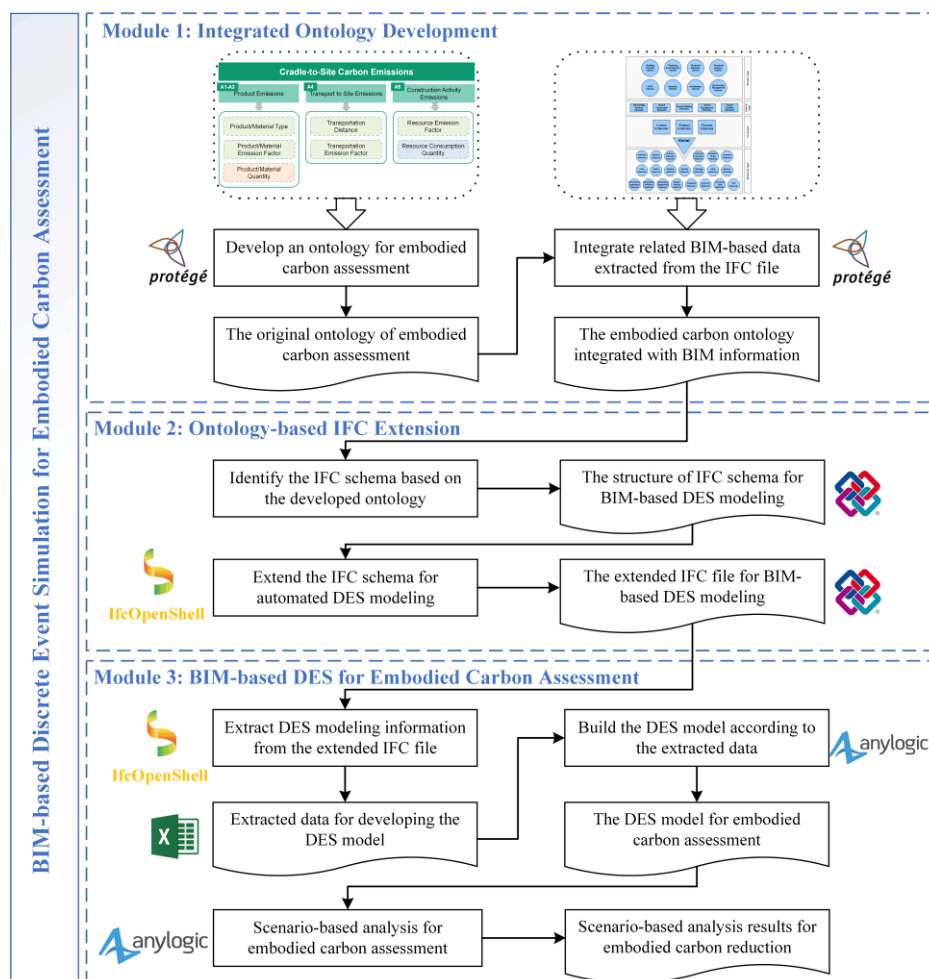


Figure 1. Method overview

2.1 Integrated Ontology Development

This section presents the ontology integration of two domains: (1) cradle-to-site EC assessment parameters to identify basic parameters for embodied carbon quantification from A1 to A5 stages, and (2) BIM elements for embodied carbon assessment to identify how BIM could store and provide such parameters required in EC assessment. Accordingly, two sub-ontologies are first developed (as shown in Figure 2) as components of the integrated ontology with class hierarchies, object property hierarchies, and data property hierarchies. In the sub-ontology for cradle-to-site embodied carbon assessment in Figure 2(a), three main cradle-to-site embodied carbon emission sources are set as individual classes in the second hierarchy, which are product carbon (A1 – A3 stages), transportation carbon (A4 stage), and construction process carbon (A5 stage). For product carbon, three main materials, concrete, formwork, and rebar, are considered in this study, and each material is equipped with its quantity and emission factor for EC assessment. Similarly, trucks and cranes are considered in this sub-ontology for transportation carbon and construction process carbon, respectively. Figure 2(b) describes all BIM elements in the format of IFC entities that are related to parameters required in cradle-to-site embodied carbon assessment, which further directs to how IFC entities can store and provide data for BIM-based DES for embodied carbon assessment. In Figure 2(b), IFC entities as classes at the second hierarchy are shown in the ontology, including *IfcMaterial*, *IfcProduct*, *IfcResource*, *IfcProcess*, *IfcPropertySet*, and *IfcQuantitySet* that can provide the necessary information in embodied carbon assessment.

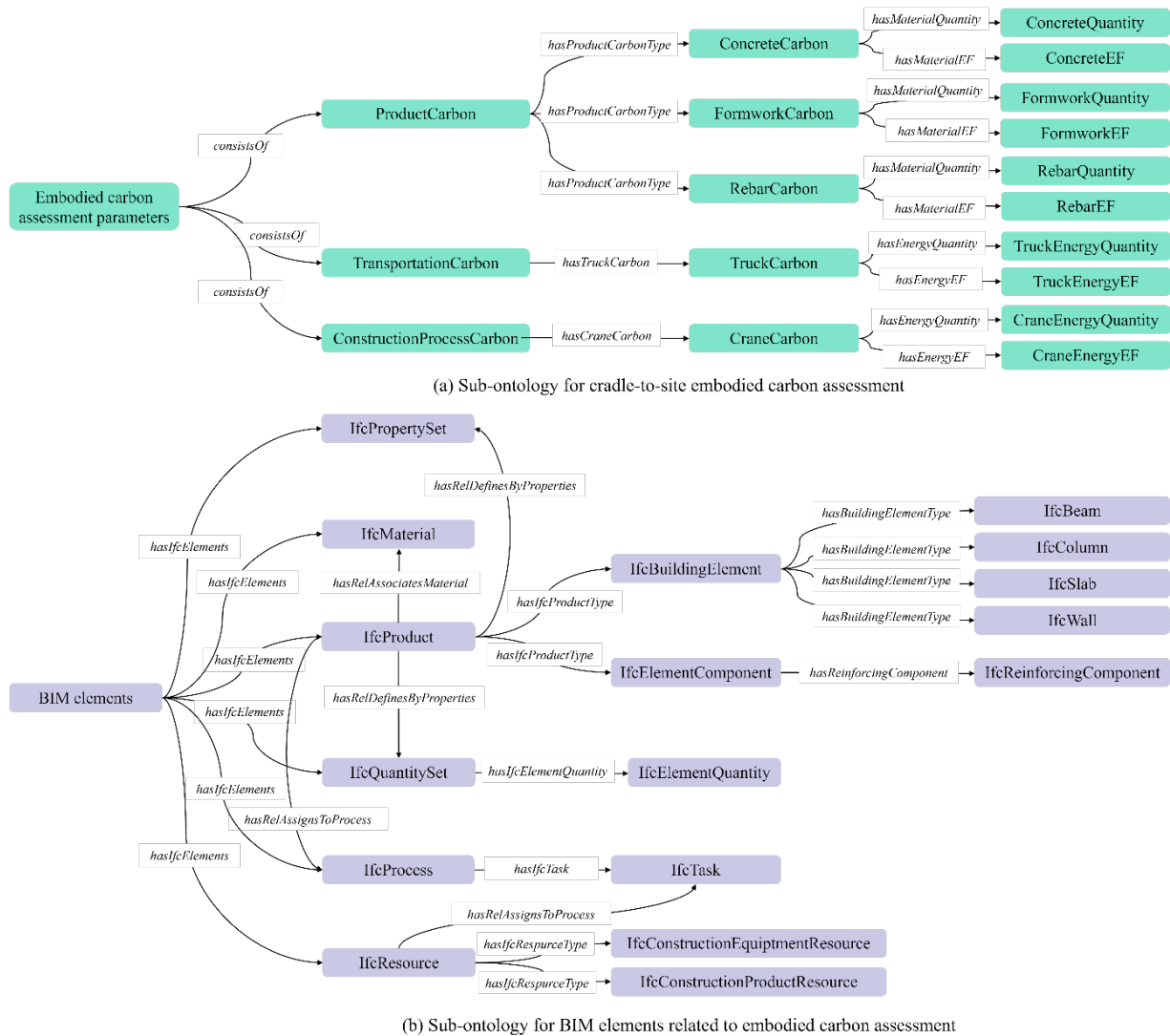


Figure 2. Sub-ontologies in related domains

Based on the two sub-ontologies above, the integrated ontology is designed with content and mapping representations between the two sub-ontologies. As shown in Figure 3, the integrated ontology keeps all classes and internal relationships in two sub-ontologies as components. Apart from these existing components, classes in the second hierarchy from two sub-ontologies are connected with external relationships, which helps to identify

mapping relationships among different domains. Three classes (product carbon, transportation carbon, and construction process carbon) from the sub-ontology that indicate three main cradle-to-site embodied carbon emission sources are connected to classes in sub-ontology of BIM elements based on how IFC entities can provide parameter data for different emission sources in embodied carbon assessment. To summarize, this integrated ontology identifies clear and structured data requirements from two domains for BIM-based DES for embodied carbon assessment and establishes mapping relationships between two domains to ensure data availability during BIM extraction, which is helpful to improve DES modeling efficiency based on direct BIM extractions.

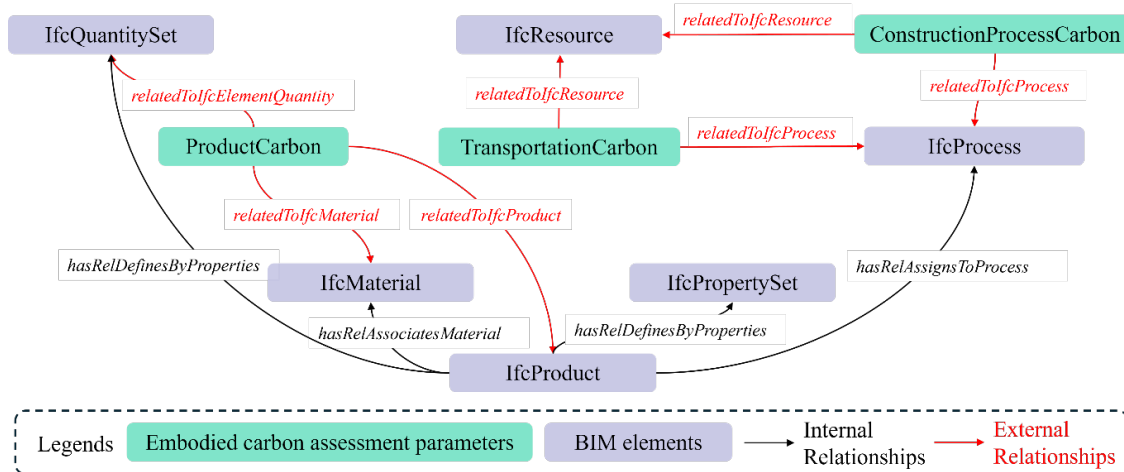


Figure 3. Integrated ontology

2.2 Ontology-based IFC extension

Although the IFC schema incorporates numerous classes referred to as “entities” to describe building information, the information identified in BIM-based DES for cradle-to-site embodied carbon assessment in the integrated ontology is beyond the current scope of the IFC schema, especially for (1) detailed quantity-related data for different material used in a single structural element, and (2) on-site construction resource-related information. To enable the IFC file to provide the necessary information required for DES modeling for embodied carbon assessment, the IFC file exported from the BIM model is semantically enriched based on the integrated ontology. The enriched IFC property sets with properties are proposed in Table. 1. For the enrichment of detailed quantity-related data, parameters of different ratios deciding the material quantities are incorporated into a new property set, designated as “Mset_ModelingSourceCommon”, which is attached to the IfcBuildingElement entities (including IfcBeam, IfcColumn, IfcWall, and IfcSlab) to provide quantity information in DES modeling. For the enrichment of on-site construction resource-related information, the enriched property sets “Mset_ModelingResourceCommon” and “Mset_ModelingResourceIndicators” are added to IfcConstructionEquipmentResource to store basic equipment parameters (such as number and loading capacities) and emission-related indicators of equipment operations, respectively. The initial IFC schema is then further extended by IfcOpenshell to incorporate all these mentioned property sets with properties to realize BIM-based DES modeling for cradle-to-site embodied carbon assessment.

Table 1. Extended property sets for DES-based modeling

Property set	Property	Description
Mset_ModelingSourceIndicators	M_ConcreteRatio	The ratio of actual concrete use to element quantity in structural elements
	M_RebarRatio	The ratio of actual rebar use to element quantity in structural elements
	M_FormworkRatio	The ratio of actual formwork use to element quantity in structural elements
Mset_ModelingResourceCommon	M_ResourceNumber	The number of construction equipment on-site
	M_ResourceCapacity4Concrete	The capacity for concrete (m ³)
	M_ResourceCapacity4Rebar	The capacity for rebar (kg)
	M_ResourceCapacity4Formwork	The capacity for formwork (m ²)

Property set	Property	Description
Mset_ModelingResourceIndicators	M_ResourceBusyEF	The EF of equipment during operation
	M_ResourceBusyEFUnit	The EF unit of equipment during operation
	M_ResourceIdleEF	The EF of equipment during idling
	M_ResourceIdleEFUnit	The EF unit of equipment during idling

2.3 BIM-based DES for Embodied Carbon Assessment

The last module in this method is developing the corresponding DES model based on the data extracted from the BIM-related files to implement embodied carbon assessment under different scenarios. The whole process is shown in Figure 4. After semantic enrichment in the IFC file based on the developed integrated ontology, all data related to DES modeling for embodied carbon assessment is extracted from the enriched IFC file via IfcOpenShell. The DES modeling-related data is then stored in an Excel file as the output of the developed extraction tool, providing data sources for building the corresponding DES model. AnyLogic is used as the DES modeling tool in this study. Based on the data in Excel, the corresponding DES model is then developed for simulation and scenario-based embodied carbon assessment to implement carbon reduction strategies. The detailed scenarios considered and final simulation results in this study are introduced in Section 3.

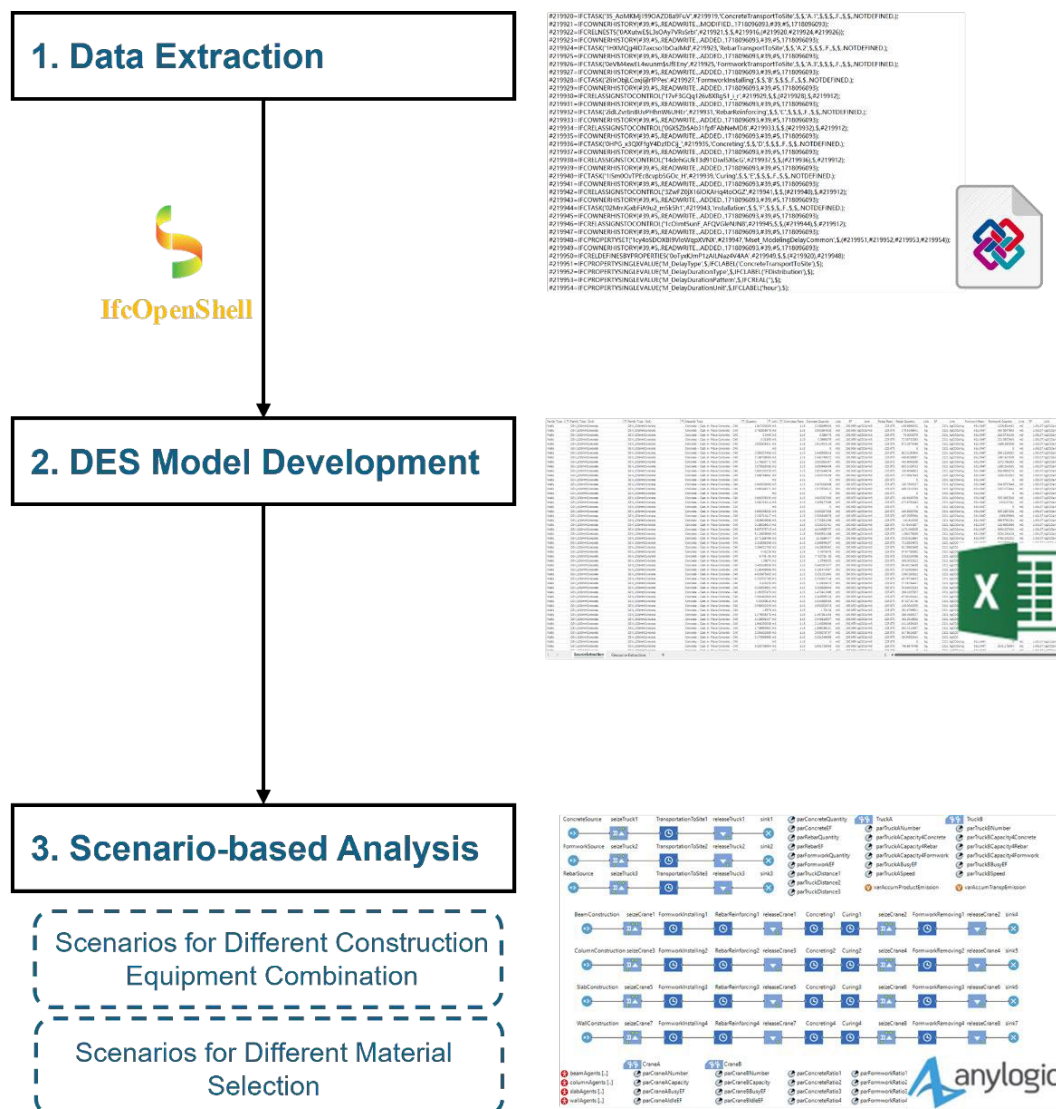


Figure 4. Process map of Module 3

3. RESULTS

To validate the workability of the proposed method of BIM-based DES for embodied carbon assessment, the ground floor of a 2-story building with a reinforced concrete structure located in Hong Kong is used for the validation. The IFC file of the case study extracted from the BIM model is semantically enriched based on the data requirements from the developed integrated ontology. Figure 5 shows the example of the extended property set “Mset_ModelingSourceCommon” with three properties with values in the IFC schema. In the next step, all data related to DES modeling stored in the enriched IFC file is extracted via IfcOpenShell and stored in the Excel file (as shown in Figure 4) to support the data sources for the DES model development. After manually building the DES model based on the Excel data, the simulation result of embodied carbon in this case study is generated, which is presented in Figure 6 with three parts following the LCA stages.

Extended Mset_ModelingSourceCommon with three properties with values

```
#217435=IFCPROPERTYSINGLEVALUE('M_ConcreteRatio',$,IFCREAL(1.15),$);
#217436=IFCPROPERTYSINGLEVALUE('M_RebarRatio',$,IFCREAL(225.98),$);
#217437=IFCPROPERTYSINGLEVALUE('M_FormworkRatio',$,IFCREAL(8.96),$);
#217438=IFCPROPERTYSET('3seG4E1M9Ag9yeAQF3l_FN',#42,'Mset_ModelingSourceCommon',$,(#217435,#217436,#217437));
#217439=IFCRELDEFINESBYPROPERTIES('0FxQ2uRq57ouuiUi$6afB$',#42,$,($419),#217438);
```

Figure 5. Example of extended IFC schema

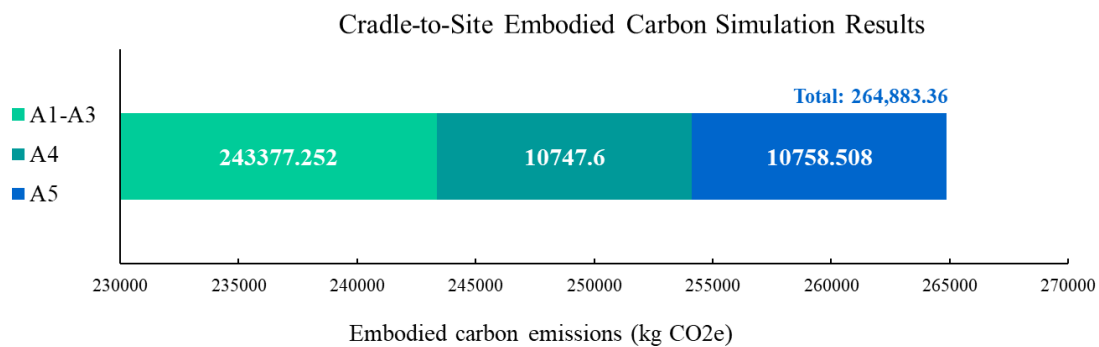


Figure 6. DES result of embodied carbon quantification (in the base scenario)

Figure 6 shows the cradle-to-site embodied carbon results of the base scenario with initial settings for both construction material types and on-site construction equipment use. To better assess how construction materials and construction equipment influence total embodied carbon emissions, this study implements a scenario-based analysis regarding different materials and equipment use in the case study. Figures 7 and 8 present the embodied carbon emission results under different scenarios. As shown in Figure 7, utilizing materials can greatly affect the total embodied carbon, especially the use of concrete with high-ground granulated blast slag (GGBS) proportions and highly recycled rebar. At the same time, emissions from material transportation should be considered together before final material selections. In Figure 8, different combinations of on-site construction equipment have slight effects on the total cradle-to-site embodied carbon but can still reduce 10% to 30% embodied carbon emissions at corresponding stages compared to the base scenario.

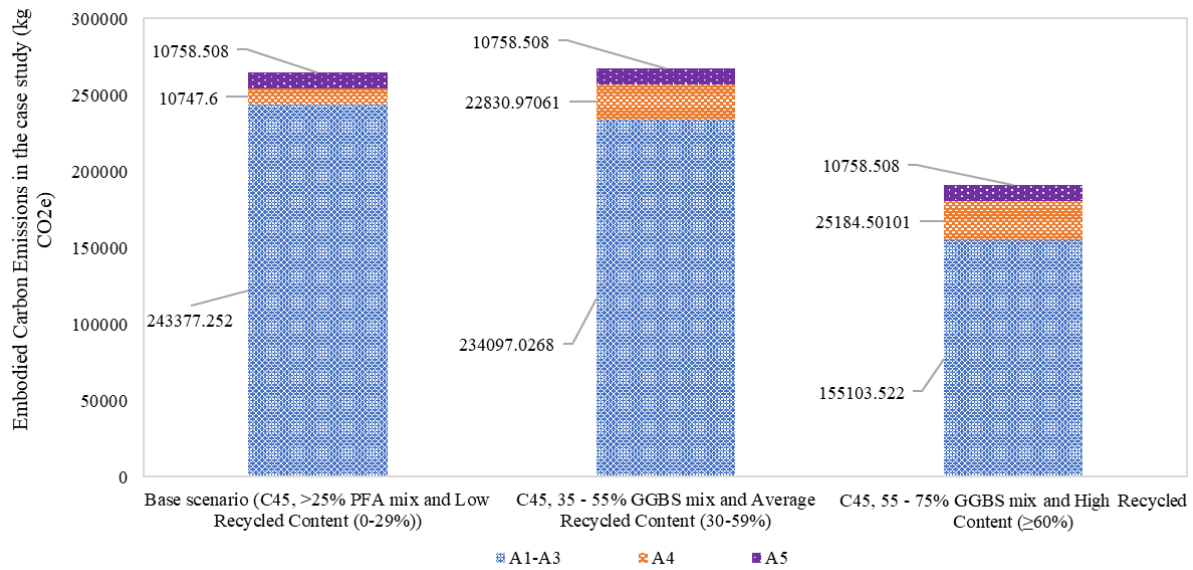


Figure 7. Embodied carbon emissions in the case study with different materials

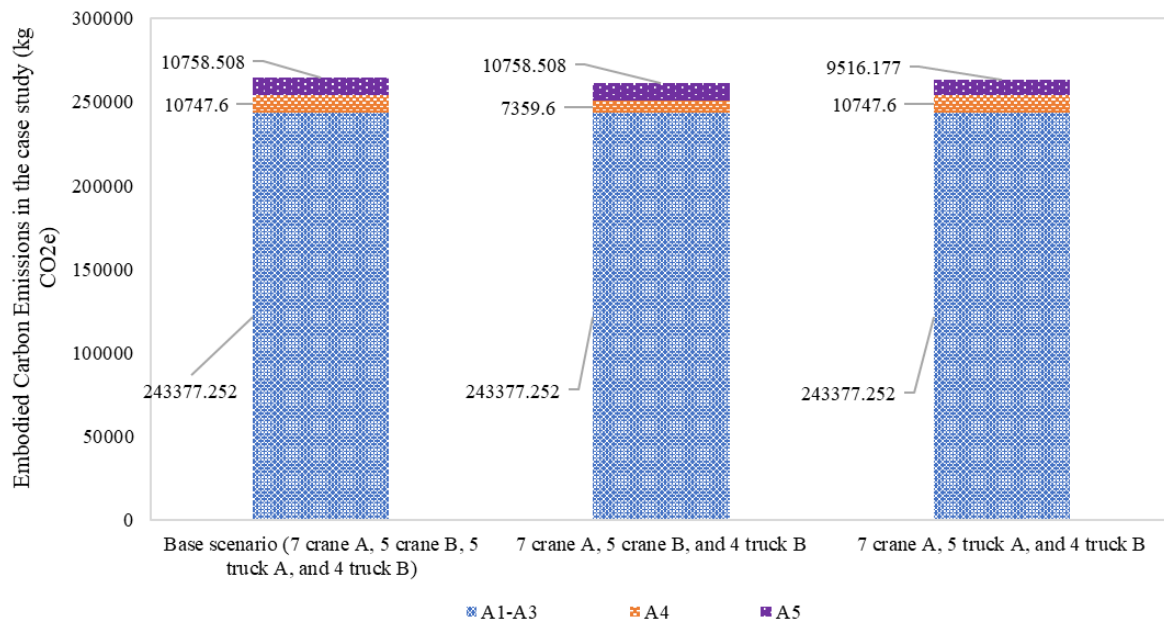


Figure 8. Embodied carbon emissions in the case study with different construction equipment quantities

4. DISCUSSION

The integration of BIM with DES for embodied carbon assessment in this paper presents a significant advancement in sustainable construction practices. By addressing the challenges associated with data collection for cradle-to-site embodied carbon assessment, this integrated approach offers a more holistic understanding of the environmental impact of construction materials and activities. The combination of BIM's detailed material and quantity information extraction capabilities with DES's dynamic simulation abilities allows for real-time analysis of energy consumption and carbon emissions throughout the project lifecycle. Compared with previous studies, the BIM-DES integration in this study improves the data richness of the single data source, the BIM model, by enriching the necessary data via the IFC file, which helps enhance the efficiency of BIM-based DES modeling for embodied carbon assessment. This integration also streamlines the assessment process, enabling stakeholders to identify carbon hotspots, optimize material selection, and enhance construction processes for improved environmental performance.

5. CONCLUSIONS

This paper proposes and develops BIM-based DES for cradle-to-site embodied carbon assessment, which mainly includes three parts: integrated ontology development, ontology-based IFC extension, and BIM-based DES

for embodied carbon assessment. The integration of BIM and DES holds great promise for advancing embodied carbon assessment in construction projects. By bridging the gap between detailed data extraction and dynamic simulation, this integrated approach provides stakeholders with valuable insights into the environmental impact of their decisions. Through scenario-based analysis and optimized carbon reduction strategies, construction industry professionals can make informed choices that minimize carbon emissions and contribute to more sustainable building practices. Moving forward, further research and implementation of BIM-DES integration are essential for improving the efficiency of DES modeling by realizing automated DES modeling based on BIM data.

REFERENCES

- British Standards Institution. (2011). BS EN 15978 Sustainability of construction works -assessment of environmental performance of buildings -calculation method, British Standards Institution, London.
- Bryde, D., Broquetas, M., & Volm, J. M. (2013). The project benefits of building information modelling (BIM). *International journal of project management*, 31(7), 971-980.
- Dashti, M. S., RezaZadeh, M., Khanzadi, M., & Taghaddos, H. (2021). Integrated BIM-based simulation for automated time-space conflict management in construction projects. *Automation in Construction*, 132, 103957.
- Huang, L., Krigsvoll, G., Johansen, F., Liu, Y., & Zhang, X. (2018). Carbon emission of global construction sector. *Renewable and sustainable energy reviews*, 81, 1906-1916.
- König, M., Koch, C., Habenicht, I., & Spieckermann, S. (2012, December). Intelligent BIM-based construction scheduling using discrete event simulation. In *Proceedings of the 2012 Winter Simulation Conference (WSC)* (pp. 1-12). IEEE.
- Robinson, S. (2005). Discrete-event simulation: from the pioneers to the present, what next?. *Journal of the Operational Research Society*, 56(6), 619-629.
- Sharrard, A. L., Matthews, H. S., & Ries, R. J. (2008). Estimating construction project environmental effects using an input-output-based hybrid life-cycle assessment model. *Journal of Infrastructure Systems*, 14(4), 327-336.
- United Nations Environment Programme, & Yale Center. (2023). Ecosystems + Architecture, Building Materials and the Climate: Constructing a New Future.
- United Nations Environment Programme (2024). Global Status Report for Buildings and Construction: Beyond foundations: Mainstreaming sustainable solutions to cut emissions from the buildings sector.
- Wang, W. C., Weng, S. W., Wang, S. H., & Chen, C. Y. (2014). Integrating building information models with construction process simulations for project scheduling support. *Automation in construction*, 37, 68-80.

ENHANCING SCAN-TO-BIM IN BRIDGE ENGINEERING THROUGH POINT CLOUD COMPLETION

Tao Yang¹, Yang Zou², and Enrique del Rey Castillo³

1) Ph.D. Candidate, Department of Civil and Environmental Engineering, University of Auckland, Auckland 1010, New Zealand. Email: tyan632@aucklanduni.ac.nz

2) Ph.D., Senior Lecturer, Department of Civil and Environmental Engineering, University of Auckland, Auckland 1010, New Zealand. Email: yang.zou@auckland.ac.nz.

3) Ph.D., Senior Lecturer, Department of Civil and Environmental Engineering, University of Auckland, Auckland 1010, New Zealand. Email: e.delrey@auckland.ac.nz.

Abstract: Automatic generation of Building Information Models (BIMs) from point clouds (i.e., Scan-to-BIM) plays a critical role in bridge maintenance and the development of Digital Twinning (DT). However, the problem of incomplete point cloud (e.g., caused by occlusions in laser scanning) significantly hinders the Scan-to-BIM accuracy. To overcome this challenge, we propose addressing the occlusion problems in bridge point clouds by introducing an additional point cloud completion task in the Scan-to-BIM process. This new task aims to take incomplete bridge point clouds, following segmentation, as input, and generate complete point clouds as output. The learning-based completion model, Point Completion Network (PCN), is adopted to validate the proposed strategy and show robust completion performance for bridge components with varying levels of occlusion. It can improve the average 11.94 Chamfer Distance (CD) and 11.05 F-score for coarse completion, 8.35 CD and 1.59 F-score for dense completion. This study contributes to Scan-to-BIM by refining the Scan-to-BIM framework in bridge engineering and defining a point cloud completion task to facilitate the development of bridge DT systems.

Keywords: Scan-to-BIM, bridge point cloud, point cloud completion, geometric digital twinning

1. INTRODUCTION

Bridges are essential in the infrastructure system but often suffer serious health issues due to prolonged usage and environmental factors (American Society of Civil Engineers ASCE, 2021). However, traditional bridge inspection methods, such as human-labour-based data collection and paper-based data management, are always subjective and inefficient, which results in a pressing demand for efficient and smart bridge maintenance, management, and operation systems. In recent years, Building Information Modelling (BIM) has been introduced into bridge management. BIM provides a robust platform for improving traditional bridge management practices by offering a centralised digital environment for storing, visualising, and managing detailed bridge information. By integrating emerging photogrammetry techniques and Artificial Intelligence (AI), as-is condition data of bridges can be rapidly captured and analysed to enhance bridge monitoring and management (Khudhair et al., 2021). In this regard, the Digital Twin (DT) was further proposed to form a more comprehensive and systematic framework for the whole life cycle management of bridges (Honghong et al., 2023; Pregnotato et al., 2022). The DT can replicate physical bridges into their virtual twin models in the operation and maintenance phase (O&M), constructing and updating them based on multiple data sources and adding new functions to simulate different operation scenarios (Honghong et al., 2023).

To build a bridge DT, it is critical to reconstruct a geometric and semantically rich BIM model, reflecting the as-is condition of bridges. The model can serve as the foundation for subsequent bridge DT applications, such as decision marking, real-time monitoring, and visualisation (Honghong et al., 2023). Obtaining the geometric information is often the first step for reconstructing a semantically rich BIM model, known as Geometric Digital Twin (gDT) (Brilakis, 2024). Various emerging reality capture tools, such as laser scanning and photogrammetry techniques, have been introduced to capture geometric information primarily through 3D point cloud data. The point cloud data can be used for constructing geometric BIM for bridges when BIM models are not generated from the design stage (Scan-to-BIM) (Bosché, 2010; Bosché et al., 2014). However, the reconstruction process is highly human-dependent, low-efficiency, and error-prone (Lu et al., 2019). Automating these tasks still remains a global challenge.

Due to the general absence of BIM models for most existing bridges, Scan-to-BIM research has attracted growing attention (Schönfelder et al., 2023). In recent years, extensive research has focused on the Scan-to-BIM of different bridge types, including RC girder bridges, steel girder bridges, masonry arch bridges, and truss bridges (Yang et al., 2024). However, most research focused on point cloud segmentation tasks, and the research on bridge geometric information reconstruction has not been fully investigated. Only a few studies focused on this topic (Justo et al., 2023; Lu & Brilakis, 2019; Mafipour & Vilgertshofer, 2023; Mehranfar et al., 2021; Walsh et al., 2013; Yan & Hajjar, 2022). The proposed methods are primarily for some sample slab-beam bridges, and the reconstruction quality is highly affected by point cloud quality problems, particularly occlusion (Zhao et al., 2023). Although some strategies were designed to address occlusion problems in their methods, the over-consideration of data quality problems in reconstruction workflow poses more limitations on the selection of reconstruction

algorithm, resulting in poor generalizability. No prior research has specifically aimed to complete the missing regions in bridge point clouds.

To address this challenge, this paper defines a new point cloud completion task in the existing Scan-to-BIM framework, enhancing geometric BIM reconstruction in bridge engineering, as shown in Figure 1. This task will take incomplete bridge point clouds following segmentation as input and aim to generate complete point clouds as output. To validate the proposed strategy, the learning-based completion model, Point Completion Network (PCN), is introduced to recover the complete point cloud from raw incomplete input with a coarse-to-fine strategy.

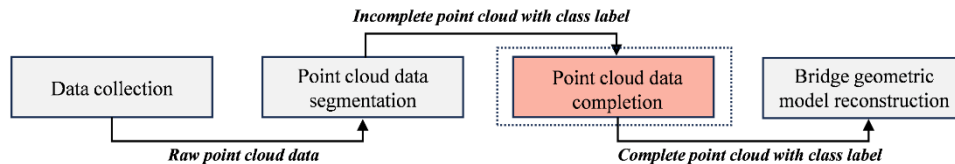


Figure 1. Role of point cloud completion in Scan-to-BIM.

2. RELATED WORK

2.1 Geometric Digital Twinning of Bridges

Reconstructing 3D bridge geometric information from point clouds can be considered as a model-to-cloud fitting problem. To achieve this, the Constructive Solid Geometry (CSG) method was first adopted to generate the 3D model of bridge piers and pier caps (Walsh et al., 2013). A shape library containing generic objects (e.g., cuboid, cylinder) was established to fit the point cloud clusters. However, the CSG method relies highly on pre-defined geometrically simple solid primitives, which fail to capture complex bridge geometric information. After that, the Swept Solid Representation (SSR) based method was proposed to create a 3D shape by sweeping a 2D profile enclosed by a contour line along a specific path in 3D space. The bounding hulls, such as convex hulls (Preparata & Hong, 1977) and α -shape (Edelsbrunner & Mücke, 1994), are commonly adopted for 2D profile fitting. Lu and Brilakis (2019) proposed an object-fitting method to generate a gDT of existing RC bridges based on Industry Foundation Classes (IFC). The 2D ConcaveHull α -shape method was adopted for shape representation and parameter extraction, and four types of point clusters, including pier, pier cap, girder, and slab, are considered. Despite the satisfactory results, some point cloud quality problems, such as occlusion and uneven density, have been revealed to cause mismatch points and problems in concave hull generation.

An early model generation for steel girder bridges is proposed by Yan and Hajja (2022). The linear skeleton model was designed to extract the specific parameter value for 3D modelling. The scan information in data collection was utilised to improve the RANSAC-based parameter fitting, addressing the occlusion in model reconstruction. However, a pre-defined dummy model is required in the initial stages, which limits its ability to reconstruct other elements with different geometric features. Moreover, the parameter extraction accuracy relies heavily on the quality of occlusion labelling, and the significant occlusion remains a challenge in model reconstruction. To address occlusion issues, Mafipour and Vilgertshofer (2023) proposed a Parametric Prototype Model (PPM) to extract the parameters from segmented point clouds. The definition of different PPMs in the initial stage can enable the reliable parameter identification of bridge components and show satisfactory performance in sustaining occlusion. However, this method relies heavily on the PPM library, and prior knowledge and statistical study are required for parameter range setting. More importantly, the occlusion may not show much influence on the parameter extraction but will significantly affect the accurate selection of a PPM.

2.2 Learning-based Point Cloud Completion

The learning-based completion method is a promising solution for dealing with point cloud occlusion, arousing growing interest (Zhuang et al., 2024). The model can learn local and global features of incomplete input and output complete object shape and semantics. Research on learning-based completion methods can be classified into volumetric shape completion and point-wise completion according to different types of input data. However, translating the point cloud into 3D volume data would result in the loss of detailed geometric information, rendering it less generalisable for intricate shapes or 3D model construction with high LoD. Thus, many point-wise completion models have been proposed to generate complete shapes from given incomplete point clouds in recent years (Huang et al., 2020; Liu et al., 2020; Tchapmi et al., 2019; Wang et al., 2022; Wen et al., 2020; Yuan et al., 2018). The pioneer completion model, PCN, was proposed by Yuan et al. (2018), where an encoder-decoder architecture was adopted to complete the point cloud in a coarse-to-fine manner. After that, different improvements on the network were proposed to achieve better results, such as the feature-points-based completion network and multi-scale generating network in PF-Net (Huang et al., 2020), the skip-attention mechanism in SA-Net (Wen et al., 2020), and multi-stages strategy in (Liu et al., 2020; Wang et al., 2022).

3. PROPOSED SOLUTION

This paper adopts raw PCN for bridge point cloud completion, addressing the occlusion problems in bridge point clouds prior to geometric reconstruction. The model is selected in this study for the following three reasons: (1) PCN is the classical and pioneering model for point cloud completion, with the most popular encoder-decoder network; (2) Despite being published in 2018, the completion performance still outperforms some newly published model in some shape types in the public dataset, e.g., plane, cabinet and car in ShapeNet (Yi et al., 2016); (3) Many completion models are limited to generating a fixed number of points e.g., 2048 or 4096 points. PCN is superior to some models by outputting dense point clouds through a coarse-to-fine strategy, significantly preserving the geometric features for BIM reconstruction.

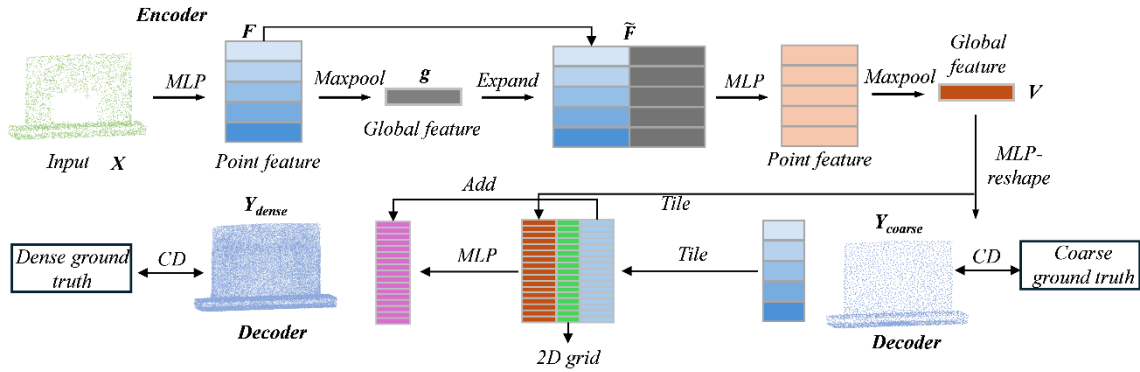


Figure 2. The architecture of the PCN model.

As shown in Figure 2, PCN follows an encoder-decoder architecture. The encoder consists of two stacked PointNet (PN) layers. The input partial point cloud X is initially encoded into k -dimensional feature vector F through a multi-layer perceptron (MLP). A point-wise max-pooling operation is then applied to obtain k -dimensional global feature g . The second PN layer takes both g and F as inputs, outputting the augmented point feature \tilde{F} through feature concatenation. Finally, another MLP is adopted to generate the final global vector V .

The PCN adopts a multi-stage point generation process, starting from a coarse point cloud Y_{coarse} and progressing to dense point cloud Y_{dense} . A fully-connected decoder is utilised for coarse point cloud ($s \times 3$) generation, where the point number of coarse output s is set as 4096 points in this study. The generated output is compared with coarse ground truth data for loss calculations. After that, a folding-based operation is used to generate dense point clouds. In this step, a patch of t points is generated in the local area centred at each point P_i from Y_{coarse} and then transformed into the global coordinates by adding P_i to the output. The final result Y_{dense} consists of $n=st$ points, where the t is set as 4 and finally output 16384 points as dense output. Then, the loss is calculated between Y_{dense} and dense ground truth data. In this study, the symmetric version of Chamfer Distance (CD) is adopted to measure the difference between the output point cloud (Y_{out}) and the ground truth point cloud (Y_{GT}), as defined by Equation (1). Accordingly, the total loss for model training is defined by Equation (2).

$$L_{CD}(Y_{out}, Y_{GT}) = \frac{1}{|Y_{out}|} \sum_{p \in Y_{out}} \min_{q \in Y_{GT}} \|p - q\|_2^2 + \frac{1}{|Y_{GT}|} \sum_{q \in Y_{GT}} \min_{p \in Y_{out}} \|q - p\|_2^2 \quad (1)$$

$$L_{total}(Y_{coarse}, Y_{dense}, Y_{GT}) = L_{CD}(Y_{coarse}, \tilde{Y}_{GT}) + \alpha L_{CD}(Y_{dense}, Y_{GT}) \quad (2)$$

where the \tilde{Y}_{GT} and Y_{GT} are the ground truth of coarse point clouds and dense point clouds, respectively. α is the weight hyperparameter. The first term indicates the loss of coarse completion, and the second term represents the loss of dense completion. More details about the PCN model can be found in (Yuan et al., 2018), and the author's code can be accessed on GitHub (<https://github.com/wentaoyuan/pcn>).

4. EXPERIMENT AND RESULTS

This section details the results of the point cloud completion model for the validation of the proposed completion strategy. Three bridge components in the substructure are considered in this study, including piers, pier caps, and abutments, as shown in Figure 3. Commonly used types are selected for each component. The pier consists of circular piers, wall piers, and rectangular piers. Three types of pier caps are considered, as shown in Figure 3(d) to (f). Additionally, two typical abutment types are included, as shown in Figure 3(g) and (h).

To generate paired partial and completed point clouds, a series of 3D models for each component type are generated beforehand. Each component was modelled as a Revit family, constrained by several design parameters,

and then, C#-based Revit API was used to generate parametric models with varying design parameters automatically. Afterwards, the point clouds were generated from mesh sampling with a pre-defined point number set as 100000 points in this study. Voxel-based down-sampling is then performed to obtain dense complete point clouds (16384 points). The partial coarse point cloud is created by occlusion generation and down-sampling. Two different partial point clouds were generated from each complete point cloud, including the partial point clouds with a small occlusion and a large occlusion. The occlusion is created in three steps:

- Seed point selection: Random seed point selection from the point cloud.
- Radius calculation. This process begins by determining the coordinate range for X, Y, and Z, from which the maximum range is denoted as L . Then, the radius for small occlusion is calculated as $[0, 0.1L]$ and for large occlusion, as $[0.1L, 0.3L]$.
- Point removal. Once the seed point and radius are determined, points within the defined radius around the seed point are removed to generate the final partial point cloud.

Finally, the partial point clouds were uniformly down-sampled to 4096 points, providing the raw input for model training.

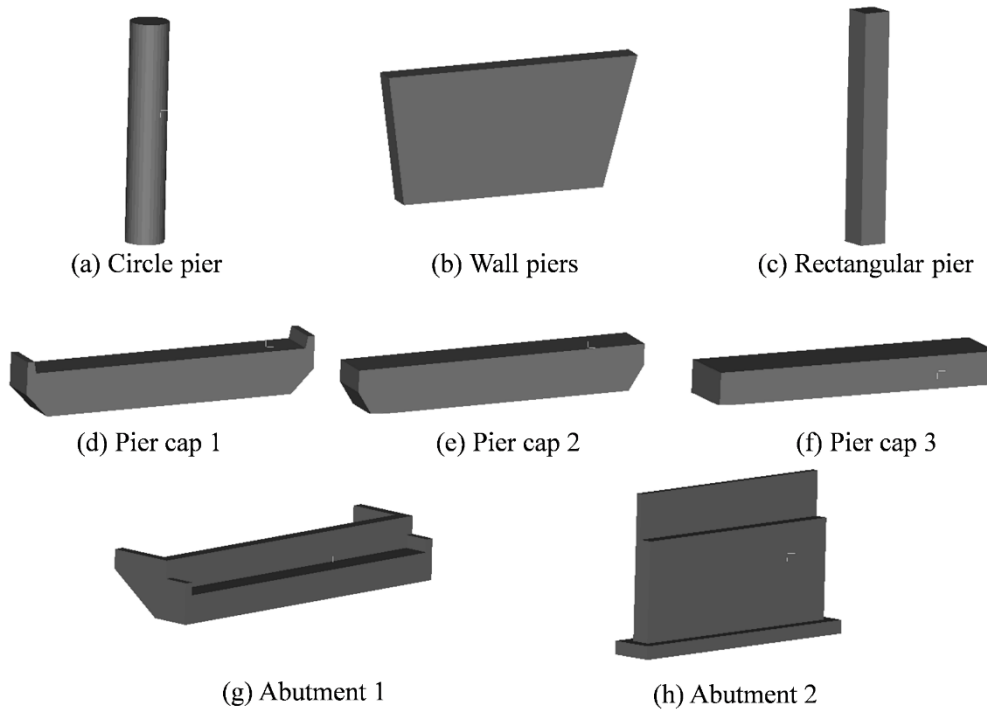


Figure 3. Component types in the training dataset.

4.1 Implementation Details

The Pytorch framework was adopted for implementation on Ubuntu 22.04 through Intel® Xeon(R) Gold 6242R CPU and a single NVIDIA Quadro RTX 5000 GPU with 16 G memory. We trained the completion model for each bridge component to achieve the best completion performance. The dataset finally consists of 22500 piers, 25000 pier caps and 29000 abutments. The dataset was randomly split, with 90% used for training, 5% for validation, and the remaining 5% reserved for testing. Finally, the test and validation datasets contain 1200 pier instances, 1300 pier cap instances, and 1500 abutment instances, respectively. All our models were trained using the Adam optimiser with an initial learning rate of 0.0001 for 150 epochs and a batch size of 16. The learning rate decayed by 0.7 for every 50,000 iterations to fine-tune model performance.

4.2 Result and Evaluation

Point-wise metrics were employed to evaluate the performance of point cloud completion. The Chamfer Distance (CD) and F1-score were adopted for quantitative evaluation (Zhuang et al., 2024). The CD calculates the average closest distance between the output point cloud Y_{out} and the ground truth point cloud Y_{GT} , where the symmetric version of CD was used for comprehensive evaluation, as defined by Equation (1). The F-score, as shown in Equation (3), can assess the percentage of correctly reconstructed points according to pre-defined d , calculated by Precision (P) in Equation (4) and Recall (R) in Equation (5).

$$F-Score(d) = \frac{2P(d)R(d)}{P(d) + R(d)} \quad (3)$$

$$P(d) = \frac{1}{|X_{out}|} \sum_{p \in X_{out}} \left[\min_{q \in X_{GT}} \|q-p\| < d \right] \quad (4)$$

$$R(d) = \frac{1}{|X_{GT}|} \sum_{p \in X_{GT}} \left[\min_{q \in X_{out}} \|q-p\| < d \right] \quad (5)$$

The completion results of point clouds with large occlusion for abutments, pier caps, and piers are shown in Figures 4, 5, and 6, respectively. It can be observed that the PCN can effectively complete the occlusion in the raw point cloud, even with a significant loss of geometric information, outputting coarse results (4096 points) and dense results (16384 points). The prediction model can recover the missing point cloud from partial input and output the coarse results with clear boundary details. However, for dense results, some loss of detailed boundary geometric information can be found in these figures from the observation components boundary. Thus, while the PCN model can accurately predict the missing regions of the partial input using its fully connected decoder, the folding-based decoder for dense point cloud generation tends to result in a loss of boundary detail. This is because the folding-based decoder aims to predict approximating smooth surfaces, which limits its ability to recover sharp features along the component boundaries.

	Partial input	Coarse completed	Dense completed	Ground truth
Pier type-1				
Pier type-2				
Pier type-3				

Figure 4. Completion results for piers.

	Partial input	Coarse completed	Dense completed	Ground truth
Abutment -1				
Abutment -2				

Figure 5. Completion results for abutments.

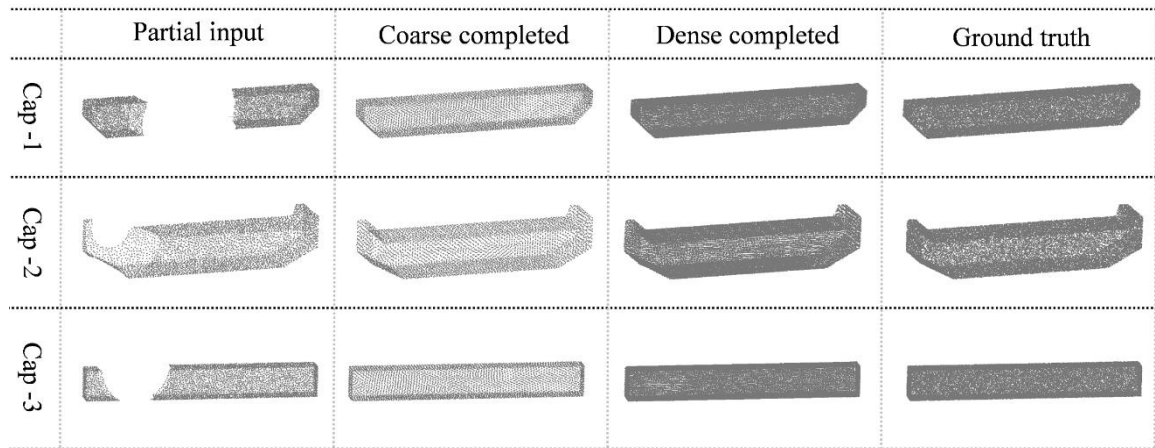


Figure 6. Completion results for pier caps.

Tables 1 and 2 provide the quantitative evaluation result for coarse and dense point cloud completion, respectively. Notably, the first row of each table compares the dense/coarse partial input with the corresponding dense/coarse ground truth, while the second row presents the calculation results for the dense/coarse ground truth and the recovered dense/coarse output from the partial input. The smaller CD values and large F-score values indicate better completion performance. For the F-score, different thresholds were set for different bridge components to effectively evaluate completion performance, as different bridge components have different scale sizes, resulting in different point cloud densities under a certain number of points. Accordingly, we set the thresholds of 0.05 m and 0.1 m for dense and coarse point cloud evaluation of piers and pier caps, while thresholds of 0.1 m and 0.2 m were applied for abutment completion evaluation.

These tables show that both results demonstrate a good performance in completing the bridge point cloud through PCN. The overall improvement of 11.94 CD and 11.05 F-score can be found for coarse completion, and 8.35 CD and 1.59 F-score can be found for dense results. Specifically, in the coarse output, improvements in CD of 16.97, 7.9, and 10.96 are achieved for abutment, pier, and pier cap, respectively, with corresponding F-score increases of 10.23, 7.94, and 14.97. For the dense output, CD reductions of 11.38, 4.47, and 9.22 are seen for abutment, pier, and pier cap, with F-score gains of 7.59, -8.09, and 5.29, respectively. The unsatisfactory result for pier completion may be attributed to the geometric loss in dense point cloud generation.

The detailed quantitative evaluation for point clouds with different scale occlusions is shown in Tables 3 and 4. Similarly, the PCN effectively mitigates occlusions in the bridge point cloud, as reflected by the improved evaluation metrics. However, only slight CD improvements or even unsatisfactory results are observed for point cloud completion with smaller occlusions, highlighted in red. This diminished performance may be due to accuracy loss during the encoder-decoder process. Given the small radius, ranging from 0 to $0.1L$, for occlusion generation, the completion for minor occlusions does not yield noticeable improvements in evaluation metrics. At the same time, the accuracy loss in the encoder-decoder would be avoidable to deteriorate the evaluation results. From the observation in the F-score, despite the poor performance in CD, the point clouds become more complete according to the increased F-score. In contrast, the model significantly improves evaluation metrics for point clouds with larger occlusions despite the presence of accuracy loss.

Table 1. Overall evaluation of coarse point cloud completion.

Dataset	CD ($\times 10^2$)				F-score (%)			
	Abutment	Pier	Pier cap	Avg.	Abutment	Pier	Pier cap	Avg.
Partial input	37.67	19.39	22.34	26.46	87.16	85.92	80.29	84.45
After completion	20.70	11.49	11.38	14.52	97.39	93.86	95.26	95.50

Table 2. Overall evaluation of dense point cloud completion.

Dataset	CD ($\times 10^2$)				F-score (%)			
	Abutment	Pier	Pier cap	Avg.	Abutment	Pier	Pier cap	Avg.
Partial input	24.78	12.18	15.32	17.42	85.86	87.49	81.26	84.87
After completion	13.40	7.71	6.10	9.07	93.45	79.40	86.55	86.46

Table 3. Evaluation of coarse point cloud completion under different occlusion conditions.

Dataset	CD ($\times 10^2$)				F-score (%)			
	Abutment	Pier	Pier cap	Avg.	Abutment	Pier	Pier cap	Avg.
Small occlusion	21.26	9.64	12.17	14.35	91.83	94.46	83.81	90.03
After completion	17.87	11.04	11.11	13.34	99.24	94.61	95.90	96.58
Large occlusion	54.68	30.38	35.11	40.05	82.33	76.29	75.87	78.16
After completion	23.63	12.00	11.72	15.78	95.46	93.02	94.44	94.30

Table 4. Evaluation of dense point cloud completion under different occlusion conditions.

Dataset	CD ($\times 10^2$)				F-score (%)			
	Abutment	Pier	Pier cap	Avg.	Abutment	Pier	Pier cap	Avg.
Small occlusion	7.41	2.21	4.67	4.76	90.39	96.31	84.97	90.55
After completion	10.39	7.37	5.77	7.84	96.12	79.91	87.97	88.00
Large occlusion	42.79	23.42	28.70	31.63	81.16	77.54	76.60	78.43
After completion	16.5	8.10	6.52	10.37	90.69	78.83	84.76	84.76

5. CONCLUSIONS

Creating a geometric and semantically rich BIM model is critical in the Scan-to-BIM, serving as the foundation for further development of bridge DT. However, generating an accurate geometric 3D parametric model from a point cloud is still a challenging task, particularly when the point cloud suffers from serious occlusion problems. Previous bridge reconstruction methods have inevitably developed corresponding strategies to mitigate the effects of point cloud occlusion and enhance reconstruction quality. The consideration of the occlusion problem in reconstruction algorithms can impact the overall design of algorithms, resulting in poor generalizability and low robustness in the reconstruction strategy. Addressing the point cloud occlusion can facilitate the overall reconstruction process, removing the limitation on the selection of reconstruction algorithms, enabling the successful application of algorithms and improving reconstruction accuracy.

Thus, this paper aims to advance the bridge gDT by introducing a new point cloud completion task within the current Scan-to-BIM framework, addressing point cloud occlusion as a distinct challenge. The task involves taking incomplete bridge point clouds with class labels as input and generating complete point clouds as output. To validate the proposed concept, we adopted the classical point cloud completion model PCN and validated the point cloud completion for bridge substructures, including abutments, piers, and pier caps. Both quantitative and qualitative results demonstrated a robust completion performance for point clouds with different scale occlusions.

The success of our study revealed the potential of addressing the point cloud occlusion problem after point cloud segmentation, relieving the pain in geometric reconstruction. Our study contributes to Scan-to-BIM by refining the Scan-to-BIM framework in bridge engineering. The new task point cloud completion can facilitate the bridge BIM reconstruction and the further development of DT systems. However, the occlusion in our study is generated through the spheres-based removal of point clouds, which cannot fully consider the different occlusion in real bridge data. Additionally, the application of point clouds for real-scanned bridge point clouds still needs further investigation. Thus, future work will explore the effects of different occlusion characteristics on bridge point cloud completion and the potential challenge when applying to real-scanned bridge point clouds. Training strategies for more complex bridge components, such as decks and girders, and completion performance for dense point clouds will be further explored in our future works.

ACKNOWLEDGMENTS

The authors would like to acknowledge the financial support from the University of Auckland and China Scholarship Council (Project No. 202206690016).

REFERENCES

- American Society of Civil Engineers ASCE. (2021). *America's Infrastructure Report Card 2021*.
- Bosché, F. (2010). Automated recognition of 3D CAD model objects in laser scans and calculation of as-built dimensions for dimensional compliance control in construction. *Advanced Engineering Informatics*, 24(1), 107–118. <https://doi.org/10.1016/j.aei.2009.08.006>
- Bosché, F., Guillemet, A., Turkan, Y., Haas, C. T., & Haas, R. (2014). Tracking the Built Status of MEP Works: Assessing the Value of a Scan-vs-BIM System. *Journal of Computing in Civil Engineering*, 28(4), 1–13. [https://doi.org/10.1061/\(asce\)cp.1943-5487.0000343](https://doi.org/10.1061/(asce)cp.1943-5487.0000343)
- Brilakis, I. (2024). *Digital Twinning the Built Environment*. December, 17–18. <https://doi.org/10.47330/cbc.2023.mhck5583>

- Edelsbrunner, H., & Mücke, E. P. (1994). Three-Dimensional Alpha Shapes. *ACM Transactions on Graphics (TOG)*, 13(1), 43–72. <https://doi.org/10.1145/174462.156635>
- Honghong, S., Gang, Y., Haijiang, L., Tian, Z., & Annan, J. (2023). Digital twin enhanced BIM to shape full life cycle digital transformation for bridge engineering. *Automation in Construction*, 147(July 2022), 104736. <https://doi.org/10.1016/j.autcon.2022.104736>
- Huang, Z., Yu, Y., Xu, J., Ni, F., & Le, X. (2020). PF-Net: Point fractal network for 3D point cloud completion. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 6(3), 7659–7667. <https://doi.org/10.1109/CVPR42600.2020.00768>
- Justo, A., Lamas, D., Sánchez-Rodríguez, A., Soilán, M., & Riveiro, B. (2023). Generating IFC-compliant models and structural graphs of truss bridges from dense point clouds. *Automation in Construction*, 149(January). <https://doi.org/10.1016/j.autcon.2023.104786>
- Khudhair, A., Li, H., Ren, G., & Liu, S. (2021). Towards future BIM technology innovations: A bibliometric analysis of the literature. *Applied Sciences (Switzerland)*, 11(3), 1–21. <https://doi.org/10.3390/app11031232>
- Liu, M., Sheng, L., Yang, S., Shao, J., & Hu, S. M. (2020). Morphing and sampling network for dense point cloud completion. *AAAI 2020 - 34th AAAI Conference on Artificial Intelligence*, 11596–11603. <https://doi.org/10.1609/aaai.v34i07.6827>
- Lu, R., & Brilakis, I. (2019). Digital twinning of existing reinforced concrete bridges from labelled point clusters. *Automation in Construction*, 105(May), 102837. <https://doi.org/10.1016/j.autcon.2019.102837>
- Lu, R., Brilakis, I., & Middleton, C. R. (2019). Detection of Structural Components in Point Clouds of Existing RC Bridges. *Computer-Aided Civil and Infrastructure Engineering*, 34(3), 191–212. <https://doi.org/10.1111/mice.12407>
- Mafipour, M. S., & Vilgertshofer, S. (2023). Automated geometric digital twinning of bridges from segmented point clouds by parametric prototype models. *Automation in Construction*, 156(May). <https://doi.org/10.1016/j.autcon.2023.105101>
- Mehranfar, M., Arefi, H., & Alidoost, F. (2021). Knowledge-based 3D reconstruction of bridge structures using UAV-based photogrammetric point cloud. *Journal of Applied Remote Sensing*, 15(04), 1–26. <https://doi.org/10.1117/1.jrs.15.044503>
- Pregnoiato, M., Gunner, S., Voyagaki, E., De Risi, R., Carhart, N., Gavriel, G., Tully, P., Tryfonas, T., Macdonald, J., & Taylor, C. (2022). Towards Civil Engineering 4.0: Concept, workflow and application of Digital Twins for existing infrastructure. *Automation in Construction*, 141(July), 104421. <https://doi.org/10.1016/j.autcon.2022.104421>
- Preparata, F. P., & Hong, S. J. (1977). Convex Hulls of Finite Sets of Points in Two and Three Dimensions. *Communications of the ACM*, 20(2), 87–93. <https://doi.org/10.1145/359423.359430>
- Schönfelder, P., Aziz, A., Faltin, B., & König, M. (2023). Automating the retrospective generation of As-is BIM models using machine learning. *Automation in Construction*, 152, 104937. <https://doi.org/10.1016/j.autcon.2023.104937>
- Tchapmi, L. P., Kosaraju, V., Rezatofighi, H., Reid, I., & Savarese, S. (2019). Topnet: Structural point cloud decoder. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2019-June*, 383–392. <https://doi.org/10.1109/CVPR.2019.00047>
- Walsh, S. B., Borello, D. J., Guldur, B., & Hajjar, J. F. (2013). Data processing of point clouds for object detection for structural engineering applications. *Computer-Aided Civil and Infrastructure Engineering*, 28(7), 495–508. <https://doi.org/10.1111/mice.12016>
- Wang, X., Ang, M. H., & Lee, G. H. (2022). Cascaded Refinement Network for Point Cloud Completion With Self-Supervision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11), 8139–8150. <https://doi.org/10.1109/TPAMI.2021.3108410>
- Wen, X., Li, T., Han, Z., & Liu, Y. S. (2020). Point Cloud Completion by Skip-Attention Network with Hierarchical Folding. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1936–1945. <https://doi.org/10.1109/CVPR42600.2020.00201>
- Yan, Y., & Hajjar, J. F. (2022). Geometric models from laser scanning data for superstructure components of steel girder bridges. *Automation in Construction*, 142. <https://doi.org/10.1016/j.autcon.2022.104484>
- Yang, T., Zou, Y., Yang, X., & del Rey Castillo, E. (2024). Domain knowledge-enhanced region growing framework for semantic segmentation of bridge point clouds. *Automation in Construction*, 165(November 2023), 105572. <https://doi.org/10.1016/j.autcon.2024.105572>
- Yi, L., Kim, V. G., Ceylan, D., Shen, I. C., Yan, M., Su, H., Lu, C., Huang, Q., Sheffer, A., & Guibas, L. (2016). A scalable active framework for region annotation in 3D shape collections. *ACM Transactions on Graphics*, 35(6), 1–12. <https://doi.org/10.1145/2980179.2980238>
- Yuan, W., Khot, T., Held, D., Mertz, C., & Hebert, M. (2018). PCN: Point completion network. *Proceedings - 2018 International Conference on 3D Vision, 3DV 2018*, 728–737. <https://doi.org/10.1109/3DV.2018.00088>

- Zhao, B., Chen, X., Hua, X., Xuan, W., & Lichti, D. D. (2023). Completing point clouds using structural constraints for large-scale points absence in 3D building reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 204(September), 163–183. <https://doi.org/10.1016/j.isprsjprs.2023.09.008>
- Zhuang, Z., Zhi, Z., Han, T., Chen, Y., Chen, J., Wang, C., Cheng, M., Zhang, X., Qin, N., & Ma, L. (2024). A Survey of Point Cloud Completion. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17, 1–22. <https://doi.org/10.1109/jstars.2024.3362476>

DEVELOPING DEFECT EMBEDDED MASONRY H-BIM USING DEEP LEARNING-BASED DETECTION AND SEGMENTATION

Xinyu TONG¹, Jingyu ZHOU², Cheng ZHANG^{3*}, Yiping DONG⁴

1) Ph.D. student, Department of Civil Engineering, Xi'an Jiaotong-liverpool University, China

2) MSc student, Department of Civil Engineering, Xi'an Jiaotong-liverpool University, China

3) Ph.D., Senior Associate Professor, Department of Civil Engineering, Design School Intelligent Built Environment Research Centre, Xi'an Jiaotong-liverpool University, China.

4) Ph.D., Associate Professor, Department of Architecture, Xi'an Jiaotong-liverpool University, China.

Abstract: The surface of ancient masonry structures is prone to various defects over time. H-BIM assists in defect inspection digitally, which improves efficiency and saves the labor force. However, existing H-BIM of masonry structures still has limitations of low defect complexity and ideal geometric shape, and the defect information is not integrated with corresponding masonry units, lacking accurate and comprehensive prediction in structural analysis. The developed model presents detailed and realistic masonry units fused with defect information and could be used for numerical simulations. A YOLOv8 model is used to detect and segment defects in masonry structures. K-fold cross-validation is employed during model training to mitigate the impact of category imbalance in the dataset. The YOLO model has also been employed to segment masonry units and extract their contours. The defect information is integrated with masonry units based on their positions. A case study was carried out in an ancient city wall in Suzhou, China. The generated masonry H-BIM assists the current and future protection of the structures, highlighting the feasibility of the method for the analysis of masonry structures.

Keywords: Deep learning, Defect detection, Defect segmentation, Masonry structure, Heritage building information modeling, Built heritage

1. INTRODUCTION

Masonry structures are a traditional construction form of built heritages that primarily utilizes materials such as bricks, stones, and mortar. Over time, these masonry units are susceptible to various types of defects due to natural or human-induced factors, including weathering or erosion, efflorescence, missing parts, and biological colonization, among others. These defects compromise the integrity of the building and may pose risks to its structural safety. Therefore, regular inspections of masonry structures are necessary to discover and address defects early, while preemptive measures need to be taken to mitigate risks in vulnerable positions. Conventional defect detection methods often rely on manual recording and annotation, lacking automation, which can be time-consuming and labor-intensive when dealing with hundreds or thousands of masonry units (Keshmiry et al., 2024). By utilizing deep learning-based artificial intelligence algorithms, various defects can be automatically detected. Another issue is that conventional risk assessment relies heavily on human experience, which may not always provide accurate and comprehensive foresight (Stepinac & Gašparović, 2020). Digital twin technology establishes a model reflecting real-world masonry structure and integrates defect information into masonry units, enabling digital management on heritage building information modeling (H-BIM) platforms, and facilitating the management and maintenance of those positions (Tan et al., 2021). The model can also be used for further numerical simulations to analyze the failure mechanisms, assess potential risks, and provide guidance for the current and future protection of the building.

2. LITERATURE REVIEW

2.1 Heritage Masonry Structural Models for Simulation

The performance of masonry structures is often analyzed through numerical simulation methods. Ferretti & Bažant (2006) evaluated the stability of ancient masonry towers in terms of size effect, moisture diffusion, and chemical reaction kinetics using the finite element method (FEM). Zhang et al. (2024) conducted thermal and moisture simulations on the ancient city wall of Nanjing, studying the impact of plants on the hygrothermal conditions of the wall. Li et al. (2021) simulated the effects of plants on the microclimate and surface crystalline weathering of the city wall. Loverdos et al. (2021) proposed an image-based approach for automatically generating numerical models of cracked masonry and established a numerical model based on the discrete element method. Pereira et al. (2023) studied the mechanical properties of masonry units under loading using numerical methods.

Existing numerical simulation modeling of masonry units mostly involve ideal geometric shapes, while some scholars have indicated that more representative masonry visualizations lead to more accurate results (Erdogmus et al., 2019; Loverdos et al., 2021). In addition, some existing research only models the morphology of masonry units and mortar joints, while information associated with the health status, such as defects, is not accurately reflected at corresponding locations, which could also affect the accuracy of the analysis.

2.2 Defect Mapping in H-BIM

The defect information is often integrated into the H-BIM for future management. Some existing research has explored the mapping process, typically through the conversion of planar photo coordinates, geographic coordinates, and 3D model coordinates (Brackenbury & Dejong, 2018; Liu et al., 2021; Tan et al., 2022; Chen et al., 2023; Pantoja-Rosero et al., 2023; Yang et al., 2024). However, existing research only simply overlays the defect positions on the model surface without deeper integration of defect information with the building components. For masonry structure models (Masonry-BIM), a more comprehensive approach could involve generating the arrangement of each masonry unit on the model and then assigning defect information to the corresponding masonry units according to the positions. This allows knowing the defect types present in each masonry unit on the H-BIM to precisely assess the risk of each unit. Furthermore, masonry units could be more precise than the entire wall surface in allocating defect attributes to them, enhancing the reliability of structural numerical simulation analysis.

2.3 Deep Learning-based Defect Detection and Segmentation

Deep learning has been employed for defect detection and segmentation of buildings to enhance inspection efficiency and accuracy. Some research utilizes computer vision techniques to detect objects in images. Huang et al. (2024) fused visible images with thermal images, developed a dataset of cracks in masonry structures, and evaluated the improvement in crack segmentation accuracy through six convolutional neural network (CNN)-based models. Wang et al. (2019, 2020) used a Faster R-CNN model based on ResNet to detect weathering and spalling in historic masonry structures and later utilized Mask-CNN to segment damage on yellow-glazed tiles. You Only Look Once (YOLO) is another model which has many applications in object detection and instance segmentation because it is adaptable to targets of various sizes and shapes with good detection speed and accuracy (Casas et al., 2024). Yan et al. (2024) employed the YOLO model to detect dislodged thin bricks on the roof pedestals of classical gardens. Li et al. (2023) used a YOLO model to identify cracks, chalking, plant damage, and ubiquinol on the surface of the Great Wall. Apart from image-based methods, there are also research based on point clouds. Valero et al. (2019) used a logistic regression classification algorithm to detect cracks, material loss, and discoloration on the facades of castles and churches.

A limitation of existing research is that the detected categories are often singular or limited, or the complexity of the surfaces is low, such as walls with uniform background colors or sparse defects. This limitation hinders deep learning models from achieving high accuracy when encountering dense defects or irregularly arranged cluttered masonry units.

2.4 Deep Learning-based Masonry Unit Segmentation

For the segmentation of masonry units, image-based or point cloud-based methods can be used. Image-based methods are relatively straightforward for regular-shaped masonry units. When masonry structures have random rubble and joints with irregular shapes and sizes, such as in ancient architectures, it will incur more training costs that need to classify and label each category of masonry unit with different sizes and materials multiple times (Wang et al., 2020). Valero et al. (2020) and Forster et al. (2023) employed a CloudCompare plugin based on the 2D Continuous Wavelet Transform (CWT) to segment masonry units and extract their contours from laser point clouds. This point cloud-based method is more suitable for irregular rather than flat wall facades (Güneş et al., 2024).

The objective of the present paper is to enhance the information integrity of masonry H-BIM, and investigate a solution of generating a lightweight masonry model that includes detailed and realistic masonry units mapped with defect information, enabling precise localization and assessment of potential risks to assist management personnel in protecting the building currently and in the future.

3. THE PROPOSED METHOD

To achieve the objective, the following methodology has been proposed, as shown in Figure 1. The procedure of the methodology includes data acquisition, masonry defect detection and segmentation, masonry unit segmentation, and integration of defect information with masonry H-BIM.

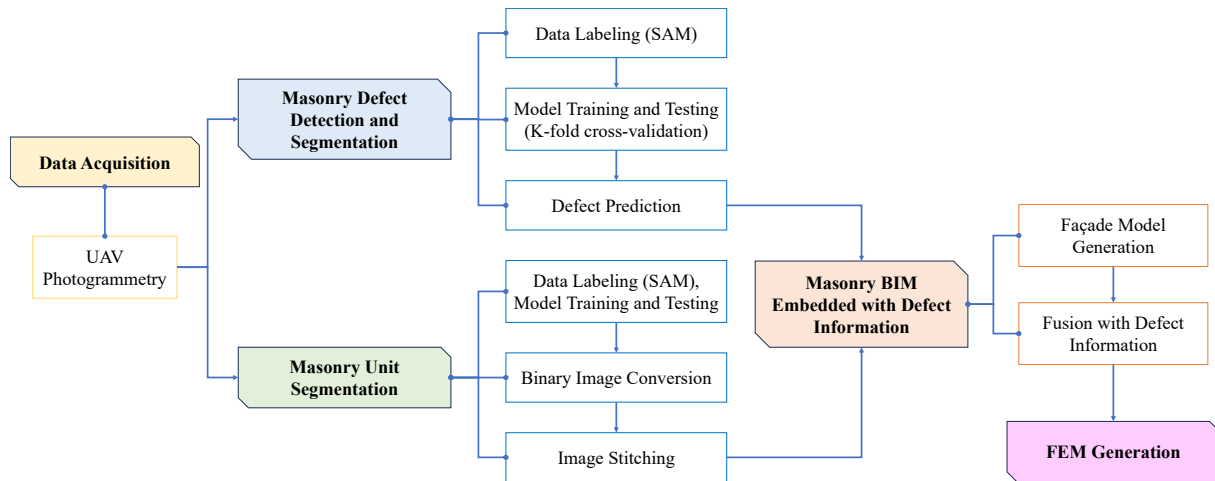


Figure 1. Proposed methodology

3.1 Masonry Defect Detection and Segmentation

A YOLOv8 model is developed and trained to detect and segment defects on masonry structures because of its good detection speed and accuracy on large-scale masonry structures. The process involves data acquisition, data labeling, model training, and model testing. Data acquisition involves capturing front views of the masonry facades using unmanned aerial vehicle (UAV) photogrammetry, followed by proportionally dividing the images into training, validation, and test sets. Images are cropped to a uniform pixel size to improve training effectiveness. During data labeling, four categories of defects in masonry structures are identified: erosion (further divided into deep erosion and superficial erosion based on the extent of damage), efflorescence, missing parts, and biological colonization. The Segment Anything (SAM) tool (Kirillov et al., 2023) is used for labeling to generate masks for the corresponding objects in the training images. Subsequently, the model is trained and tested for detection and segmentation. To mitigate the impact of category imbalance in the dataset, K-fold cross-validation is employed during training. This method involves dividing the original dataset into K subsets, iteratively using one subset as the validation set while the remaining K-1 subsets are used for training the model. The model is trained and validated multiple times, and the average of the results is used as the best weight. After multiple adjustments and optimizations of parameters such as initial learning rate, epochs, batch size, etc., the model file with the best weight is obtained for defect detection on the target masonry structures.

3.2 Masonry Unit Segmentation

Another YOLO model is developed for the segmentation of masonry units. Data labeling, model training, and model testing are also performed. In data labeling, the labeled objects became each masonry unit, which is the brick and stone block. Masks are added on the blocks in SAM. The model with the optimal weight after training is used for prediction. The images used in the prediction dataset in this step is the same as the prediction dataset used in 3.1, in order to maintain a consistent coordinate system in subsequent processing. The predicted images are converted with a black background and the masks are set to white to derive the binary images of the masonry structure, which clearly distinguishing between blocks and joints (white for blocks, black for joints), as shown in Figure 2. The binary images are then spliced to form an entire side of the wall.

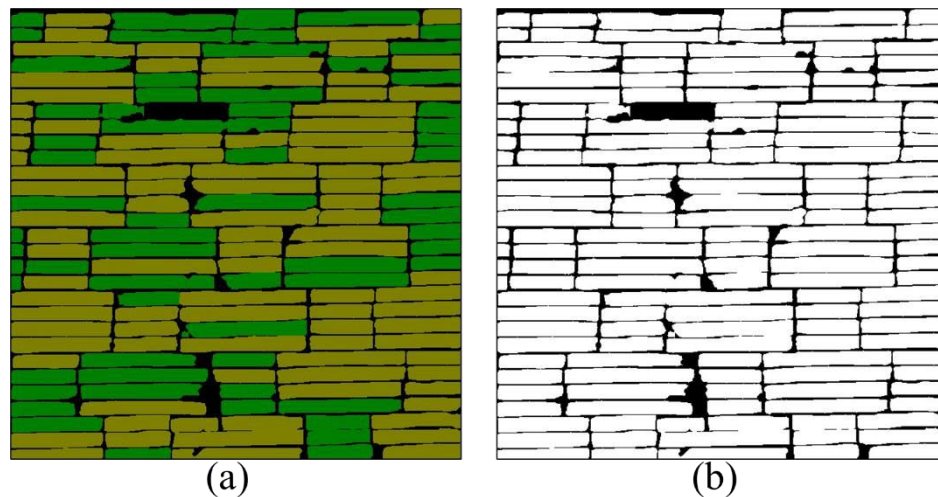


Figure 2. (a) The segmentation result of masonry units (different color represent different materials); (b) The concerted binary images of the masonry arrangement

3.3 Masonry H-BIM Generation and Defect Fusion

Images of the structure are aligned according to their position and the image point cloud is generated after alignment. The facades are segmented from the image point cloud and then using the "point to mesh" method, these parts are fitted into plane surfaces to generate lightweight facade models. Then, the binary images obtained in 3.2 are used for the arrangement of the masonry units on the facade models. Based on the black-and-white values, the blocks and joints can be automatically divided by parametric language and then projected onto the surface of the facade model.

After the generation of the façade model, the defect information is fused with the masonry units based on their positions. The coordinate system is set as shown in Figure 3. The coordinates of the defect masks in each predicted image are transformed into global coordinates, and masonry units at corresponding positions are assigned typical defect attributes. The final obtained model accurately reflects the shape and size of masonry units, also integrate relevant defect information at the corresponding locations.

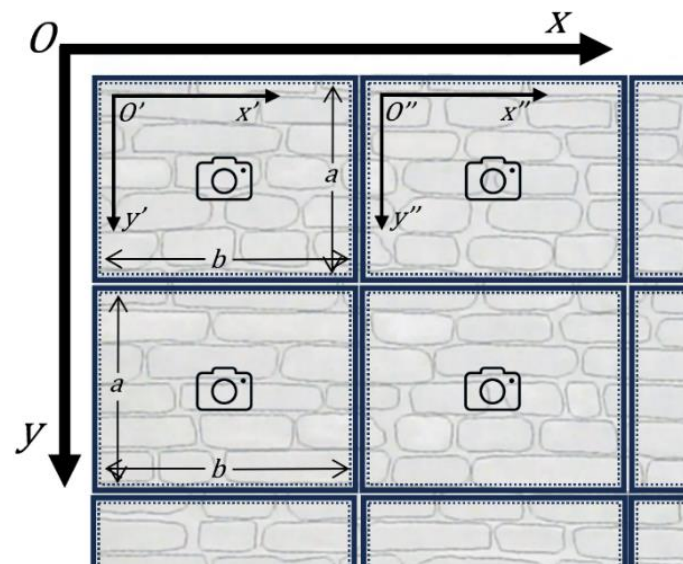


Figure 3. Set of the coordinate system

3.4 FEM Generation from the H-BIM

The obtained defect-integrated masonry model is then imported into FEM software for analysis. Material properties of bricks and joints are defined, and contact coefficients between elements such as friction coefficients are set. Different value of coefficients is assigned to each element based on the position of defect, such as Young's modulus, thermal and moisture conductivity coefficients, etc. Finally, the simulation is conducted under typical external environmental conditions.

4. PRELIMINARY RESULTS

The results involve a case study of an ancient city wall in Suzhou, China. The construction materials of the structure include bricks, stones, earth, mortar, and a small amount of wood. The stones are diverse in type, shape, and size. The ancient wall exhibits various defect categories such as erosion, efflorescence, missing parts, and biological colonization, with a high quantity and density. The complexity of this ancient wall poses significant research challenges, highlighting the feasibility of the methods proposed in this article for the analysis of masonry structures.

4.1 Data Preparation

The ancient walls were captured through UAV photogrammetry, as shown in Figure 4. The obtained images were used for YOLO model training and defect prediction. To prepare the YOLO dataset, images of three sides of the walls were used for model training and testing, while the last part of the wall was used for defect prediction and generating the H-BIM model of the masonry structure. During image acquisition, the camera needs to be kept as parallel to the wall as possible. For the images used for prediction, each image was removed distortion, and then all images were stitched to form a complete picture of the entire ancient wall surface. In the stitched complete picture, parts that do not belong to the wall, such as the ground and sky, were removed, as shown in Figure 5. This step was taken to facilitate the set of the wall coordinate system and match it with the point cloud coordinate system.



Figure 4. UAV photogrammetry at the ancient walls



Figure 5. Regions to be removed and to be corrected distortion in the stitched picture.

4.2 Masonry Defect Detection and Segmentation

500 uniformly cropped images were used for training. Five defect categories were defined for mask labeling, including *Deep erosion*, *Superficial erosion*, *Efflorescence*, *Missing part*, and *Biological colonization*. In the data training process, YOLOv8m-seg was selected as the pre-trained model, which offered moderate performance with acceptable accuracy without wasting resources. The dataset was divided into five subsets for K-fold cross-validation to mitigate the impact of sample category imbalance. The obtained training weight file best.pt was used for defect detection and segmentation on the other part of the ancient wall. Some prediction results are shown in Figure 6. In the images, red masks with the label "bio" stands for "*Biological colonization*", light blue masks with label the "deep" stands for "*Deep erosion*", white masks with the label "superficial" stands for "*Superficial erosion*", green masks with the label "salt" stands for "*Efflorescence*", and dark blue masks with the label "hole" stands for "*Missing part*". It can be observed that the trained model accurately detected and segmented the majority of defects. There were a few inaccurate detections that need further improvement. For example, some small plant leaves were not detected, some large areas of biological colonization or efflorescence could not be completely segmented, some bricks without defects were misidentified as erosion, and there were cases of repeated boxing in some locations.



Figure 6. Defect detection and segmentation results

4.3 Masonry Unit Segmentation and H-BIM Generation

The contours of the masonry units in the ancient wall were detected using the trained YOLO model and converted into binary images, as shown in Figure 7(a). The wall surface was extracted and the binary image was projected onto the wall surface, then blocks and joints were automatically generated through parametric language,

as shown in Figure 7(b). Subsequently, the defect information was assigned to the masonry units. Since image distortion had been corrected and non-wall parts had been removed in the previous step, the coordinate systems of the ancient wall surface in the images and the modeled wall surface from the point cloud were consistent. This allowed for the direct location of defect information to the corresponding masonry units on the model, as shown in Figure 7(c).

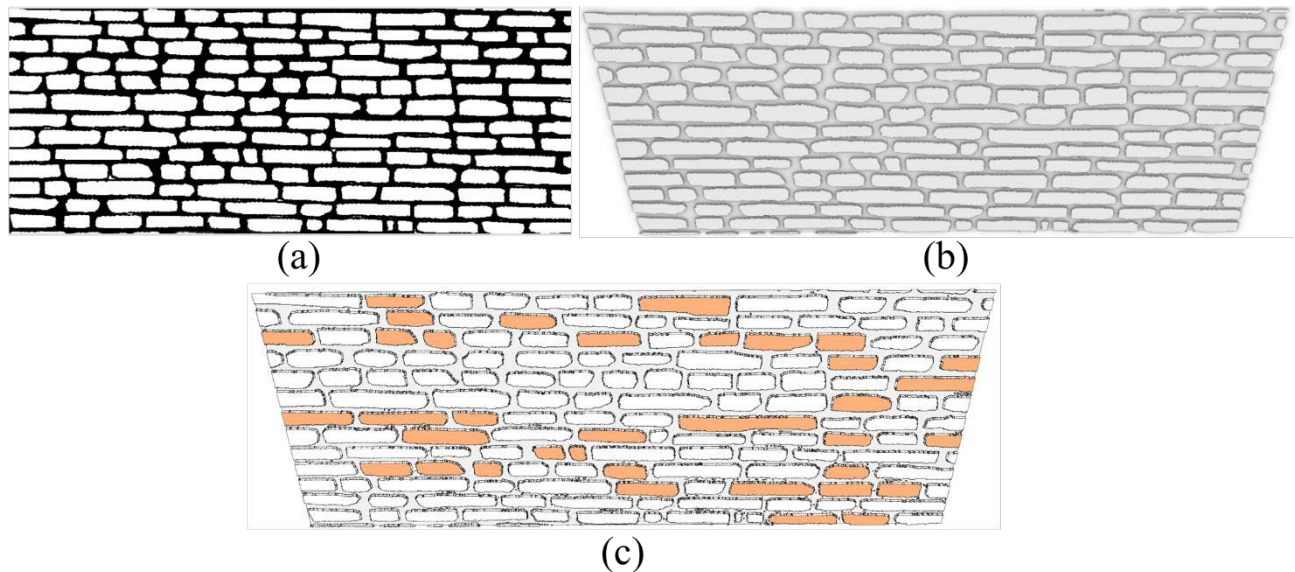


Figure 7. (a) Binary image of the entire wall; (b) Facade model; (d) Facade model with efflorescence information.

5. DISCUSSION

The results involve a case study of an ancient city wall in Suzhou, China. The proposed model presents detailed and realistic masonry units, fused with defect information. The YOLOv8 model is used to detect and segment defects on masonry structures. K-fold cross-validation is employed during model training to mitigate the impact of category imbalance in the dataset. The defect information is fused with the masonry units based on their positions. In future work, the masonry H-BIM model integrated with defect information will be imported into numerical simulation software for analysis. To ensure the reliability of the analysis, subsequent tasks will include investigating the properties of the filled earth inside the ancient wall, determining the loads on the wall and recording the surrounding environment, etc., in order to establish a comprehensive numerical simulation model. The complexity of this ancient city wall poses great research challenges, demonstrating the feasibility of the proposed method for masonry structure analysis.

ACKNOWLEDGMENTS

The authors would like to thank the support from the Design School Intelligent Built Environment Research Centre, Xi'an Jiaotong-Liverpool University.

REFERENCES

- Brackenbury, D., & Dejong, M. (2018). Mapping mortar joints in image textured 3D models to enable automatic damage detection of masonry arch bridges. In 17th international conference on computing in civil and building engineering (Vol. 173, pp. 530-45).
- Casas, E., Ramos, L., Romero, C., & Rivas-Echeverría, F. (2024). A comparative study of YOLOv5 and YOLOv8 for corrosion segmentation tasks in metal surfaces. *Array*, 100351.
- Chen, J., Lu, W., & Lou, J. (2023). Automatic concrete defect detection and reconstruction by aligning aerial images onto semantic - rich building information model. *Computer - Aided Civil and Infrastructure Engineering*, 38(8), 1079-1098.
- Erdogmus, E., Pulatsu, B., Can, B., & Ozkan, K. (2019, June). Analysis of the last standing arch of the Roman Aqueduct at Blaundos. In *Proceedings of the thirteenth North American masonry conference*.
- Ferretti, D., & Bazant, Z. P. (2006). Stability of ancient masonry towers: Moisture diffusion, carbonation and size effect. *Cement and Concrete Research*, 36(7), 1379-1388.
- Forster, A., Valero, E., Bosché, F., Hyslop, E., & Wilson, L. (2023). Digital Toolkit to Assist the Interpretation of Traditional Masonry Construction. *International Journal of Architectural Heritage*, 18(5), 725-739.

- Güneş, M. C., Mertan, A., Sahin, Y. H., Unal, G., & Özkar, M. (2024). Semi-automated minimization of brick-mortar segmentation errors in 3D historical wall reconstruction. *Automation in Construction*, 167, 105693.
- Huang, H., Cai, Y., Zhang, C., Lu, Y., Hammad, A., & Fan, L. (2024). Crack detection of masonry structure based on thermal and visible image fusion and semantic segmentation. *Automation in Construction*, 158, 105213.
- Keshmiry, A., Hassani, S., Dackermann, U., & Li, J. (2024). Assessment, repair, and retrofitting of masonry structures: A comprehensive review. *Construction and Building Materials*, 442, 137380.
- Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., ... & Girshick, R. (2023). Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 4015-4026).
- Li, Q., Zheng, L., Chen, Y., Yan, L., Li, Y., & Zhao, J. (2023). Non-destructive testing research on the surface damage faced by the Shanhaiguan Great Wall based on machine learning. *Frontiers in Earth Science*, 11, 1225585.
- Li, Y., Xia, C., Wu, R., Ma, Y., Mu, B., Wang, T., ... & Hokoi, S. (2021). Role of the urban plant environment in the sustainable protection of an ancient city wall. *Building and Environment*, 187, 107405.
- Liu, D., Xia, X., Chen, J., & Li, S. (2021). Integrating building information model and augmented reality for drone-based building inspection. *Journal of Computing in Civil Engineering*, 35(2), 04020073.
- Loverdos, D., Sarhosis, V., Adamopoulos, E., & Drougkas, A. (2021). An innovative image processing-based framework for the numerical modelling of cracked masonry structures. *Automation in Construction*, 125, 103633.
- Pantoja-Rosero, B. G., Achanta, R., & Beyer, K. (2023). Damage-augmented digital twins towards the automated inspection of buildings. *Automation in Construction*, 150, 104842.
- Pereira, M., D'Altri, A. M., de Miranda, S., & Glisic, B. (2023). Automatic multi-leaf nonperiodic block-by-block pattern generation and computational analysis of historical masonry structures. *Engineering Structures*, 283, 115945.
- Stepinac, M., & Gašparović, M. (2020). A review of emerging technologies for an assessment of safety and seismic vulnerability and damage detection of existing masonry structures. *Applied Sciences*, 10(15), 5060.
- Tan, Y., Li, G., Cai, R., Ma, J., & Wang, M. (2022). Mapping and modelling defect data from UAV captured images to BIM for building external wall inspection. *Automation in Construction*, 139, 104284.
- Tan, Y., Li, S., Liu, H., Chen, P., & Zhou, Z. (2021). Automatic inspection data collection of building surface based on BIM and UAV. *Automation in Construction*, 131, 103881.
- Valero, E., Bosché, F. (2020). Masonry segmentation plugin for CloudCompare, [software]. Cyberbuild Lab. The University of Edinburgh. doi:10.7488/ds/2892.
- Valero, E., Forster, A., Bosché, F., Hyslop, E., Wilson, L., & Turmel, A. (2019). Automated defect detection and classification in ashlar masonry walls using machine learning. *Automation in construction*, 106, 102846.
- Wang, N., Zhao, X., Zhao, P., Zhang, Y., Zou, Z., & Ou, J. (2019). Automatic damage detection of historic masonry buildings based on mobile deep learning. *Automation in Construction*, 103, 53-66.
- Wang, N., Zhao, X., Zou, Z., Zhao, P., & Qi, F. (2020). Autonomous damage segmentation and measurement of glazed tiles in historic buildings via deep learning. *Computer - Aided Civil and Infrastructure Engineering*, 35(3), 277-291.
- Yan, L., Chen, Y., Zheng, L., & Zhang, Y. (2024). Application of computer vision technology in surface damage detection and analysis of shedthin tiles in China: a case study of the classical gardens of Suzhou. *Heritage Science*, 12(1), 72.
- Yang, L., Liu, K., Ou, R., Qian, P., Wu, Y., Tian, Z., ... & Yang, F. (2024). Surface Defect-Extended BIM Generation Leveraging UAV Images and Deep Learning. *Sensors*, 24(13), 4151.
- Zhang, T., Xia, C., Ma, Y., Zhang, R., Han, S., Kong, Z., ... & Li, Y. (2024). Effects of tree density variations on outdoor heritage conservation: Numerical study of an ancient brick city wall with four orientations. *Building and Environment*, 254, 111392.

A Digital Twin-Driven Approach for Detecting Overbreak and Underbreak in Drill-and-Blast Tunnels

Zhiyang Pan¹, Yichong Zhang², Yujie Lu³, Bing Tang⁴

1) B.Eng. Candidate, SWJTU-Leeds joint school, Southwest Jiaotong University and University of Leeds, Chengdu, China.
Email: cn21zp@leeds.ac.uk

2) Ph.D. Candidate, Shanghai Research Institute for Intelligent Autonomous Systems, Tongji University, Shanghai, China
Email: zhangyichong@tongji.edu.cn

3) Ph.D., Prof., Department of Building Engineering, College of Civil Engineering, Tongji University, Shanghai, China.
Email: lu6@tongji.edu.cn

4) Senior Engineer, Department of Engineering, Zhejiang Quzhou-Lishui Railway Co., Ltd., Zhejiang, China.
657774378@qq.com

Abstract: This study presents a computational workflow for detecting overbreak and underbreak in drill-and-blast tunnels, grounded in the principles of Digital Twin technology, with the objective of enhancing blasting quality control during tunnel construction. Utilizing 3D laser scanning technology, a 3D model of the real-world tunnel is constructed and integrated with the design model for comparative analysis. The differences in the contours of both models are computed, facilitating the automated assessment of overbreak and underbreak volumes in blasting sections. The specific workflow encompasses data acquisition, point cloud processing, model construction and optimization, as well as model integration and analysis, thereby establishing an efficient and precise system for detecting overbreak and underbreak. In comparison to traditional manual section measurement methods, the proposed approach not only significantly reduces labor workload but also substantially enhances detection accuracy. This technology offers reliable technical support for tunnel blasting quality assessment, effectively addressing the challenges of high labor input in drill-and-blast tunnel construction.

Keywords: Tunnel, Digital Twin, 3D Reconstruction, Quality Control.

1. INTRODUCTION

Drill-and-blast excavation is the most widely employed method in mountain tunnel construction, owing to its high adaptability and cost-effectiveness, especially in the construction of long-distance tunnels in complex geological conditions. However, mountain tunnel projects often involve the simultaneous construction of multiple tunnels over extensive distances, presenting considerable challenges for the efficient management of tunnel groups. Existing quality control technologies for drill-and-blast tunnel construction often fail to meet the required standards of efficiency and precision. Coupled with the uncertainty of geological conditions and the subjectivity of manual operations, this frequently results in overbreak and underbreak.

Overbreak and underbreak refer to the deviations between the actual excavation profile and the design tunnel profile. Overbreak occurs when the excavation profile exceeds the design boundary, whereas underbreak occurs when the excavation profile falls short of the design boundary. The primary causes of these deviations include conservative blasting parameters, unpredictable geological conditions, and variability in manual operations. Blasting designs often employ conservative parameters to ensure safety, limiting the ability to control excavation boundaries. Moreover, the drilling location and explosive charge placement in drill-and-blast methods are highly dependent on the operators' subjective judgment, with variations in their technical skill and experience impacting the consistency of the results. Furthermore, the inherent complexity of the rock mass makes it more difficult to control blasting effects, resulting in overbreak and underbreak as common issues in drill-and-blast tunnel construction.

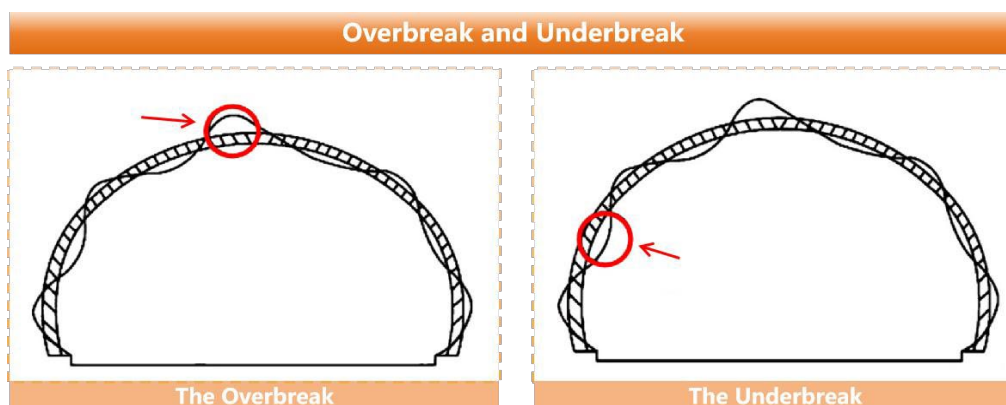


Figure 1. Schematic diagram of the overbreak and underbreak

The consequences of overbreak and underbreak are complex and multifaceted. Overbreak increases the volume of broken rock, thereby increasing costs for rock removal and waste disposal. Underbreak, in contrast, typically requires manual rock correction, which delays subsequent blasting cycles and, in turn, prolongs the project timeline. Traditional analysis of overbreak and underbreak depends on the manual selection of multiple tunnel sections for individual measurements, a time-consuming process susceptible to human error, thereby further extending the construction period. Therefore, an efficient and accurate detection method for overbreak and underbreak is essential for enhancing the quality and efficiency of drill-and-blast tunnel construction.

To address this challenge, this study proposes a detection method for overbreak and underbreak based on Digital Twin technology. Digital Twin technology facilitates the creation of a digital representation of a physical entity, enabling computational analysis in a virtual environment. In this study, 3D laser scanning technology is utilized to acquire the actual 3D model of the tunnel, which is subsequently compared with the design model. By analyzing the contour differences between these two models, the volumes of overbreak and underbreak are computed. Compared to traditional manual measurement methods, this approach offers superior efficiency and accuracy, significantly improving quality management in tunnel construction and providing an effective solution for the intelligent management of tunnel construction quality.

2. RELATED WORK

2.1 The Influence of Overbreak and Underbreak

In mountain tunnel construction, the drill-and-blast method is one of the most widely used excavation techniques, and the evaluation of construction quality is closely tied to the levels of overbreak and underbreak. These deviations directly affect construction progress and lead to increased support and maintenance costs. Specifically in drill-and-blast tunneling, due to the complexity of geological conditions and the uncontrollable nature of construction parameters, the cost of addressing underbreak is typically higher than that of overbreak. As a result, the principle of “better over than under” is often adhered to, making overbreak more common than underbreak in tunnel construction (Koopialipoor et al., 2019). Overbreak and underbreak primarily arise from geological conditions, deviations between design and actual conditions, and human influence during operations (Van Eldert, 2017; Verma, 2018). Overbreak and underbreak caused by geological conditions can be partially controlled by improving preliminary investigation methods and optimizing blast designs, while deviations arising from design or execution issues can be mitigated by adjusting blast parameters (Singh, S.P. & Xavier, P., 2005). Overbreak and underbreak have significant adverse effects on the overall stability of tunnels, potentially leading to rock instability, water leakage, and increased support costs. In recent years, an increasing body of research has focused on optimizing drill-and-blast parameters and evaluating the impact of overbreak and underbreak on tunnel construction. For instance, Lei Mingfeng et al. (2023) proposed an instance segmentation algorithm for muck block size based on deep learning, which has shown promising engineering applications in optimizing blast parameters for mountain tunnels. Additionally, Ding Xiang (2022), using the Jian Mountain Tunnel project on the Zhonglan Railway as a case study, employed discrete element numerical simulation to examine the effects of interlayer dip angles on blast-induced overbreak and underbreak, resulting in optimized blast design solutions.

2.2 Measurement Methods and Technological Advances in Overbreak and Underbreak

With the continuous development of tunnel overbreak and underbreak measurement technologies, two primary approaches are currently employed: the traditional total station method and digital model analysis (Kim, Y. & Bruland, A., 2019). Since the 1980s, total stations have been widely used in tunnel engineering; however, this method requires substantial manual measurement, making it time-consuming and prone to significant errors and inefficiencies, which negatively impact project costs. Consequently, total stations have been progressively replaced by more efficient three-dimensional (3D) model-based methods (Alhaddad, M., 2016; Huang, Y. et al., 2022). The 3D model analysis approach primarily includes close-range photogrammetry and 3D laser scanning technologies, both of which assess overbreak and underbreak by collecting 3D data and comparing it with the design drawings (Wei, Z., 2023). Close-range photogrammetry involves the use of professional optical cameras to capture images, and recent technological advances have facilitated faster data transmission and processing (Zhaoyang Jin et al., 2021). Meanwhile, 3D laser scanning, an emerging technology, utilizes laser ranging to collect 3D coordinates (Haishan Zhu & Sheng Li, 2021). The data is then processed and compared with the design drawings to assess levels of overbreak and underbreak. Recent studies indicate that overbreak and underbreak detection technology based on 3D laser scanning can automatically identify the quantity, location, and volume of overbreak and underbreak (Xuan Xie et al., 2024) and offers significantly higher accuracy than traditional total station or close-range photogrammetry methods (Wei, Z., 2023). Additionally, some studies have proposed analytical methods based on fitting point cloud surfaces to design surface normals to accurately

calculate tunnel overbreak and underbreak volumes (Fang et al., 2024).

2.3 Prediction and Optimization of Tunnel overbreak and underbreak

In addition to existing measurement technologies, the prediction and optimization of overbreak and underbreak using artificial intelligence and machine learning methods have become key research focuses in recent years. For instance, Biao He et al. (2023) developed a hybrid model based on random forests to predict overbreak volume, which has been validated in multiple engineering projects. Amin Hekmatnejad et al. (2024) proposed a “Universal Discontinuity Index” (UDI) to predict the geometric characteristics of tunnel overbreak and underbreak. Furthermore, G.M. Foderà et al. (2020), in their study of the Brenner Base Tunnel (BBT), developed an operational method to estimate overbreak volume, distinguishing between technical overbreak caused by deficiencies in drill-and-blast design and execution, and geological overbreak influenced by rock mass characteristics.

These predictive models and algorithms leverage big data and deep learning techniques to forecast tunnel overbreak and underbreak, providing robust support for the optimization of tunnel blast parameters. However, due to the unpredictability of geological conditions, subjective construction practices, and inconsistent adherence to construction specifications, overbreak and underbreak in tunnels remain difficult to fully mitigate. These factors pose significant challenges to the accurate prediction of overbreak and underbreak. Current machine learning and deep learning methods primarily base predictions on numerical data and distribution patterns of overbreak and underbreak, which fail to account for all influencing factors. Therefore, the rapid detection and localization of overbreak and underbreak after each blast cycle remain crucial for on-site construction management. Real-time data on tunnel overbreak and underbreak plays a vital role in designing appropriate corrective and repair strategies.

3. METHODOLOGY

The fusion analysis of the design model and point cloud model near the tunnel face facilitates the accurate calculation of overbreak and underbreak. This study employs the concept of Digital Twin to establish a mapping between physical space and digital space. The corresponding design model and reality-based model in the digital space are derived from the physical space, as shown in the figure 2.

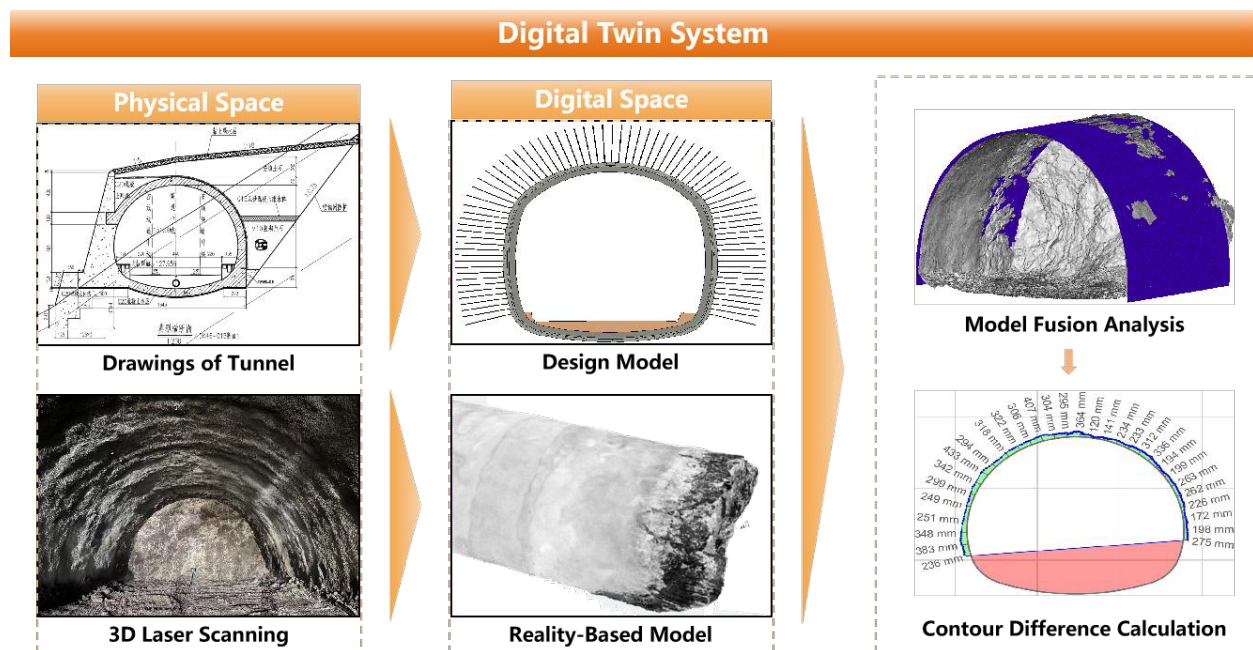


Figure 2. Construction of the Digital Twin System

3.1 Data Collection and Point Cloud Processing

This study primarily employs three-dimensional laser scanning technology. 3D laser scanning is a non-contact measurement technique based on laser ranging. It functions by emitting laser beams and receiving reflected signals, accurately measuring the spatial coordinates of the target object. Using high-density point cloud data, 3D laser scanning technology can generate detailed digital models of real-world objects and is widely applied in fields such as engineering surveying, building inspection, and tunnel engineering. The Faro Focus 350 employed in this study is a high-precision, stationary 3D laser scanner, specifically designed for rapid and

efficient scanning of large and complex environments. This device has a scanning range of up to 300 meters and a measurement speed of 976,000 points per second, enabling the rapid generation of high-resolution point clouds. Additionally, it offers exceptional measurement accuracy, with errors as small as one millimeter, making it ideal for engineering scenarios requiring high precision, such as tunnels.

During data collection, the first step involves selecting appropriate scanning locations on the tunnel construction site to ensure complete coverage of the tunnel's interior profile. Next, scanning parameters such as resolution, scanning angle, and range are adjusted to meet the precision requirements. For overbreak and underbreak detection in tunnels, higher resolution is typically selected to capture detailed tunnel profile information. After scanning, the point clouds are aligned and stitched together between scan stations, followed by refined processing, such as denoising, filtering, and downsampling, to generate a reality-based model of the tunnel. By integrating the design model with the reality-based model, digital twin technology maps the physical and digital spaces, enabling accurate calculations of overbreak and underbreak volumes. This series of steps and technological applications ensures that overbreak and underbreak detection in drill-and-blast tunnel construction is precise and efficient, providing strong support for quality control in tunnel engineering.

3.2 Establishment of Reality-Based Model By Using 3D Laser Scanning Technology

The establishment of a reality-based model using 3D laser scanning technology is a critical step in the entire process. Its core lies in the collection, processing, and reconstruction of high-precision point clouds to achieve detailed modeling and analysis of tunnel structures. This process involves several steps, including preliminary processing and refinement of point cloud data, 3D reconstruction, and the final generation of the reality-based model.

Once point cloud data are collected, the first step is to process the raw data. Due to the complexity of the field measurement environment, point cloud data may contain significant noise, redundant data, and local anomalies. The primary task of point cloud processing is to enhance data clarity, reduce noise interference, and optimize the overall quality of the point cloud data.

Denoising is the first step in point cloud data processing. In this study, the Radius Outlier Removal (ROR) method is employed. By setting a fixed radius, any point with fewer neighboring points within this radius than a preset threshold is considered a noise point. This method effectively removes outliers, ensuring the continuity and accuracy of the point cloud.

To optimize point cloud resolution and reduce redundant data, the Octree method is employed for downsampling. This method divides the Cartesian coordinate system into eight quadrants, layering the original point cloud. Representative points are selected from each layer to replace multiple points within that region, thus reducing the data volume while preserving the geometric features of the point cloud. This processing efficiently encodes nodes and enhances the processing efficiency and decoding speed of the point cloud data.

After point cloud processing, the next step is 3D reconstruction using the point cloud. A commonly used method in 3D reconstruction is triangulation, where the point cloud is transformed into a polygonal mesh model that represents the surface geometry of the target. Poisson surface reconstruction uses normal vector information from the point cloud to solve Poisson equations and generate a continuous, smooth surface, which is particularly useful for point cloud data with noise or local gaps.

Once the initial mesh is generated through triangulation, surface fitting and mesh optimization are performed to improve the smoothness and accuracy of the model. Surface fitting can be achieved using mathematical methods such as NURBS (Non-Uniform Rational B-Splines) or B-spline surfaces, which smooth complex surfaces and eliminate sharp surface irregularities. Mesh optimization adjusts the vertices and edges of the mesh to ensure the generated model maintains geometric accuracy while achieving higher mesh quality.

Through point cloud processing and 3D reconstruction, the reality-based tunnel model is generated (Fig. 3). This model is not only a digital replica of the physical world but also serves as the foundation for further analysis. In overbreak and underbreak detection, comparing the geometric profiles of the reality-based and design models is a key step. By overlaying the reality-based model with the design model, accurate calculations of overbreak and underbreak volumes in the surrounding rock can be made. This 3D laser scanning-based analysis method significantly enhances detection precision and efficiency, providing reliable technical support for tunnel construction quality control.

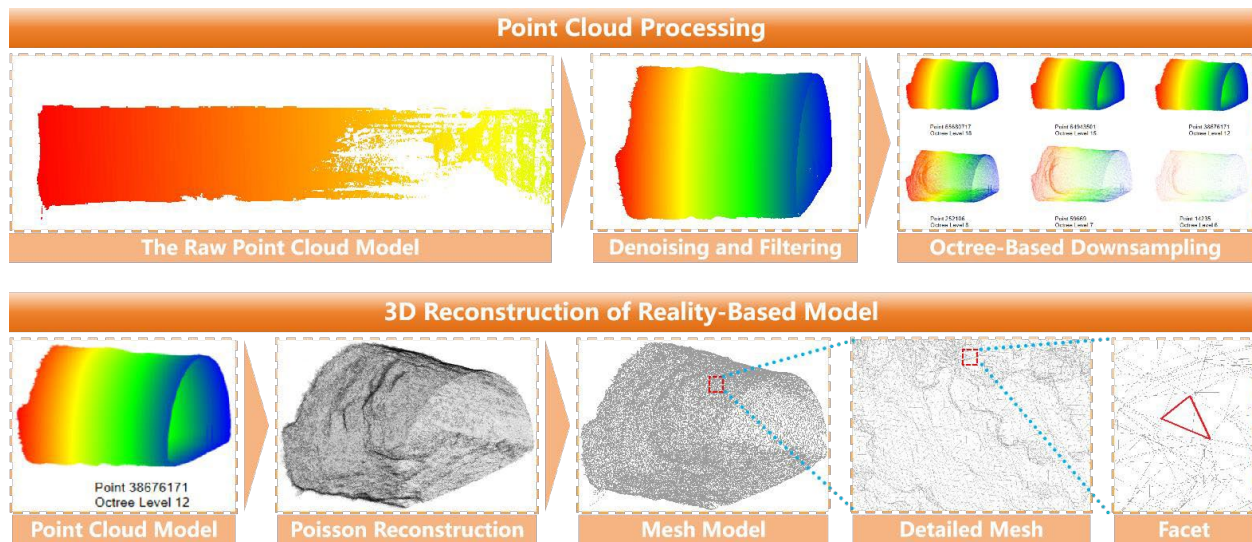


Figure 3. Tunnel Reality-Based Model Construction Process

3.3 Calculation Method for Overbreak and Underbreak

Since the design model is based on drawings, its coordinate system typically follows a standard geodetic coordinate system, while the reality-based model, generated through 3D laser scanning technology, uses an internal coordinate system. To achieve accurate overbreak and underbreak calculations in the tunnel, it is essential to align these two models within the same coordinate system. This ensures geometric consistency and relative accuracy between the models. Therefore, a coordinate system datum transformation is a crucial step before performing overbreak and underbreak calculations.

The Bursa seven-parameter coordinate transformation model can be used to align the two coordinate systems. This model, based on affine transformation, uses seven parameters - *translation along three axes*, *rotation*, and a *scale factor* - to achieve precise conversion from the internal to the geodetic coordinate system. The core of the Bursa model involves calculating transformation parameters using known control points that can be accurately identified in both coordinate systems. In this study, key points are extracted from the point cloud model by identifying critical nodes such as the tunnel crown, invert, and sidewalls. The known coordinate points are then obtained and matched to the design drawings. Using optimization methods such as the least squares method, the parameters are calculated, and the coordinate transformation subsequently maps the reality-based model to the geodetic coordinate system.

Once the datum transformation is complete, both the design and reality-based models will share the same coordinate system, ensuring geometric consistency in space. This forms the foundation for subsequent overbreak and underbreak calculations. Using this coordinate transformation method, the reality-based tunnel model is accurately aligned with the design model, enabling precise identification of overbreak and underbreak regions. This enables the construction and detailed analysis of the digital twin model.

3.3.1 Calculation of Overbreak and Underbreak Distances

The calculation of overbreak and underbreak volumes is fundamentally based on the geometric comparison between the reality-based and design models. Both models are represented as 3D mesh structures composed of multiple triangular facets, where each facet is defined by three vertices, and each with unique spatial coordinates. Consequently, the calculation process involves determining the distance from the vertices of the reality-based model's triangular facets to the corresponding facets of the design model, as shown in Figure 4. The following outlines the procedure for calculating overbreak and underbreak distances:

(1) **Extraction and Sorting of Triangle Facets in the Design Model:** The triangular facets in the design model are extracted and sorted into a sequence within a triangular facet group. One triangle facet is selected from this group at a time for subsequent calculations.

(2) **Check the Vertex Projection:** For each selected triangular facet A in the design model, corresponding vertices $P_i (x_i, y_i, z_i)$ in the reality-based model are sequentially searched. For a vertex in the reality-based model, it is necessary to determine whether the vertex can project onto the triangular facet A of the design model. If the projection of the vertex falls inside the triangular facet A, the vertex is recorded in the array $[P_1, P_2, P_3, \dots, P_n]$.

(3) **Point-to-Facet Distance Calculation:** If the projection of the vertex P_i falls inside the triangular facet A of the design model, the distance from the vertex P_i to the triangular facet A is calculated. The distance D between a point $P(x_p, y_p, z_p)$ and a triangle plane can be calculated using the point-to-plane distance formula. The

resulting distance D_i is stored in the corresponding distance array for that triangular facet A.

(4) **Retrieving Other Vertices and Updating Distance Groups:** The next vertex in the reality-based model is retrieved, and the projection check and distance calculation process is repeated. If the vertex can be projected onto the triangular facet of the design model, its distance to the facet is calculated and stored in the corresponding distance array for that triangular facet, until all vertices have been checked. If there are n valid vertices within the facet A, its corresponding distance group is created as $[D_1, D_2, D_3, \dots, D_n]$.

(5) **Calculation of Overbreak and Underbreak for Each Triangle Facet:** For each triangular facet in the design model, the final overbreak and underbreak distance is determined by the minimum value in its corresponding distance array, denote as $D_{\min} = \min(D_1, D_2, D_3, \dots, D_n)$. The D_{\min} represents the overbreak or underbreak for the triangular facet, the symbols "+" and "-" represent overbreak and underbreak, respectively, reflecting the geometric deviation between the reality-based model and the design model in the corresponding region.

(6) **Global Overbreak and Underbreak Distance Calculation:** The above process is repeated for all triangular facets in the design model, sequentially retrieving and calculating distances, until the overbreak and underbreak volumes for all facets are determined. Then each triangular facet in the design model corresponds to an overbreak or underbreak distance. By aggregating the overbreak and underbreak distances from all facets, a comprehensive overbreak and underbreak analysis for the entire tunnel can be conducted.

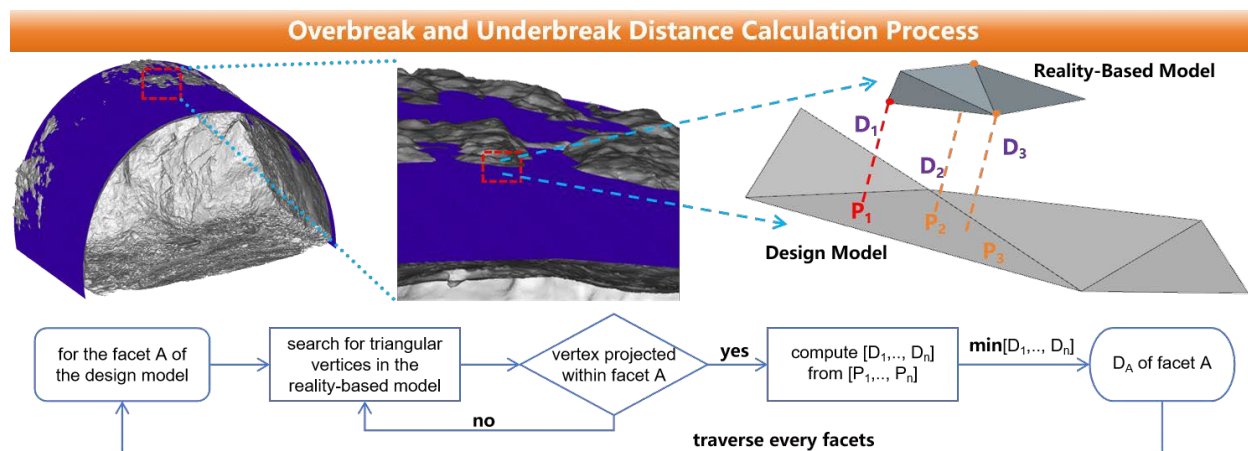


Figure 4. Overbreak and Underbreak Distance Calculation

3.3.2 Area and Volum Calculation of Overbreak and Underbreak

In drill-and-blast tunnel engineering, the overbreak area serves as a critical indicator for assessing the geometric deviation of tunnel cross-sections. The calculation of the overbreak area for a specific cross-section involves introducing a sliding plane, perpendicular to the longitudinal axis of the tunnel, which intersects both the reality-based model and the design model, as shown in Figure 5. The overbreak area for a given cross-section is computed through the following steps:

(1) **Introduction of the Sliding Plane:** The sliding plane is defined as the chosen tunnel cross-section. The position of the reference plane is established according to the longitudinal coordinates of the tunnel.

(2) **When the sliding plane intersects the vertices:** If the sliding plane intersects the vertices of the triangular facets in the reality-based model, the overbreak distance can be calculated based on the results presented in section 3.3.1.

(3) **When the sliding plane intersects the facet:** If one or two vertices of the triangular facet lie on opposite sides of the sliding plane, the sliding plane intersects the facet. Given the equation of the sliding plane and its perpendicular orientation to the longitudinal axis, the intersection points between the plane and the triangular facet occur at the same longitudinal position. Using the line equation and the coordinates of the intersection points, the coordinates of intersections $A(x_a, y_a, z_a)$ and $B(x_b, y_b, z_b)$ between the sliding plane and the two edges of the triangular facet can be computed.

(4) **Elimination of Redundant Intersection Points:** Since triangular facets in the real-world model are adjacent, one edge of an intersected triangular facet may be counted twice. Sorting the intersection points based on their polar coordinate angles removes redundant points, resulting in a unique array of intersection points between the reference plane and the real-world model.

(5) **Selection of Triangular Facets from the Design Model:** After obtaining the array of intersection points for the real-world model, the triangular facets from the design model that intersect with the reference plane are identified. This results in a set of triangular facets from the design model.

(6) **Distance and Area Calculation:** The distances L between adjacent intersection points in the array

are measured. Using the results from section 3.3.1, the distances D between the intersection points and the triangular facets of the design model are computed. By integrating these distances, the overbreak and underbreak area for the sliding plane's cross-section is computed.

(7) **Calculation of Overbreak and Underbreak Volume Along Steps:** By shifting the sliding plane along the longitudinal axis of the tunnel, the overbreak and underbreak areas for each cross-section are determined. The total overbreak and underbreak volume for the tunnel blast section is obtained by integrating the overbreak areas as the sliding plane moves, thereby completing the entire calculation process.

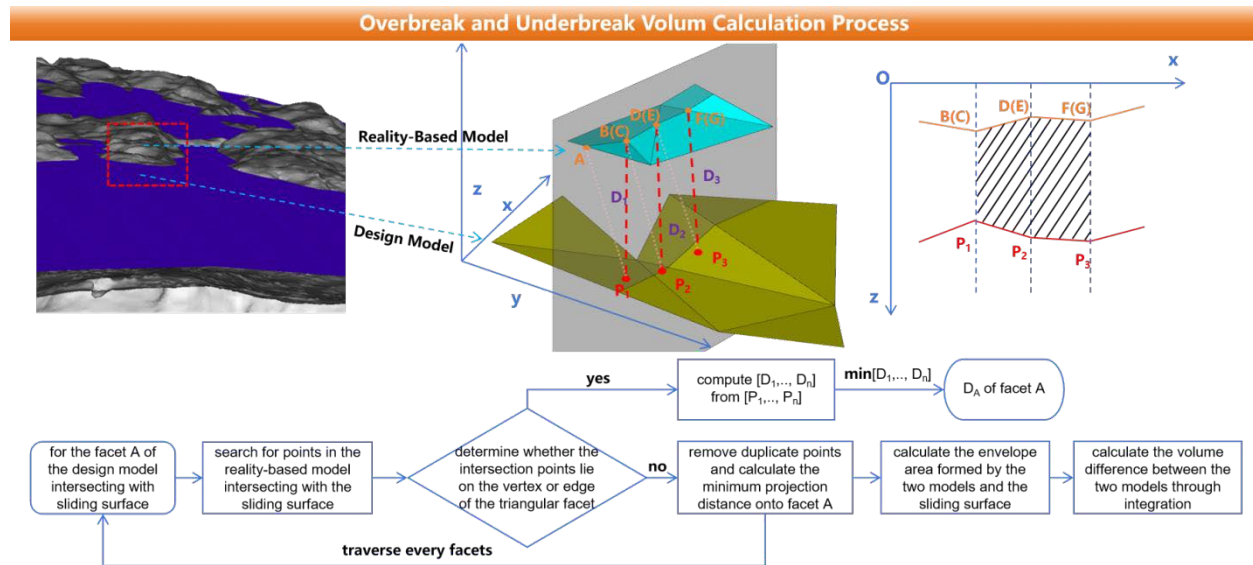


Figure 5. Overbreak and Underbreak Volum Calculation

Through the aforementioned calculation process, the overbreak distance is first calculated based on the geometric relationship between the reality-based model and the design model. Next, the overbreak area for each cross-section is determined using the polygonal method, and the total overbreak volume for the tunnel is obtained through numerical integration. This series of methods ensures the accuracy and operability of model comparison and analysis, effectively quantifying the volume of the overbreak area and providing a theoretical basis for refined management and construction optimization in tunnel engineering.

4. RESULT

To enhance the visualization of the calculation results, this study divides the tunnel cross-sections by angle and assigns a unique identifier to each cross-section. By calculating the overbreak and underbreak areas at each position along the cross-section, the overbreak and underbreak for the entire cross-section can be visualized, facilitating statistical analysis of the areas and their distribution patterns. Furthermore, by introducing sliding step lengths (d_1 , d_2 , d_3), the overbreak and underbreak areas for multiple cross-sections within each step length can be computed, and these areas are then integrated to calculate the total volume, as shown in Figure 6. Theoretically, the finer the division of the tunnel cross-sections, i.e., the greater the number of identifiers assigned to each cross-section, the more precise the calculated overbreak and underbreak areas. Similarly, for a given tunnel segment, the smaller the step length (d), the more accurate the calculated overbreak and underbreak volume. This method allows for flexible adjustment based on the requirements for calculation time and accuracy, thereby optimizing the process of overbreak and underbreak detection. It provides greater flexibility and robustness in actual construction project management, offering an efficient and precise method for evaluating tunnel blasting quality.

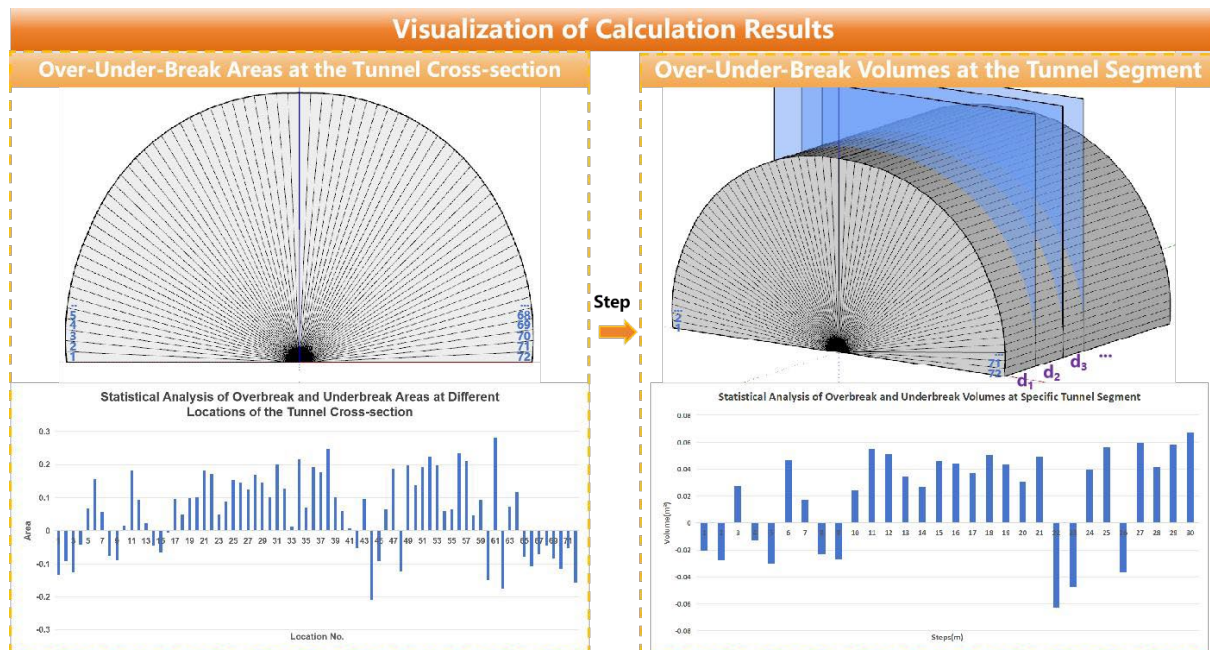


Figure 6. Calculation and Statistics of Tunnel Overbreak and Underbreak Area and Volume

5. CONCLUSIONS

This study proposes an efficient method for detecting tunnel overbreak and underbreak by applying 3D laser scanning technology, point cloud processing, and triangulation techniques to establish a three-dimensional reality-based model of the tunnel, while also constructing the design model based on blueprints. Utilizing digital twin theory, issues traditionally addressed through manual measurements in the physical space are mapped into the digital space, where tunnel overbreak and underbreak detection is achieved through digital model processing and analysis. The detection results are visualized to guide tunnel blasting quality control in the physical space. Compared to traditional measurement methods that require significant human resources, the 3D reconstruction and integrated model analysis, driven by digital twin technology, reduce the reliance on labor-intensive management practices for tunnel construction by replacing manual labor with efficient computational methods. This approach also mitigates subjective errors associated with different personnel, thereby enhancing measurement accuracy. The digital overbreak and underbreak detection method provided in this study supports efficient and precise quality control of drill-and-blast tunnel construction and lays a foundation for achieving intelligent tunnel construction management.

ACKNOWLEDGMENTS

This study was supported by Zhejiang Quzhou-Lishui Railway Co., Ltd., through the project "Theoretical and Applied Research on Digital Twin Modeling of Railway Infrastructure."

REFERENCES

- [1] Koopialipoor, M., Ghaleini, E. N., Tootoonchi, H., Jahed Armaghani, D., Haghighi, M. & Hedayat, A. (2019). Developing a new intelligent technique to predict overcut in tunnels using an artificial bee colony-based ANN, *Environ. Earth Sci.*, 78 (5), 1–14. <https://doi.org/10.1007/s12665-019-8163-x>.
- [2] Van Eldert, J. (2017). Measuring of Over-Break and the Excavation Damage Zone in Conventional Tunneling, World Tunnel Congress 2017. Bergen, Norway.
- [3] Verma, H. K., Samadhiya, N. K., Singh, M., Goel, R. K., & Singh, P. K. (2018). Blast induced rock mass damage around tunnels. *Tunnelling and Underground Space Technology*, 71, 149–158. <https://doi.org/10.1016/j.tust.2017.08.019>.
- [4] Singh, S.P. & Xavier, P. (2005). Causes, impact and control of over-break in underground excavations. *Tunnelling and Underground Space Technology*, 20(1), 63-71. <https://doi.org/10.1016/j.tust.2004.05.004>.
- [5] LEI Mingfeng, ZHANG Yunbo, QIN Guifang, et al. A Study on Neural Network Evaluation Model of Blasting Effect in Mountain Tunnel and Decision-making Method for Blasting Parameter Optimization[J]. *Modern Tunnelling Technology*, 2023, 60 (2): 54-61. <https://doi.org/10.13807/j.cnki.mtt.2023.02.007>
- [6] Ding, X. (2022). Research on the Control Technology of Over-under-excavation of Tunnel Blasting in Interbedded Rock Masses. *Journal of Railway Engineering Society*, 39(3), 75–80.

- [7] Kim, Y. & Bruland, A. (2019). Analysis and Evaluation of Tunnel Contour Quality Index. *Automation in Construction*, 99(MAR.), 223-237. <https://doi.org/10.1016/j.autcon.2018.12.008>.
- [8] Alhaddad, M. (2016). Photogrammetric Monitoring of Cast-iron Tunnels and Applicability of Empirical Methods for Damage Assessment. Ph.D. Dissertation, The University of Cambridge. Cambridge.
- [9] Huang, Y., Liu, F., Wang, J., Zhang, S., & Tang, Q. (2022). A Photogrammetric System for Tunnel Underbreak and Overbreak Detection. *IEEE Transactions on Intelligent Transportation Systems*, 23, 22217-22226. <https://doi.org/10.1109/TITS.2022.3157588>
- [10] Wei, Z., Zhou, Z., Yu, M., & Wang, Y. (2023). Research on Detection Method for Tunnel Underbreak and Overbreak Based on 3D Laser Scanning Point Cloud Integral Analysis. *Tiedao Xuebao/Journal of the China Railway Society*, 45(1), 135–140. <https://doi.org/10.3969/j.issn.1001-8360.2023.01-017>.
- [11] Chaoyang Jin, Yi Chen, Sulong Zhang, Wei Ye, Qingqing Fu. Application and research of tunnel over-under-excavation monitoring technology based on image recognition[J]. Chinese Science and Technology Journal Database (Full Text Edition) Engineering Technology, 2021(3):122-124
- [12] Haishan Zhu, Sheng Li. Application of 3D laser scanning technology in tunnel as-built measurement[J]. Geospatial information, 2021, 19(1):96-98I0007
- [13] Xuan Xie, Xiaoping Zhang, Shaohui Tang et al. Accurate calculation of tunnel overbreak and underbreak volume based on quasimonte carlo method [J/OL]. Geomatics of Wuhan University (Engineering Science), 1-10 [2024-11-14]. <http://kns.cnki.net/kcms/detail/42.1675.T.20240325.1652.002.html>.
- [14] Huang, Y., Liu, F., Wang, J., Zhang, S., Tang, Q. (2022). A Photogrammetric System for Tunnel Underbreak and Overbreak Detection, *IEEE Trans. Intell. Transp. Syst.*, 23(11), 22217-22226. <https://doi.org/10.1109/tits.2022.3157588>.
- [15] Fang, W., Chen, W., Love, P.E., Luo, H., Zhu, H., & Liu, J. (2024). A status digital twin approach for physically monitoring over-and-under excavation in large tunnels. *Adv. Eng. Informatics*, 62, 102648. <https://doi.org/10.1016/j.aei.2024.102648>
- [16] He, Biao & Jahed Armaghani, Danial & Lai, Sai & He, Xuzhen & Asteris, Panagiotis & Sheng, Daichao. (2024). A deep dive into tunnel blasting studies between 2000 and 2023-A systematic review. *Tunnelling and Underground Space Technology*, 147, 105727. <https://doi.org/10.1016/j.tust.2024.105727>.
- [17] He, Biao & Jahed Armaghani, Danial & Lai, Sai. (2023). Assessment of tunnel blasting-induced overbreak: A novel metaheuristic-based random forest approach. *Tunnelling and Underground Space Technology*, 133, 104979. <https://doi.org/10.1016/j.tust.2022.104979>.
- [18] Hekmatnejad, A., Rojas, E., Saavedra, C. & Crespín, B. (2022). Presentation of the Universal Discontinuity index (UDi) system and its application to predict the geometry of over-excavation along a tunnel at New El Teniente mine. *Engineering Geology*, 311, 106901. <https://doi.org/10.1016/j.enggeo.2022.106901>
- [19] Foderà, Giuseppe & Voza, Antonio & Barovero, G. & Tinti, Francesco & Boldini, Daniela. (2020). Factors influencing overbreak volumes in drill-and-blast tunnel excavation. A statistical analysis applied to the case study of the Brenner Base Tunnel – BBT. *Tunnelling and Underground Space Technology*, 105, 103475. <https://doi.org/10.1016/j.tust.2020.103475>.
- [20] Wang, J. & Wu, Fan & Zhou, Xiaobo & Hu, Lisheng. (2021). Evaluation of drill and blast excavation quality for a tunnel. *IOP Conference Series: Earth and Environmental Science*, 861, 042059. <https://doi.org/10.1088/1755-1315/861/4/042059>.
- [21] Zhang, Yunliang & Cao, Wei & Wang, Jian, et al. (2010). Blasting Technique for Controlling Excessive and Insufficient Excavation in Tunnel with Horizontal Layered Rockmass[J]. Journal of Railway Science and Engineering, 7(5): 70-74



Part 2

Construction Safety and Worker Well-being

Creation of a Web-Based Tool for the Visual Inspection of Building Equipment

Angelina Aziz¹, and Markus König²

1) Ph.D. Candidate, Department of Civil and Environmental Engineering, Ruhr University Bochum, Universitaetsstrasse 150, 44801 Bochum, Germany. Email: angelina.aziz@ruhr-uni-bochum.de

2) Prof. Dr.-Ing., Department of Civil and Environmental Engineering, Ruhr University Bochum, Universitaetsstrasse 150, 44801 Bochum, Germany. Email: koenig@inf.bi.rub.de

Abstract: Fire safety inspection is essential for protecting occupants from fire hazards. Traditional inspection methods often rely on subjective assessments, often leading to potential errors and inadequate maintenance. This study introduces a visual fire safety inspection tool that integrates self-trained Machine Learning (ML) services to enhance the accuracy and efficiency of documenting Fire Safety Equipment (FSE) using images. The ML services were incorporated into a web-based application built with the React framework, featuring a backend developed using FastAPI and MongoDB for efficient processing and scalability. The tool achieved a high mean average precision (mAP) over 80% on testing datasets. It offers a robust environment for fire safety experts to validate and compare models, also providing insights into the impact of active learning algorithms. Despite the tool's high accuracy, challenges such as slow loading times and application freezing were identified, with proposed solutions focusing on optimizing backend and frontend processes. The integration of ML services demonstrates significant potential for improving fire safety inspections, with future work aimed at refining models, expanding real-time monitoring capabilities, and ensuring compatibility with Building Information Modeling (BIM) systems and conventional smartphones.

Keywords: Fire safety inspection, computer vision, automation, web-based application

1. INTRODUCTION

Traditionally, fire safety equipment (FSE), such as fire extinguishers, smoke detectors, and safety signs, are inspected manually. These inspections often rely on subjective judgment and manual data entry, increasing the likelihood of human error and incomplete documentation. As buildings grow in complexity, the need for more accurate, efficient, and automated facility management and maintenance becomes essential to ensure safety compliance and reduce risks (Sanzana et al., 2022; Yousefli et al., 2020).

The facility management sector has increasingly adopted Computer-Aided Facility Management (CAFM) software to manage support activities, often complemented by Computerized Maintenance Management System (CMMS) software for more detailed monitoring and maintenance scheduling (Rudl, 2021). However, over the years, researchers have explored various specific strategies to improve facility management processes, particularly during the Operation and Maintenance (O&M) phase of buildings. Previous literature concentrated on developing mobile and web-based applications to facilitate this process. For instance, (Alhaj et al., 2022) proposed an occupant-centric facility maintenance management system, demonstrating the value of mobile applications for streamlining building maintenance tasks. Similarly, (Wen et al., 2017) introduced a web-based hybrid BIM cost-estimating system for fire safety engineering, which includes a "Fire Safety Equipment Property Database" module, underscoring the importance of digital tools for fire safety inspections. Yousefli et al. (2020) developed an automated Multi-Agent Facility Management system (MAFMS) that leverages a web-based application and multi-agent technology to integrate components such as workflow prioritization and resource scheduling. This system effectively automates facility management processes, demonstrating significant improvements in time-sensitive environments like hospitals. Using BIM and robotics, Chen et al. (2023) of this study focus on a knowledge-driven approach through a formalized model called the integrated Scene-Task-Agent (iSTA). This model synergizes information from BIM and robotics to enable autonomous facility inspections. However, while these systems offer significant improvements in overall facility management, they often fall short when it comes to automating specific fire safety inspection subjects, which still require manual data collection and input.

The integration of machine learning (ML) into the fire safety inspection processes offers an opportunity to enhance both the accuracy and efficiency of these inspections. By automating tasks such as detecting fire safety equipment in images, identifying safety signs, and recognizing inspection tags, ML can streamline the inspection process. In particular, computer vision techniques have shown significant promise in the detection and documentation of FSE through image analysis (Aziz et al., 2022; Corneli et al., 2020; Heinbach & Aziz, 2023). However, there is a notable gap in the availability of image-based web tools that can quickly analyze FSE and evaluate captured images within the application, enabling rapid responses and direct data storage.

To address these limitations, this paper presents the concept of a web-based visual fire safety inspection tool that integrates ML services and demonstrates it through a proof-of-concept implementation. By automating the detection and documentation of FSE using images, this tool aims to decrease documentation efforts and streamline the inspection process. The tool leverages computer vision techniques to detect key FSE components directly from images. The web application is built using modern web technologies, with a frontend developed in

React (Clark et al., 2022), and a backend powered by FastAPI (Ramírez, 2023) and MongoDB (Stephens, 2023) for data management. The integration of ML services into this application allows fire safety managers to automate routine evaluation tasks, compare model performance, and validate results.

2. METHOD

The development of the visual fire safety inspection tool was focused on creating a scalable and user-friendly web-based application that automates FSE documentation through the integration of ML services. As illustrated in Figure 1, the workflow starts with users, typically fire safety inspectors, uploading images of FSE such as fire extinguishers, safety signs, and inspection tags through the tool's web interface. Once images are uploaded, the user requests the ML service to automatically detect and document key fire safety features within the images. This is handled using computer vision techniques that analyze the images.

Additionally, the tool incorporates an active learning component that plays a critical role in improving the performance of the ML models. As shown in Figure 1, the system can request annotation from the user on the most informative image data. These are the data that the model finds uncertain or challenging. Once annotated, the data is sent back into the system, where it is stored in the database for further use like retraining the ML models improving their accuracy and performance in detecting FSE.

The tool should manage data storage and facilitate seamless communication between the user and ML services. Inspection results generated by the ML system, including detected FSE and any accompanying documentation, are directly downloadable and archived in the database. This process of archiving results allows users to retrieve and review inspection reports at any time, so that the tool supports long-term data management.

In this workflow, the combination of automated detection, user feedback through active learning, and periodic model retraining creates a tool that not only reduces manual documentation efforts but also adapts and improves overtime, meeting the evolving needs of fire safety inspections.

The next subchapters outline the key components of the tool's creation, from setup and interface design to the integration of ML services and backend architecture.

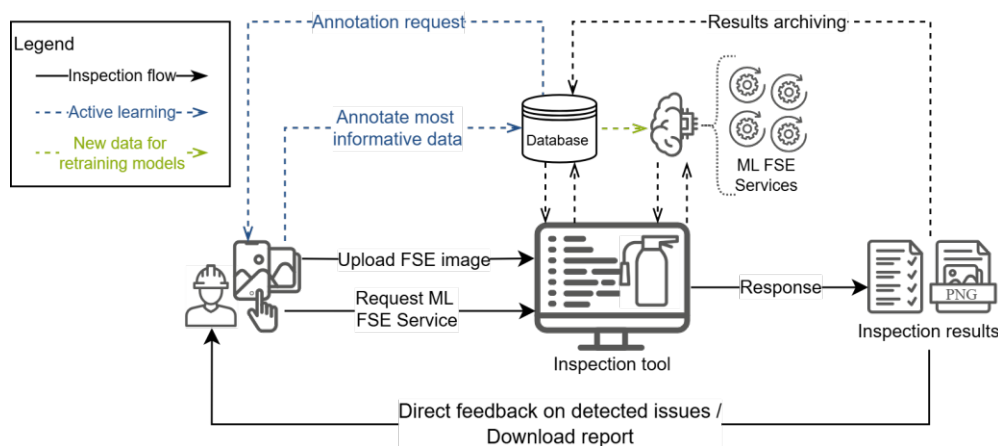


Figure 1. Workflow of the web-based visual fire safety inspection tool integrating ML services and active learning.

2.1 Web Application Setup

The web application was developed using a modular, component-based architecture to ensure flexibility and scalability for both the frontend and backend. As illustrated in Figure 2, the frontend was built using React (Clark et al., 2022), a widely used JavaScript library known for efficiently rendering dynamic user interfaces. React's Virtual DOM optimizes updates in response to user actions, making it an ideal framework for handling large datasets of images and annotations. This is particularly important in the context of fire safety inspections, where the application needs to manage a significant volume of images without sacrificing performance.

The backend was developed using FastAPI (Ramírez, 2023), a modern Python framework designed for high-performance asynchronous operations. FastAPI was chosen for its ease of use and its ability to handle concurrent requests, ensuring that multiple users can interact with the application seamlessly. As shown in Figure 2, FastAPI communicates with the MongoDB (Stephens, 2023) database to store, retrieve, and manage project data, user information, and image annotations. MongoDB was selected for its flexibility in handling dynamic, document-based data, which is generated during on-site inspections and can scale with increasing data volumes.

Additionally, FastAPI integrates with the CVAT (Computer Vision Annotation Tool) (Sekachev, B., Manovich, N., Zhiltsov, M., Zhavoronkov, A., Kalinin, D., Hoff, B., ... & Sidnev, D., 2020), which is used for the annotation of new FSE images. This allows images and their associated annotations to be sent to CVAT for processing, as depicted in Figure 2, and then returned to FastAPI for further handling and storage. The interaction

between FastAPI and MongoDB ensures that annotated images are properly archived and available for later retrieval, while CVAT facilitates the efficient annotation of new data.

To ensure consistent deployment and operation across various environments, the entire application was containerized using Docker. Docker simplifies the deployment process by bundling all components, including the FastAPI backend, MongoDB database, and CVAT integration, into containers that can be deployed on any machine without complex setup procedures. This containerized architecture guarantees that the system performs reliably and efficiently, regardless of the deployment environment.

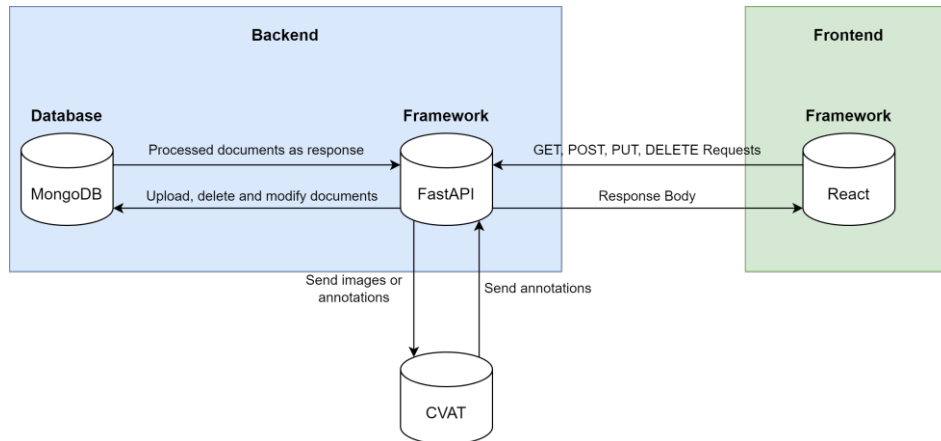





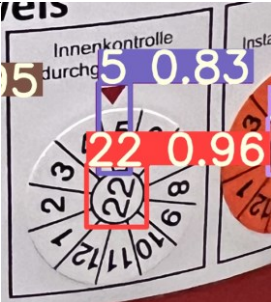
Figure 2. Workflow of the web-based visual fire safety inspection tool integrating ML services and annotation.

2.2 ML Services Integration

The core feature of the web application is its ability to automate the detection and documentation of FSE using ML models. Several custom-trained ML services were integrated into the application to facilitate this process. These services are built on object detection models such as YOLOv5 (Jocher et al., 2022) and YOLOv7 (Wang et al., 2023), selected for their high precision in object detection and their adaptability to the specific needs of fire safety inspections. YOLOv5 and YOLOv7 were chosen based on their proven performance in prior studies on fire safety equipment detection (Aziz et al., 2023; Aziz & König, 2024; Bayer & Aziz, 2022), offering a reliable balance of accuracy and speed ($mAP > 80\%$). Newer models were not adopted to maintain consistency with these validated results and avoid additional retraining efforts.

As illustrated in Table 1, the ML services integrated into the tool cover a range of critical detection tasks, enabling the system to identify FSE, safety signs, inspection tags, and maintenance information. These services were developed based on prior ML-based studies (Aziz et al., 2023; Aziz & König, 2024; Bayer & Aziz, 2022) and are trained on custom datasets. One of the core services is *FSE Detection*, which identifies FSE such as fire extinguishers, fire blankets, smoke detectors and manual call points within images. This service is powered by a custom-trained YOLOv5 model, specifically optimized for real-time object detection. It achieves a mean Average Precision (mAP) of 80.1% at Intersection over Union (IoU) thresholds between 0.5 and 0.95, ensuring reliable identification of critical equipment in various environments. Another essential service is *Safety Sign Detection*, responsible for recognizing safety signage typically found near fire safety equipment.

Table 1. Performance metrics of ML FSE services integrated into the visual inspection tool, detailing the models used and their mAP at various IoU thresholds on test datasets.

FSE Detection	Safety Sign Detection	Inspection Tag Detection	Maintenance Information Detection
			
Used model: YOLOv5	Used model: YOLOv7	Used model: YOLOv7	Used model: YOLOv7
$mAP@0.5:0.95 = 80.1\%$	$mAP@0.5 = 85.2\%$	$mAP@0.5:0.95 = 85.5\%$	$mAP@0.5:0.95 = 86.1\%$

This service uses a YOLOv7 model to detect important safety signs such as exit signs and fire extinguisher location markers. The model performs with an mAP of 85.2% at a 0.5 IoU threshold. *Inspection Tag Detection* is a service focused on detecting inspection tags and manufacturer tag affixed to fire extinguishers and other equipment using the YOLOv7. With an mAP of 85.5% at IoU thresholds between 0.5 and 0.95, this service is highly accurate.

Finally, *Maintenance Information Detection* is another key service integrated into the tool. It is designed to recognize and extract date information from inspection tags, particularly focusing on fields such as the month and year of the last or next inspection. Like the inspection tag service, this functionality is driven by a YOLOv7 model and achieves a high level of accuracy, with an mAP of 86.1% at IoU thresholds of 0.5 to 0.95. These ML services, integrated into the backend architecture, allow users to upload images of FSE and receive automated feedback on the equipment's condition and inspection status.

2.3 User Interface

The user interface of the web application was designed with a focus on usability, scalability, and performance, ensuring that users can effectively manage their images, annotations, and models within various projects. As outlined in Figure 3, the component hierarchy of the application reflects a modular architecture that allows for the smooth management of fire safety inspection tasks, while also accommodating active learning algorithms to minimize the need for manually labeled images.

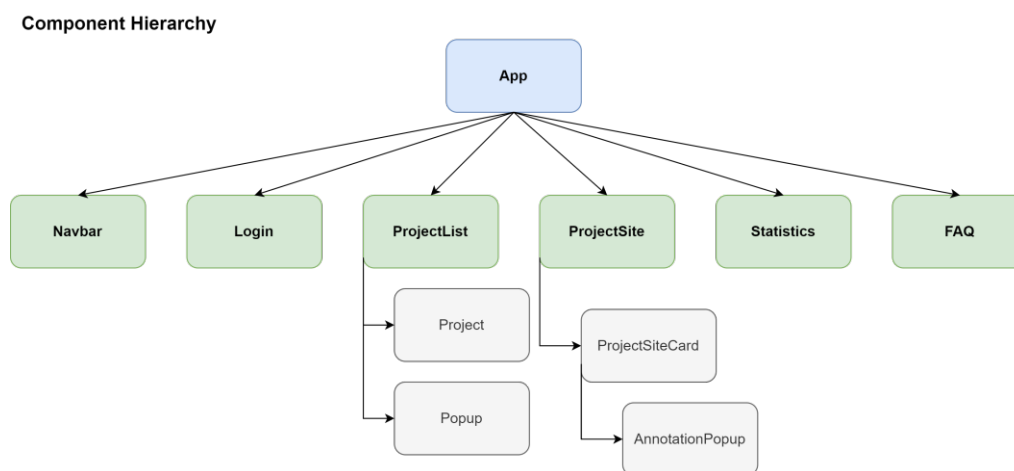


Figure 3. The component hierarchy of the web application.

The *App Component* serves as the root of the application, acting as the central entry point where essential states—such as user credentials, access tokens, and project data—are managed and passed down to other components. This architecture ensures that the system remains both scalable and responsive, making it capable of handling a large number of projects and associated datasets. After logging in via the *Login Component*, where users can either sign in or register for a new account, users are redirected to the *ProjectList Component*.

The *ProjectList Component* offers an overview of all current projects associated with the user's account, including functionalities to create, delete, or search for specific projects. Each project is represented by a *Project Component*, which displays important information, such as the project's image, title, description, and additional metadata. If a new project needs to be created, the *Popup Component* is rendered, allowing the user to enter the necessary details, such as the project's name and description.

Once a project is selected, the user is redirected to the *ProjectSite Component*, which contains the primary functionalities of the application. Within the *ProjectSite Component*, each project is represented visually by a *ProjectSiteCard*, which serves as an interface summarizing the key details of the project. This includes the project's name, associated images, annotation progress, and model versions. The *ProjectSiteCard* provides users with quick access to essential information and actions, such as uploading new images, viewing statistics, or managing annotations.

The active learning mechanism implemented in this study builds upon the approach described by Soutana and Aziz (2023) and is designed to enhance the performance of pretrained object detection models while reducing the annotation burden. The human-in-the-loop framework plays a central role in this process, enabling users to iteratively refine model performance by providing annotations for the most uncertain or challenging samples. Specifically, the algorithm employs class weighting to address class imbalance by assigning higher importance to underrepresented classes, ensuring their inclusion in the training set. Diversity sampling clusters images containing rare classes, improving the likelihood of detecting and training on these less frequent instances. Additionally, image quality assessment via Laplacian variance reduces the influence of low-quality images by prioritizing

higher-quality samples for annotation. This active learning cycle allows for continuous improvement of pretrained models by leveraging user-provided feedback to retrain models with a focus on the most informative and representative data.

The *Statistics Component* enables users to evaluate and validate their models, providing visual comparisons between different versions of their object detection models. Users can easily compare fine-tuned models against the original ones using performance metrics such as mAP, a critical measure of the effectiveness of the active learning algorithms used during the annotation and training process. This functionality allows users to continuously refine their models.

2.4 Backend Architecture

FastAPI (Ramírez, 2023), a modern Python framework, was chosen for the backend due to its ability to support asynchronous programming and its compatibility with key Python libraries, such as Keras and TensorFlow, which are essential e.g., for implementing the active learning algorithms used in this study. FastAPI's simplicity and performance make it an ideal choice for handling large numbers of user requests while maintaining high throughput.

The communication between the frontend and backend is handled through a REST API, which facilitates various operations via GET, POST, PUT, and DELETE requests. To improve performance and handle concurrent requests efficiently, the backend leverages FastAPI's support for asynchronous programming. By using the `async` and `await` keywords, the backend can manage multiple user requests concurrently without blocking the main application. Figure 4 depicts an example showcasing how upcoming requests are processed. For long-running synchronous tasks, such as training a YOLO model or ranking images using active learning, background tasks are assigned to dedicated worker threads using Python's `ThreadPoolExecutor`. This ensures that resource-intensive tasks run in the background, allowing other user requests to be processed simultaneously. Users are kept informed of the status of these tasks through visual feedback, such as loading indicators.

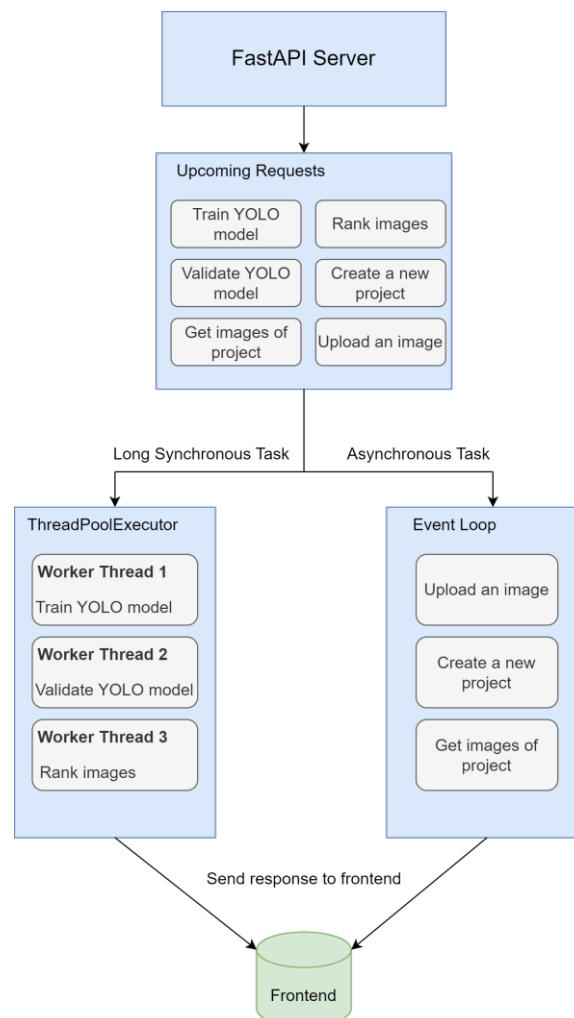


Figure 4. Visualization on how certain requests are handled.

To handle the storage and retrieval of data, the backend uses MongoDB, a document-oriented NoSQL database. MongoDB was selected for its flexibility and scalability, which make it well-suited for managing the unstructured data generated during fire safety inspections, such as images, annotations, and model results. The use of Binary JSON (BSON) allows MongoDB to efficiently store complex data types, enabling quick access and manipulation of data. The structure of the MongoDB database is visualized in the Entity-Relationship Diagram (ER) shown in Figure 5, which illustrates the relationships between various tables and their attributes.

The *Users* table contains essential user information, such as username, email, and password, which are required for authentication. This table is linked to other core components—*Projects*, *Images*, *Models*, and *Annotations*—through foreign keys. Each project, represented in the *Projects* table, has attributes such as the project name, description, and storage paths for images and annotations. The project is also linked to a user and can contain multiple models and annotations, as indicated by the foreign keys in the corresponding tables. The *Images* table stores metadata for each uploaded image, including its name, file type, and storage path. Additionally, it includes a *ranking* attribute, which is generated during the active learning process to prioritize images based on their importance for training. The *Annotations* table links each annotation to the image it was derived from through a one-to-one relationship, allowing each image to have at most one corresponding annotation. The *Models* table contains information about the trained ML models, such as the file type, storage path, and a Boolean *selected* attribute, which indicates whether a model has been chosen for validation or further training.

By structuring the database in this way, the backend ensures that all data components—users, projects, images, annotations, and models—are well-organized and easily accessible. This schema design enables efficient querying and manipulation of data, ensuring that the application can scale effectively as the number of users, projects, and data points grows.

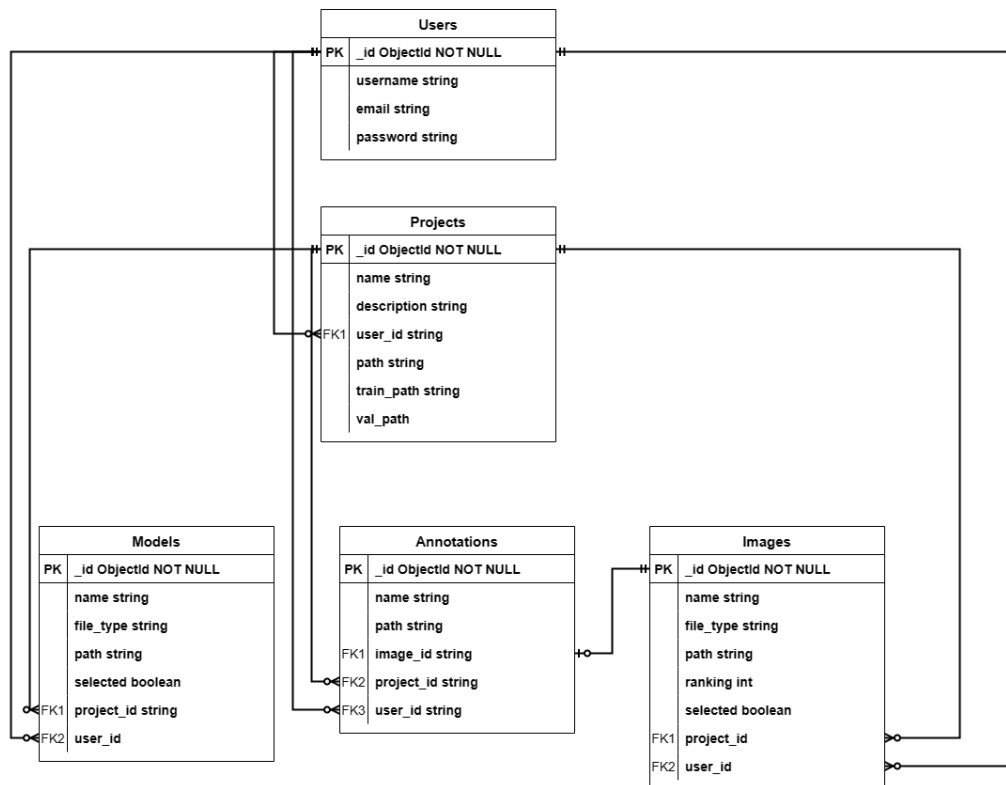


Figure 5. ER Diagram of the database.

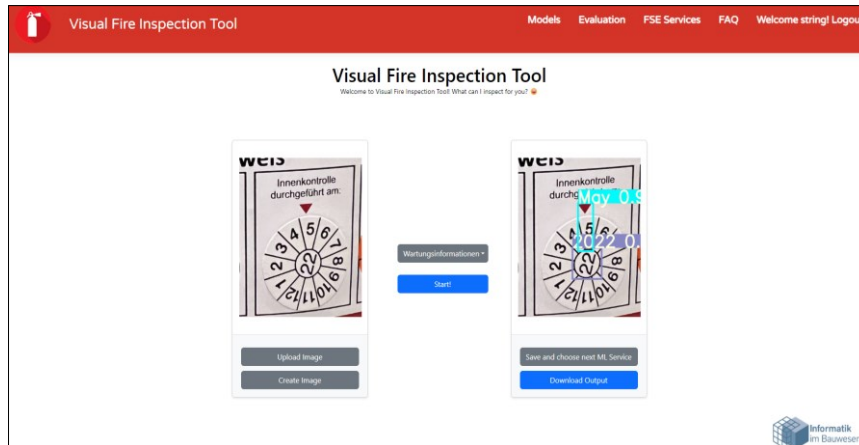
3. RESULTS

The web-based tool was thoroughly tested to evaluate the performance of the integrated ML FSE services. As shown in Table 2, each ML service demonstrated fast inference times once the models were loaded into the web application. However, the *FSE Detection* service, which relies on a YOLOv5 model, exhibited a slower execution time compared to other FSE services utilizing the more recent YOLOv7 models. The results confirm that the more advanced YOLOv7 models used for tasks like *Safety Sign Detection*, *Inspection Tag Detection*, and *Maintenance Information Detection* offer faster and more efficient performance. The user interface for analyzing a test FSE image and selecting an ML FSE service is shown in Figure 6. After successful detection, the original image is displayed with the detection results, and the output can also be downloaded directly.

Table 2. Execution time of the respective ML services.

ML FSE Service	Execution Time (s)
FSE Detection	2.5
Safety Sign Detection	1.36
Inspection Tag Detection	1.52
Maintenance Information Detection	1.32

In addition to performance testing, the active learning approach integrated into the tool was effective in reducing the annotation burden for users. The system allowed users to select the most informative images for annotation, improving the overall efficiency of the annotation process. By leveraging active learning, users were able to prioritize high-value images, which in turn enhanced the model's ability to learn from fewer labeled samples.


 Figure 6. Execution of ML service - *Maintenance Information Detection* in the tool.

Testing also revealed that the integration between the frontend and backend functioned smoothly. The communication with CVAT for annotation purposes was seamless, enabling users to upload images, annotate them, and train models in a streamlined manner. This smooth interaction was critical in ensuring that the tool maintained a responsive and user-friendly experience.

The *Statistics Component* of the tool proved valuable in helping users visually compare model performance. It provided clear insights into the effect of active learning on model accuracy, allowing users to monitor and validate model improvements over time. As shown in Figure 7, the mAP comparison graph displayed the results of different trained models over various epochs, enabling users to see the progression of each model's performance.

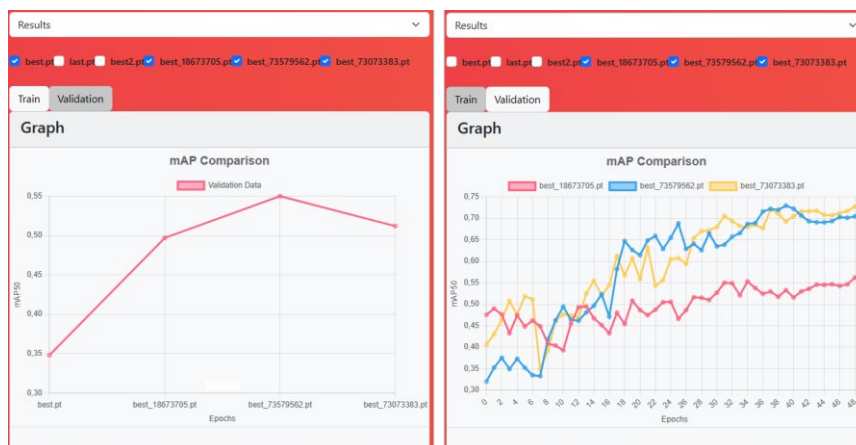


Figure 7. Visual mAP comparison of trained models in the tool. The left diagram shows mAP scores for different models on the training dataset, while the right diagram tracks validation mAP progression over epochs during training, with each line representing a different model.

4. DISCUSSION

The testing of the visual fire safety inspection tool highlighted the strong performance of the integrated ML models in detecting FSE and documenting other FSE-related inspection subjects. Models such as the YOLOv5-based FSE detection model achieved an mAP of over 80%, proving their effectiveness under various conditions. However, several performance bottlenecks were identified during system testing that need to be addressed to optimize the user experience and overall system efficiency.

One of the main issues observed was slow loading times, particularly when handling large datasets or performing inference on multiple images. This was more apparent when switching between models trained on different architectures, such as YOLOv5 and YOLOv7. These delays likely stemmed from how the system manages requests and processes large amounts of data. Additionally, application freezing occurred during extensive image uploads or batch processing, which was traced back to synchronous backend functions causing blocking during intensive operations. Optimizing these backend processes by converting synchronous operations to asynchronous ones could significantly reduce these performance bottlenecks. Transitioning synchronous backend operations to asynchronous processes using tools like Python's *asyncio* could enhance the system's responsiveness. Additionally, integrating distributed computing for batch processing and leveraging cloud-based GPU resources could improve performance when dealing with large datasets or real-time applications.

Despite these challenges, the tool offered several benefits. The *ProjectSite Component* provided users with an intuitive way to navigate between projects, upload images, and run predictions. The integration with the MongoDB database ensured that all project data, images, and annotations were securely stored and easily retrieved, even during system restarts or failures. This was essential in maintaining data integrity throughout the inspection process.

The integration of the tool's active learning algorithm proved to be valuable in improving model accuracy while reducing the annotation burden on users. By focusing on selecting the most informative images for annotation, the active learning approach helped prioritize high-impact data for training. However, this process was computationally expensive, and slowdowns were observed when handling large batches of data. Streamlining the active learning process and improving model retraining efficiency could further enhance the tool's performance and usability.

5. CONCLUSION

The development of the web-based visual fire safety inspection tool, integrating ML services, represents a practical step toward automating fire safety inspections. It improves the accuracy and efficiency of detecting FSE and streamlines the documentation process, helping to reduce human error and save time. The tool's use of ML models, particularly YOLOv5 and YOLOv7, demonstrated high accuracy in identifying FSE, safety signs, inspection tags and maintenance information. Its user-friendly interface, combined with active learning, allows for efficient image annotation, model validation, and the creation of custom datasets, making it adaptable to various inspection needs.

Looking ahead, there are several key areas for future improvement. First, integrating real-time monitoring capabilities into the tool, such as connecting it with BIM, would enable building managers to access up-to-date fire safety documentation. Additionally, adding real-time notifications for FSE maintenance and inspections would further automate the fire safety management process.

Another important enhancement would be to ensure the tool's compatibility with smartphones and mobile devices. This would enable on-site inspectors to perform fire safety inspections directly from their devices, eliminating the need for bulky equipment or manual data entry. Expanding the tool's accessibility to mobile platforms would streamline the inspection process, making it more efficient and practical for fieldwork.

In conclusion, the integration of ML services into fire safety inspections has the potential to revolutionize the field by automating routine tasks, improving documentation accuracy, and enabling real-time monitoring. While the tool has already demonstrated its capabilities in partially automating FSE documentation, ongoing research and development will focus on optimizing performance, expanding functionality, and ensuring scalability for larger operations.

REFERENCES

- Alhaj, M. B., Liu, H., Abudayyeh, O., & Sulaiman, M. (2022). Development of a mobile application for occupant-centric facility maintenance management. In *2022 IEEE World AI IoT Congress (AIIoT)* (pp. 323–329). IEEE. <https://doi.org/10.1109/AIIoT54504.2022.9817248>
- Aziz, A., Herbers, P., Bayer, H., & König, M. (2023). Fully Autonomous Fire Safety Equipment Inspection Missions on a Legged Robot. *Computing in Civil Engineering 2023*, 804–812.
- Aziz, A., & König, M. (2024). Bilderkennungsmethoden für eine teilautomatisierte Inspektion von

- Brandschutzanlagen. *Bautechnik*, 101(3), 159–165. <https://doi.org/10.1002/bate.202300099>
- Aziz, A., König, M., Zengraf, S., & Schulz, J.-U. (2022). Instance Segmentation of Fire Safety Equipment Using Mask R-CNN. *International Conference on Computing in Civil and Building Engineering*, 121–135. https://doi.org/10.1007/978-3-031-35399-4_10
- Bayer, H., & Aziz, A. (2022). Object Detection of Fire Safety Equipment in Images and Videos Using YOLOv5 Neural Network. *Proceedings of 33. Forum Bauinformatik*.
- Chen, J., Lu, W., Fu, Y., & Dong, Z. (2023). Automated facility inspection using robotics and BIM: A knowledge-driven approach. *Advanced Engineering Informatics*, 55, 101838. <https://doi.org/10.1016/j.aei.2022.101838>
- Clark, A., Lunyov, A., Abramov, D., McCabe, D., White, E., Bonta, J., Savona, J., Story, J., Middleton, K., Tan, L., Ruan, L., Wei, L., Carroll, M., Chen, M., Zhang, M., Hanlon, R., Susla, S., Gunasekaran, S., Keegan, S., . . . and Zheng, Y. (2022). React: Software (Version 18.2.0). <https://github.com/facebook/react>
- Corneli, A., Naticchia, B., Vaccarini, M., Bosché, F., & Carbonari, A. (2020). Training of YOLO Neural Network for the Detection of Fire Emergency Assets. *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*(37), 836-843.
- Heinbach, J.-H., & Aziz, A. (2023). *Visual partial inspection of fire safety equipment using machine learning*. Ruhr-Universität Bochum. <https://doi.org/10.13154/294-10096>
- Jocher, G., Chaurasia, A., Stoken, A., Borovec, J., Kwon, Y., Michael, K., & Jain, M. (2022). Ultralytics/yolov5: V7.0 - YOLOv5 SOTA Realtime Instance Segmentation. *Zenodo*. Advance online publication. <https://doi.org/10.5281/zenodo.7347926>
- Ramírez, S. (2023). Fastapi: Software (Version 0.100.1). <https://github.com/tiangolo/>
- Rudl, L. (2021). Analysis of CAFM software for construction companies in the Czech. *IOP Conference Series: Earth and Environmental Science*, 900(1), 12037. <https://doi.org/10.1088/1755-1315/900/1/012037>
- Sanzana, M. R., Maul, T., Wong, J. Y., Abdulrazic, M. O. M., & Yip, C.-C. (2022). Application of deep learning in facility management and maintenance for heating, ventilation, and air conditioning. *Automation in Construction*, 141, 104445. <https://doi.org/10.1016/j.autcon.2022.104445>
- Sekachev, B., Manovich, N., Zhiltsov, M., Zhavoronkov, A., Kalinin, D., Hoff, B., ... & Sidnev, D. (2020). *OpenCV/CVAT: v1.1.0*. Zenodo.
- Soultana, A., & Aziz, A. (2023). Active learning approach for object detection in technical building equipment images. *34. Forum Bauinformatik*, 2023.
- Stephens, A. (2023). MongoDB: Software (Version r.87.0.0-rc10), Article Version r.87.0.0-rc10. <https://github.com/mongodb/>
- Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2023). YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, 7464–7475.
- Wen, M.-H., Huang, W.-C., & Wu, Y.-W. (2017). Development of a web-based hybrid BIM cost-estimating system for fire safety engineering. *Civil, Architecture and Environmental Engineering*, 152–161.
- Yousefli, Z., Nasiri, F., & Moselhi, O. (2020). Maintenance workflow management in hospitals: An automated multi-agent facility management system. *Journal of Building Engineering*, 32, 101431. <https://doi.org/10.1016/j.jobbe.2020.101431>

AUTOMATIC ERGONOMIC ASSESSMENT CONSIDERING AWKWARD POSTURE AND EXTERNAL LOAD

Fei Tian¹ and Yantao Yu²

1) M.Phil., Candidate, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong SAR, China. Email: ftianad@connect.ust.hk

2) Ph.D., Assistant Professor, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong SAR, China. Email: ceyantao@ust.hk

Abstract: The construction industry is characterized by physically demanding tasks and the adoption of awkward postures, both of which contribute to a high incidence of work-related musculoskeletal disorders (WMSDs). Despite the significance of these factors, few studies considered the external load estimation that considers the actual weights being lifted and carried in ergonomic assessments. This research aims to enhance the accuracy of WMSD risk evaluations by integrating external load estimation into ergonomic assessments. We utilized skeleton tracking technology to automatically evaluate awkward postures based on the Rapid Upper Limb Assessment (RULA) framework, a method for evaluating the exposure of workers to ergonomic risk factors. Concurrently, we analyzed electromyography (EMG) signals measuring muscle activity to extract pertinent features for estimating external loads, which were subsequently integrated into the overall ergonomic assessment. Experimental results demonstrate that the Multi-Layer Perceptron-Back Propagation algorithm outperforms alternative machine learning classification methods, achieving an accuracy rate of 98.3%.

Keywords: Ergonomic, External Load, EMG, Awkward Posture, WMSD, RULA

1. INTRODUCTION

Construction workers frequently encounter challenging physical conditions, such as prolonged exposure to uncomfortable postures and the necessity of lifting and transporting heavy materials (Park et al., 2018). These adverse conditions related to ergonomic risk factors not only heighten the possibility of developing Work-Related Musculoskeletal Disorders (WMSDs) but also result in increased error rates and a decline in overall productivity (Li et al., 2024). Consequently, it is imperative to systematically evaluate the ergonomic risks associated with various construction activities for enhancing health and safety management practices, ultimately contributing to the sustainable development of the construction industry (Liao et al., 2023).

Among the various methodologies employed to measure exposure to WMSD risks, posture-based ergonomic assessment stands out as one of the most widely utilized techniques (Janowitz et al., 2006). Traditional methods of manual observation involve an ergonomist who observes workers' postures and movements in real-time or through recorded video footage and then assigns scores with ergonomic assessment tool (Lowe et al., 2014). However, these manual observation techniques are often criticized for being labor-intensive and time-consuming, demanding substantial effort and susceptible to observer bias (Seo & Lee, 2021), which hinders the efficiency and reliability of ergonomic assessments in dynamic construction environments. In response to these limitations, recent advancements in ergonomic assessment have sought to employ wearable sensors and vision-based method to automate the recognition of potentially hazardous postures during ongoing tasks (Wang et al., 2015). Wearable sensor systems, such as joint angle measurement systems (Rodrigues et al., 2022) and inertial measurement units (IMUs) (Yan et al., 2017), typically concentrate on monitoring specific joints and movements. The focus on isolated joints may overlook how various body segments interact during complex tasks. Vision-based methods offer a more holistic approach to ergonomic assessment by evaluating the entire configuration of the body. Yu et al. (2019b) and Roberts et al. (2020) explored innovative approaches to three-dimensional (3D) pose estimation by utilizing two-dimensional (2D) video inputs. With advancements in technology that can capture joint and activity information in real-time, Lin et al. (2022) developed a system that automatically selects appropriate assessment scales and calculates risk scores using image-based motion capture techniques. Hossain et al. (2023) employed deep learning techniques to predict ergonomic risk levels by analyzing 3D coordinates of human body positions.

Despite the substantial body of research dedicated to posture estimation, the consideration of external load in ergonomic assessments is often overlooked. This oversight is particularly significant in construction settings, where workers frequently engage with external loads. Some studies (Fortini et al., 2020; Lorenzini et al., 2019; Ventura et al., 2021) have explored external loads with the focus on individual body joints, which may fail to capture the broader implications of external loads on the worker's entire musculoskeletal system. Yu et al. (2019a) developed a smart insole to estimate external load by using total weight minus the worker's self-weight. However, it assumes that all external load is transmitted through the feet, rendering it ineffective for postures such as leaning against walls or sitting on the ground. In contrast, electromyography (EMG) offers a more comprehensive solution, as it offers insights into muscle activation patterns and the physiological demands placed

on the body (Wang et al., 2015). By quantifying the electrical activity of muscles, EMG can provide a clear understanding of how external loads affect worker performance across various postures. Recent research by Kumari et al. (2023) exemplifies the potential of EMG in this context. Their analysis of EMG data focused on agricultural workers engaged in pushing and pulling operations, revealing significant changes in muscle activity corresponding to varying external loads. This approach highlights the utility of EMG in capturing the dynamic relationship between external loads and muscle engagement, offering a more nuanced understanding of ergonomic stressors.

Our study proposes a framework that combines skeleton tracking technology with EMG signals to deliver a comprehensive assessment of worker well-being. By employing skeleton tracking, the system captures the 3D coordinates of body joints and employs the Rapid Upper Limb Assessment (RULA) method for automatic evaluation of human posture and movement. Concurrently, the analysis of EMG signals provides critical insights into the estimation of external loads placed on workers. This study will examine 40 widely utilized electromyography (EMG) features to identify the optimal combination that most accurately reflects external load. By integrating EMG analysis into ergonomic assessments, the framework enhances the monitoring of ergonomic risks and alerts workers to the potential risks of overexertion before injuries occur. Section 2 provides details of our proposed method to monitor ergonomic and estimate external load. The experimental design and results will be presented in Section 3, while the discussion will be illustrated in Section 4. The conclusion will be finally drawn in Section 5.

2. METHOD

2.1 Ergonomic Assessment

RULA is a widely recognized framework for evaluating ergonomics that aims to identify potentially harmful postures and movements that could contribute to the development of WMSDs (Nayak & Kim, 2021). RULA considers biomechanical and postural load requirements of job demands on the neck, trunk, and upper extremities. By using RULA, employers can take proactive steps to manage the risks associated with MSDs, creating a safe and comfortable work environment for workers.

Despite its utility, the RULA assessment has significant limitations, particularly in dynamic and fast-paced environments like construction sites. Since RULA only allows the evaluator to assess worker's posture at one point in time, it usually selects the representative postures, like the most difficult postures based on the worker interview and initial observation, or postures maintained for the longest period, or postures where the highest force loads occur (Middlesworth, 2012). The assessment generally involves analyzing a static image of the chosen posture, which can take approximately 10 to 20 minutes, depending on the complexity of the task being performed. However, construction workers frequently change their postures and perform a variety of tasks quickly, making it difficult for evaluators to capture an accurate representation of ergonomic risks at any single moment. This static approach can lead to incomplete assessments and may overlook critical factors affecting worker safety and comfort. Therefore, our method aims to enhance this process by automatically calculating the RULA score, allowing for more efficient and timely instructions. The model proposed in this project can detect real-time motion of human and generate 3D-coordinate of 32 joints. The joint data can subsequently be visualized in the form of a skeletal representation and used to evaluate the ergonomic level automatically based on RULA.

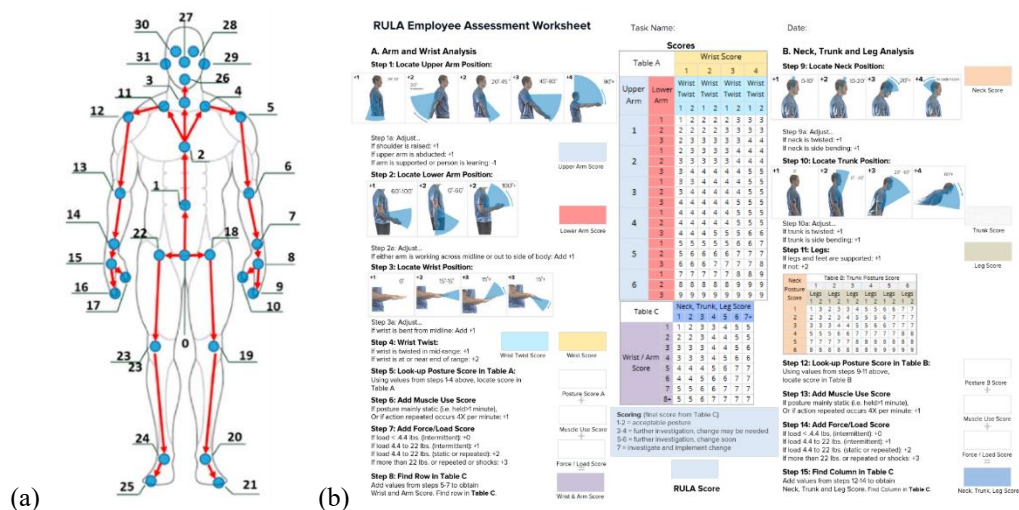


Figure 1. (a) Human Skeleton Model (b) RULA

Azure Kinect developed by Microsoft is capable of capturing depth images and tracking skeletal joints, allowing for the representation of position and orientation in three-dimensional space (Ahad et al., 2021). The skeleton data consists of a set of P joints $J = [J_1, J_2, J_3, \dots, J_P]$ where P equals to the number of joints. In our method, the dataset contains 32 joints, as shown in Figure 1(a). We calculated a position-based kinematics feature vector for every skeleton frame, with each joint denoted as a three-dimensional vector $J_i = [J_{ix}, J_{iy}, J_{iz}]$ within the Kinect coordinate system. The skeleton vector \mathbf{v} be calculated by subtracting the coordinates of the final joint point J_F from the initial joint point J_S . Some RULA criteria are determined by analyzing the amplitude of body segment movement, such as the swing of upper arm, as shown in Figure 1(b). In the Azure coordination, there are three directions of vectors $\mathbf{v}_i = \{\mathbf{v}_x, \mathbf{v}_y, \mathbf{v}_z\}$. The amplitude of movement can be represented as the angle between the skeleton vector and direction of vector, calculated by Equation (1).

$$\cos(\theta) = \frac{\mathbf{v} \cdot \mathbf{v}_i}{|\mathbf{v}| |\mathbf{v}_i|} \quad (1)$$

When assessing the upper arm position, humans are supposed to stand on the planes parallel to x-z plane according to the coordinate systems in the depth camera. The direction of vector \mathbf{v}_i should be \mathbf{v}_y that is $[0, 1, 0]$. The vector of left upper arm \mathbf{v} can be determined by subtracting the coordinates of the final point of left elbow from the initial point of left shoulder. The score of left upper arm will be one from 20-degree extension to 20-degree flexion ($20 \leq \theta$). The score of left upper arm will be two for 20-to-45-degree extension or flexion ($20 < \theta \leq 45$). The score of left upper arm will be three for 45-to-90-degree extension or flexion ($45 < \theta \leq 90$). The score of left upper arm will be three for more than 90-degree extension or flexion ($90 < \theta$).

2.2 Classification Method

In this study, we frame the estimation of external load as a classification problem, employing electromyography (EMG) features as input variables to explore robust classification techniques. Our objective is to effectively categorize the EMG signals into three distinct load levels: 0 kg, 2 kg, and 5 kg. The integration of machine learning algorithms into this framework enhances the capacity to detect and classify high-dimensional data, facilitating a systematic analysis of the complex patterns inherent in the EMG signals. Given the variability of signal patterns caused by external influences, such as changes in electrode positioning, classifiers are particularly well-suited to manage these fluctuations and mitigate the risk of overfitting (Asghari & Hu, 2007). To address the requirements for real-time processing and long-term operation, we will compare several advanced classification methods, including Multi-Layer Perceptron-Back Propagation (MLP-BP) neural networks (Reyes-Fernandez et al., 2024), K-Nearest Neighbors (KNN) (Bukhari et al., 2020), and Support Vector Machine (SVM) (Liu et al., 2021).

(1) MLP-BP

MLP, as a type of artificial neural network, can establish mathematical models through a finite number of iterations, which suits the study of complex nonlinear characteristics (Dellacasa Bellingegni et al., 2017). It contains three types of layers: the input layer to present data to the network, hidden layer to perform weight vector computation on input data, and output layer to predict the response value. Each layer is composed of different numbers of neurons, and the number of neurons in the input layer and output layer corresponds to the number of variables. The number of hidden layers and neurons in hidden layers plays an imperative role in the output response. MLP uses the BP algorithm to train the network by propagating the error back through the layers of the network and adjusting the weights and biases of each neuron in the network, such that the error is minimized (Kotsiantis, 2007). MLP-BP has advantages in its accuracy and its excellent generalization capability, which can deal with incomplete, noisy, and fuzzy data (Kukreja et al., 2016). Meanwhile, it belongs to computationally intensive algorithms, which are time-consuming.

(2) KNN

The fundamental principle of KNN is that classification of unknown instances can be implemented by relating the unknown to the known based on some distance function (Paul et al., 2017). KNN algorithm finds the K-nearest neighbors to the unknown instance, and the label of the unknown instance is allocated according to the majority vote of the K-nearest neighbors. It can be considered as two stages: a) use Euclidean distance to make pre-grouping of data and create subclusters and b) use a similar measure to merge the subclusters hierarchically (Rechy-Ramirez & Hu, 2011). As a non-parametric algorithm, KNN is intuitive and easy to implement. Without training steps, new data can be added seamlessly, which allows a quick respond to changes in the input. Nevertheless, due to computation complexity and memory limitation, KNN has poor performance of large dataset and high dimensionality (Soofi & Awan, 2017).

(3) SVM

SVM is a statistical learning system for classification and regression analysis, which is essentially to determine a hyperplane or a boundary that separates the training data into classes (Subasi, 2013). New measured data is then stored and fed back to the model so that the influence of unknown disturbances frequently appeared in the process can be compensated, which improves the accuracy and achieves self-optimization. SVM is applicable not only to training data with many features relative to the number of training instances (Kotsiantis, 2007), but also to a wide range of classification problems such as high dimensional and not linearly separable

problems. However, SVM will be limited to the dataset with more data or noise.

2.3 Load Estimation Based On EMG Features

Electromyography (EMG) is an instrument that allows the measurement and analysis of the electrical signals that are produced by muscle activity. Reaz et al. (2006) illustrated that a muscle is composed of bundles of specialized cells capable of contraction and relaxation. Skeletal muscle attaches to the bone, as one of the three significant muscle tissues, consists of thousands of muscle fibers wrapped together by connective tissue sheaths. Skeletal muscle can receive and respond to stimuli. Its contraction is initiated by electrical impulses that travel between the central and peripheral nervous systems and muscles, which facilitates the support and movement of the skeleton. When skeletal muscles contract, the electrical activity of the muscle fibers active can be detected by surface electrodes. After amplification, filtering, and processing, EMG signals can be acquired, which provides information about muscle function, like muscle activation and muscle recruitment patterns.

In this study, EMG is used to provide a estimation of external load. Due to the enormous number of inputs and randomness of the EMG signal, it is impractical to feed the raw signals to a classifier directly (Asghari & Hu, 2007). Feature extraction converts the raw signals acquired into a relevant data structure by eliminating background noise and highlighting the important data (Rechty-Ramirez & Hu, 2011). The investigation of feature extraction characteristics in both the time domain and frequency domain has gained significance in the classification of EMG signals (Phinyomark et al., 2012). 40 feature extraction methods that are widely applied are introduced in Table 1.

Table 1. EMG Feature and Equation

no	Symbol	Feature Description	Equation
1	IEMG	Integrated EMG	$IEMG = \sum_{i=1}^N x_i $
2	MAV	Mean Absolute Value	$MAV = \frac{1}{N} \sum_{i=1}^N x_i $
3	MMAV	Modified Mean Absolute Value	$MMAV = \frac{1}{N} \sum_{i=1}^N w_i x_i $ $w_i = \begin{cases} 1, & \text{if } 0.25N \leq i \leq 0.75N \\ 0.5, & \text{otherwise} \end{cases}$
4	MMAV2	Modified Mean Absolute Value 2	$MMAV2 = \frac{1}{N} \sum_{i=1}^N w_i x_i $ $w_i = \begin{cases} 1, & \text{if } 0.25N \leq i \leq 0.75N \\ \frac{4i}{N}, & \text{elseif } i < 0.25N \\ \frac{4(i-N)}{N}, & \text{otherwise} \end{cases}$
5	SSI	Simple Square Integral	$SSI = \sum_{i=1}^N x_i^2$
6	VARE	Variance of EMG	$VARE = \frac{1}{N-1} \sum_{i=1}^N x_i^2$
7	TM	Temporal Moment	$TM = \left \frac{1}{N} \sum_{i=1}^N x_i^3 \right $
8	RMS	Root Mean Square	$RMS = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}$
9	VO	V-Order	$x_i = (\gamma m_i^\alpha) n_i$ $VO = \left(\frac{1}{N} \sum_{i=1}^N x_i^v \right)^{\frac{1}{v}}$
10	LD	Log Detector	$LD = e^{\frac{1}{N} \sum_{i=1}^N \log(x_i)}$
11	WL	Waveform Length	$WL = \sum_{i=1}^{N-1} x_{i+1} - x_i $
12	AAC	Average Amplitude Change	$AAC = \frac{1}{N} \sum_{i=1}^{N-1} x_{i+1} - x_i $
13	DASDV	Difference Absolute Standard Deviation Value	$DASDV = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N-1} (x_{i+1} - x_i)^2}$
14	ZC	Zero Crossing	$ZC = \sum_{i=1}^{N-1} [\text{sgn}(x_i \times x_{i+1}) \cap x_i - x_{i+1} \geq \text{threshold}]$ $\text{sgn}(x) = \begin{cases} 1, & \text{if } x \geq \text{threshold} \\ 0, & \text{otherwise} \end{cases}$
15	MYOP	Myopulse Percentage Rate	$MYOP = \frac{1}{N} \sum_{i=1}^N [f(x_i)]$ $f(x) = \begin{cases} 1, & \text{if } x \geq \text{threshold} \\ 0, & \text{otherwise} \end{cases}$
16	WA	Willison Amplitude	$WA = \sum_{i=1}^{N-1} [f(x_i - x_{i+1})]$ $f(x) = \begin{cases} 1, & \text{if } x \geq \text{threshold} \\ 0, & \text{otherwise} \end{cases}$
17	SSC	Slope Sign Change	$SSC = \sum_{i=2}^{N-1} [f[(x_i - x_{i-1}) \times (x_i - x_{i+1})]]$ $f(x) = \begin{cases} 1, & \text{if } x \geq \text{threshold} \\ 0, & \text{otherwise} \end{cases}$

18	SKEW	Skewness	$SKEW = \frac{M_3}{M_2} \sqrt{M_2}$
			$M_k = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^k$
19	KURT	Kurtosis	$KURT = \frac{M_4}{M_2 M_2}$
20	MFL	Maximum Fractal Length	$MFL = \log_{10}(\sqrt{\sum_{i=1}^{N-1} (x_{i+1} - x_i)^2})$
21	DVARV	Difference Variance Value	$DVARV = \frac{1}{N-2} \sum_{i=1}^{N-1} (x_{i+1} - x_i)^2$
22	IQR	Interquartile Range	$IQR = Q_3 - Q_1$
23	MAD	Mean Absolute Deviation	$MAD = \frac{1}{N} \sum_{i=1}^N x_i - \bar{x} $
24	AR	Auto-Regressive Model	$x_i = -\sum_{p=1}^P a_p x_{i-p} + w_i$
25	AE	Average Energy	$AE = \frac{1}{N} \sum_{i=0}^{N-1} x_i^2$
26	VAR	Variance	$VAR = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2$
27	SD	Standard deviation	$SD = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}$
28	CARD	Cardinality	$y_i = \text{sort}(x_i)$
			$CARD = \sum_{i=1}^{N-1} y_i - y_{i+1} , \text{ if } y_i - y_{i+1} > \text{threshold}$
29	EMAV	Enhanced Mean Absolute Value	$EMAV = \frac{1}{N} \sum_{i=1}^N x_i^p $
			$p = \begin{cases} 0.75, & \text{if } 0.2N \leq i \leq 0.8N \\ 0.5, & \text{otherwise} \end{cases}$
30	EWL	Enhanced Wavelength	$EWL = \sum_{i=2}^N (x_i - x_{i-1})^p $
			$p = \begin{cases} 0.75, & \text{if } 0.2N \leq i \leq 0.8N \\ 0.5, & \text{otherwise} \end{cases}$
31	NZC	New Zero Crossing	$NZC = \begin{cases} 1, & \text{if } x_i > \text{threshold and } x_{i+1} < \text{threshold} \\ & \text{or } x_i < \text{threshold and } x_{i+1} > \text{threshold} \\ 0, & \text{otherwise} \end{cases}$
			$\text{threshold} = 4(\frac{1}{10} \sum_{i=1}^{10} x_i)$
32	ASS	Absolute value of the Summation of Square Root	$ASS = \sum_{i=1}^N (x_i)^{1/2} $
33	MSR	Mean value of the Square Root	$MSR = \frac{1}{N} \sum_{i=1}^N (x_i)^{1/2}$
34	ASM	Absolute Value of Summation of the exp th root	$ASM = \sum_{i=1}^N (x_i)^{exp} $
			$exp = \begin{cases} 0.75, & \text{if } 0.25N \leq i \leq 0.75N \\ 0.25, & \text{otherwise} \end{cases}$
35	DAMV	Difference Absolute Mean Value	$DAMV = \frac{1}{N-1} \sum_{i=1}^{N-1} x_{i+1} - x_i $
36	LDAMV	Log Difference Absolute Mean Value	$LDAMV = \log(\frac{1}{N-1} \sum_{i=1}^{N-1} x_{i+1} - x_i)$
37	LDASDV	Log Difference Absolute Standard Deviation	$LDASDV = \log \sqrt{\frac{1}{N-1} \sum_{i=1}^{N-1} (x_{i+1} - x_i)^2}$
38	COV	Coefficient of Variation	$COV = \frac{\sqrt{\frac{1}{N-1} \sum_{i=1}^{N-1} (x_i - \bar{x})^2}}{\frac{1}{N} \sum_{i=1}^{N-1} x_i}$
39	LCOV	Log Coefficient of Variation	$LCOV = \log \frac{\sqrt{\frac{1}{N-1} \sum_{i=1}^{N-1} (x_i - \bar{x})^2}}{\frac{1}{N} \sum_{i=1}^{N-1} x_i}$
40	LTKEO	Log Teager Kaiser Energy Operator	$LTKEO = \log \sum_{i=0}^{N-2} (x_i^2 - x_{i-1} x_{i+1})$

3. EXPERIMENTS AND RESULTS

3.1 Experiment Design

To examine the feasibility of the EMG-based parameters in classifying workers' external loads, we conducted an experiment and recorded the electrical activity of 10 healthy subjects while performing tasks with different external loads. 5 males and 5 females were randomly selected, aged between 19 and 22 years old. The surface of the collection site (skin) was cleaned with alcohol and coated with glycerin or conductive paste to reduce skin surface impedance and enhance electrical conductivity.

Subjects were asked to stand with the upper arm perpendicular to the ground without the abduction or adduction while the lower arm parallel to the ground without the extension or flexion, as Figure 2 shown below. Three tasks required subjects to carry 0 kg, 2 kg, and 5 kg on their hands respectively. Subjects were asked to

maintain the posture for around 6 seconds for one exercise. EMG data were collected after carrying the weight, which meant that the record did not include the process of putting up or down. A short period of rest was provided between each exercise to mitigate the cascading effect of fatigue. In total we collected 10 subjects \times 3 exercises \times 4 repetitions = 120 EMG recordings.

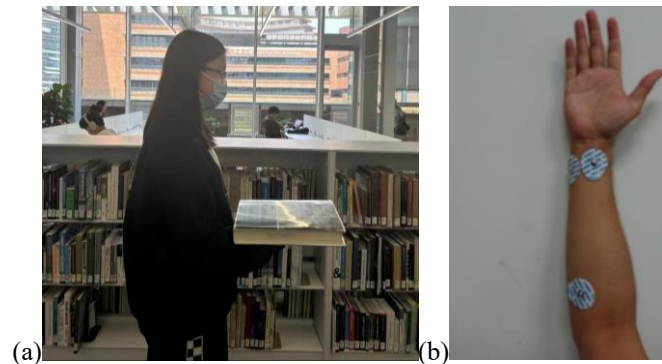


Figure 2. (a) Posture of the Subject (b) Positions of Surface Electrode

3.2 Experiment Results

(1) Pose Estimation Results

The automatic ergonomic assessment system has successfully implemented the real-time skeleton monitoring and visualization techniques to evaluate an individual's posture and movement dynamics, shown in Figure 3. This approach facilitates simultaneous monitoring of both the left and right sides of the body, and then selecting the higher score for the overall ergonomic assessment. Throughout the assessment process, essential metrics such as joint angles, posture deviations, and movement patterns are meticulously evaluated and presented in an intuitive format, enhancing the comprehension of ergonomic status.

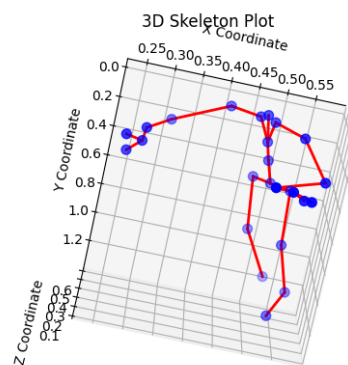


Figure 3. 3D Skeleton Plot

(2) Load Estimation Results

The surface EMG signals collected will inevitably be mixed with noise which can cause invalid features and disturb the classification. Three main types of noise in surface EMG signals include the inherent noise of the electronic components (0 to several kHz), power frequency interference (50 Hz or 60 Hz), and the baseline to drift (0 to 20 Hz) (Li et al., 2020). Therefore, preprocessing is required to denoise signals and enhance feature extraction. Bandwidth of 20 to 500 Hz Bandpass filter and 50 Hz notch filter were applied in this study to remove noise interference. The EMG signal output when the subjects carry different weight is shown in Figure 4.

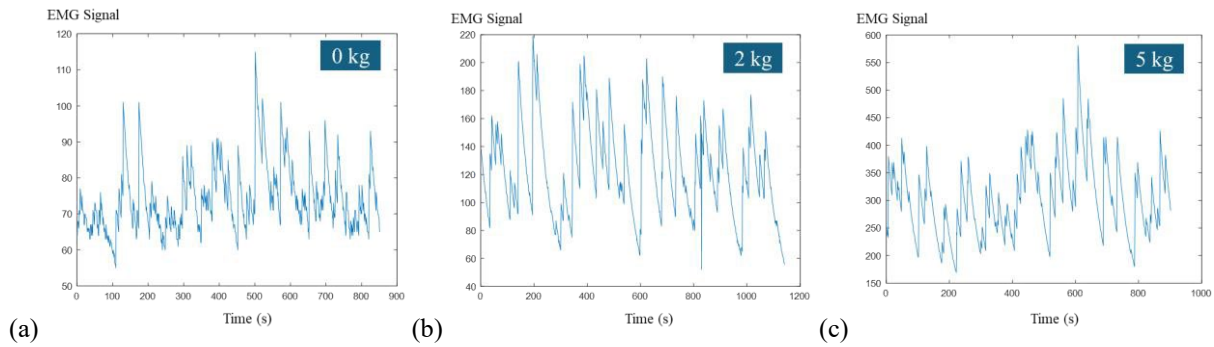


Figure 4. EMG Output with (a) 0 kg (b) 2 kg (c) 5kg

The processing of redundant and irrelevant features can diminish model accuracy while also prolonging execution time during classification tasks. To mitigate these issues, a feature selection algorithm is employed between the feature extraction and classification phases, allowing for the automatic identification of critical features (Baby et al., 2021). Our study utilized the Extra Trees classifier to discern the most significant features from the original set. This ensemble learning method enhances the tree splitting technique by reducing the variances inherent in many tree-based and neural network algorithms (Ossai & Wickramasinghe, 2022). Each tree is constructed from the original dataset. At each test node, extra tree is provided with a random sample of k features in that each decision tree selects the most relevant feature to split the data according to the Gini index, which assists the further formulation of multiple de-correlated decision trees (Sharma et al., 2022). Ultimately, all features are ranked in descending order according to their Gini index scores, from which a specified number of top features is selected based on their importance. In this study, 0 kg, 2 kg, and 5 kg were labeled with No-Load, Little-Load, and High-Load respectively. With the extra tree classifier, seven features with the most importance were selected: MSR, LD, CARD, SKEW, EMAN, ASM, LDASDV, as shown in Figure 5.

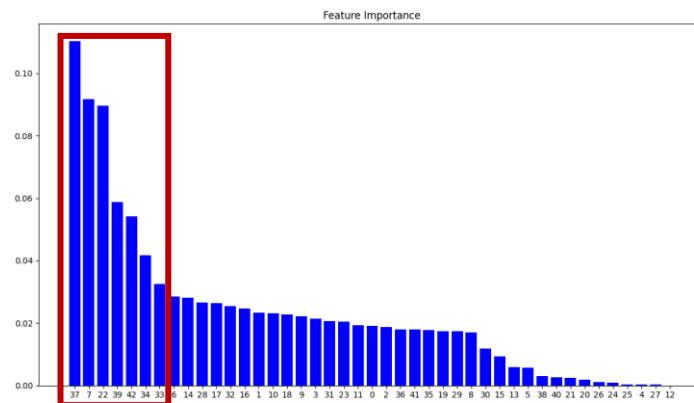


Figure 5. Feature Selection

Given the available data, 84 trials were fed in for training while the remaining 36 trials were used for testing. In this study, MLP-BP had 7 neurons in the input layer that was equal to the number of features selected and 3 neurons for output layer that was the same as the number of labels. MLP-BP applied two hidden layers: the first layer had 30 neurons, and the second layer had 10 neurons. For this configuration, the Rectified Linear Unit (ReLU) activation function and sigmoidal activation function were used in the hidden and output layers respectively. The accuracy of MLP-BP model achieved 98.3% compared with KNN (94.4%) and SVM (97.2%).

(3) Ergonomic Assessment Considering Awkward Posture and External Load

The system is designed to automatically evaluate the ergonomic score for each body part based on the assessed posture of the worker. This evaluation process incorporates both the individual scores derived from body positioning and the estimated external load levels experienced during tasks. By integrating these two critical factors, the system provides a more comprehensive ergonomic assessment. The final score is determined using the RULA worksheet, which systematically evaluates the combined effects of awkward postures and external loads. This holistic approach not only highlights potential risks associated with specific postures but also addresses the impact of external loads on overall worker well-being. Consequently, the resulting ergonomic score serves as a valuable tool for identifying areas that may require intervention, ultimately promoting safer and more efficient work practices.

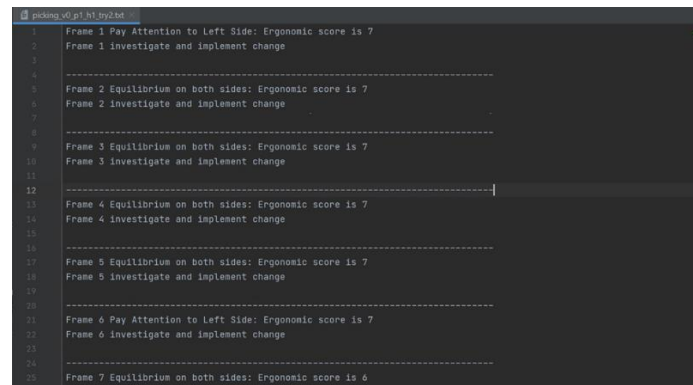


Figure 6. Ergonomic Assessment Result

4. DISCUSSION

By evaluating external load levels, the system can effectively identify instances of overloading and encourage workers to seek assistance as needed, thereby enhancing their overall well-being and minimizing the risk of injury. Furthermore, incorporating external load assessments into the RULA framework significantly enriches the evaluation of ergonomic risk factors. While traditional ergonomic assessments primarily focus on postural evaluation and movement patterns, the integration of load variables offers a more comprehensive understanding of the physical demands placed on workers. This holistic approach leads to a more nuanced and accurate assessment of ergonomic risks, facilitating the design of targeted interventions that can mitigate these hazards and ultimately foster a safer and more productive work environment.

The ergonomics model developed in this study can be seamlessly integrated with a digital display, providing real-time feedback on the ergonomic status for each assessment frame. For instance, when scores reach 5 or 6, the system indicates that further investigation and adjustments are necessary, prompting the display to alert the worker with a yellow light. A score of 7 signifies an urgent need for intervention, triggering a red-light alert. Additionally, the system can identify which side of the body presents greater ergonomic risks, enabling targeted corrective actions.

5. CONCLUSIONS

In this study, we present a framework that combines skeleton tracking technology with EMG signals to provide a thorough assessment of worker well-being. Skeleton tracking provided 3D coordinates of body joints and achieved automated evaluation of human posture and movement. This research also selected seven EMG features that are most pertinent to external load classification through Extra Trees classifier. By analyzing these features, we can effectively estimate external load levels in real time, thereby enhancing the ergonomic assessment process. Our results indicate that the MLP-BP algorithm outperforms alternative classification methods, achieving an exceptional accuracy rate of 98.3%. This finding highlights the potential of advanced machine learning techniques to improve safety and efficiency in HRC.

However, this research has limitations regarding the experimental setting, which was conducted in a controlled laboratory environment. While this setting allows for precise measurement and analysis, it may not fully capture the complexities and variabilities present in real-world scenarios. Future studies should aim to conduct field experiments to validate the framework under diverse working conditions, accounting for factors such as varying work environments, different task demands, and the presence of multiple external influences. Additionally, involving a more diverse range of test subjects in future studies would provide a more comprehensive view of the proposed work, particularly considering individuals with different demographics, such as age, gender, and physical capability. This diversity could enhance the robustness of the findings and make the framework applicable to a broader population. By expanding the research to include these real-world settings and diverse participants, we can improve the applicability of our findings and further refine the framework to better support worker safety and well-being in actual workplace contexts.

ACKNOWLEDGMENTS

This work was supported by the Young Scientists Fund of the National Natural Science Foundation of China [grant number 72201226], the Public Policy Research Project funded by the Chief Executive's Policy Unit of the Government of the Hong Kong Special Administrative Region [grant number 2023.A7.030.23C], and the Construction Safety Fund supported by the Hong Kong Construction Association.

REFERENCES

- Ahad, M. A. R., Ahmed, M., Das Antar, A., Makihara, Y., & Yagi, Y. (2021). Action recognition using kinematics posture feature on 3D skeleton joint locations. *Pattern Recognition Letters*, 145, 216–224. <https://doi.org/10.1016/j.patrec.2021.02.013>
- Soofi, A. A., & Awan, A. (2017). Classification techniques in machine learning: applications and issues. *Journal of Basic & Applied Sciences*, 13, 459–465. <https://doi.org/10.6000/1927-5129.2017.13.76>
- Asghari Oskoei, M., & Hu, H. (2007). Myoelectric control systems—A survey. *Biomedical Signal Processing and Control*, 2(4), 275–294. <https://doi.org/10.1016/j.bspc.2007.07.009>
- Baby, D., Devaraj, S. J., Hemanth, J., & Anishin Raj, M. M. (2021). Leukocyte classification based on feature selection using extra trees classifier: A transfer learning approach. *Turkish Journal of Electrical Engineering and Computer Sciences*, 29, 2742–2757. <https://doi.org/10.3906/elk-2104-183>
- Bukhari, W. M., Yun, C. J., Kassim, A. M., & Tokhi, M. O. (2020). Study of K-Nearest Neighbour classification performance on fatigue and non-fatigue EMG signal features. *International Journal of Advanced Computer Science & Applications*, 11(8), 41–47. <https://doi.org/10.14569/IJACSA.2020.0110806>
- Dellacasa Bellingegni, A., Gruppioni, E., Colazzo, G., Davalli, A., Sacchetti, R., Guglielmelli, E., & Zollo, L. (2017). NLR, MLP, SVM, and LDA: A comparative analysis on EMG data from people with trans-radial amputation. *Journal of Neuroengineering and Rehabilitation*, 14(1), 82–82. <https://doi.org/10.1186/s12984-017-0290-6>
- Fortini, L., Lorenzini, M., Kim, W., De Momi, E., & Ajoudani, A. (2020). A Real-time Tool for Human Ergonomics Assessment based on Joint Compressive Forces. *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 1164–1170. <https://doi.org/10.1109/RO-MAN47096.2020.9223565>
- Hossain, M. S., Azam, S., Karim, A., Montaha, S., Quadir, R., De Boer, F., & Altaf-Ul-Amin, M. (2023). Ergonomic Risk Prediction for Awkward Postures from 3D Keypoints Using Deep Learning. *IEEE Access*.
- Janowitz, I. L., Gillen, M., Ryan, G., Rempel, D., Trupin, L., Swig, L., Mullen, K., Rugulies, R., & Blanc, P. D. (2006). Measuring the physical demands of work in hospital settings: Design and implementation of an ergonomics assessment. *Applied Ergonomics*, 37(5), 641–658. <https://doi.org/10.1016/j.apergo.2005.08.004>
- Kotsiantis, S. B. (2007). Supervised machine learning: A review of classification techniques. *Informatica (Ljubljana)*, 31(3), 249–268.
- Kukreja, H., Bharath, N., Siddesh, C. S., & Kuldeep, S. (2016). An introduction to artificial neural network. *Int J Adv Res Innov Ideas Educ*, 1, 27–30.
- Kumari, S., Tewari, V. K., Kumar, S., & Sahni, R. K. (2023). An Electromyographical Approach to Evaluate the Effect of Load on Agricultural Worker during Push–Pull Operation. *Journal of Biosystems Engineering*, 48(4), 402–411. <https://doi.org/10.1007/s42853-023-00200-1>
- Li, K., Zhang, J., Wang, L., Zhang, M., Li, J., & Bao, S. (2020). A review of the key technologies for sEMG-based human-robot interaction systems. *Biomedical Signal Processing and Control*, 62, 102074–. <https://doi.org/10.1016/j.bspc.2020.102074>
- Li, Z., Yu, Y., Xia, J., Chen, X., Lu, X., & Li, Q. (2024). Data-driven ergonomic assessment of construction workers. *Automation in Construction*, 165, 105561–. <https://doi.org/10.1016/j.autcon.2024.105561>
- Liao, L., Liao, K., Wei, N., Ye, Y., Li, L., & Wu, Z. (2023). A holistic evaluation of ergonomics application in health, safety, and environment management research for construction workers. *Safety Science*, 165, 106198–. <https://doi.org/10.1016/j.ssci.2023.106198>
- Lin, P. C., Chen, Y. J., Chen, W. S., & Lee, Y. J. (2022). Automatic real-time occupational posture evaluation and select corresponding ergonomic assessments. *Scientific Reports*, 12(1), 2139–2139. <https://doi.org/10.1038/s41598-022-05812-9>
- Liu, Q., Liu, Y., Zhang, C., Ruan, Z., Meng, W., Cai, Y., & Ai, Q. (2021). sEMG-Based Dynamic Muscle Fatigue Classification Using SVM With Improved Whale Optimization Algorithm. *IEEE Internet of Things Journal*, 8(23), 16835–16844. <https://doi.org/10.1109/JIOT.2021.3056126>
- Lorenzini, M., Kim, W., De Momi, E., & Ajoudani, A. (n.d.). A Real-time Graphic Interface for the Monitoring of the Human Joint Overloadings with Application to Assistive Exoskeletons. In *Wearable Robotics: Challenges and Trends* (pp. 281–285). Springer International Publishing. https://doi.org/10.1007/978-3-030-01887-0_54
- Lowe, B. D., Weir, P., & Andrews, D. (2014). Observation-based posture assessment: review of current practice and recommendations for improvement. https://stacks.cdc.gov/view/cdc/24085/cdc_24085_DS1.pdf
- Middleworth, M. (2012). A step-by-step guide rapid upper limb assessment (RULA). *Ergonomics Plus*, 13.
- Nayak, G. K., & Kim, E. (2021). Development of a fully automated RULA assessment system based on computer vision. *International Journal of Industrial Ergonomics*, 86, 103218–.

- <https://doi.org/10.1016/j.ergon.2021.103218>
- Ossai, C. I., & Wickramasinghe, N. (2022). GLCM and statistical features extraction technique with Extra-Tree Classifier in Macular Oedema risk diagnosis. *Biomedical Signal Processing and Control*, 73, 103471–. <https://doi.org/10.1016/j.bspc.2021.103471>
- Park, J., Kim, Y., & Han, B. (2018). Work Sectors with High Risk for Work-Related Musculoskeletal Disorders in Korean Men and Women. *Safety and Health at Work*, 9(1), 75–78. <https://doi.org/10.1016/j.shaw.2017.06.005>
- Paul, Y., Goyal, V., & Jaswal, R. A. (2017). Comparative analysis between SVM & KNN classifier for EMG signal classification on elementary time domain features. *2017 4th International Conference on Signal Processing, Computing and Control (ISPCC)*, 2017–, 169–175. NEW YORK: IEEE. <https://doi.org/10.1109/ISPCC.2017.8269670>
- Phinyomark, A., Phukpattaranont, P., & Limsakul, C. (2012). Feature reduction and selection for EMG signal classification. *Expert Systems with Applications*, 39(8), 7420–7431. <https://doi.org/10.1016/j.eswa.2012.01.102>
- Reaz, M. B. I., Hussain, M. S., & Mohd-Yasin, F. (2006). Techniques of EMG signal analysis: Detection, processing, classification and applications. *Biological Procedures Online*, 8(1), 11–35. <https://doi.org/10.1251/bpo115>
- Rechy-Ramirez, E. J., & Hu, H. (2011). Stages for Developing Control Systems using EMG and EEG signals: A survey. *School of computer science and electronic engineering, University of Essex*, 1744-8050.
- Reyes-Fernandez, M. C., Posada-Gomez, R., Martinez-Sibaja, A., Aguilar-Lasserre, A. A., & Flores Cuautle, J. J. A. (2024). Wrist EMG Monitoring Using Neural Networks Techniques. *Journal of Sensors*, 2024, 1–9. <https://doi.org/10.1155/2024/5526158>
- Roberts, D., Torres Calderon, W., Tang, S., & Golparvar-Fard, M. (2020). Vision-Based Construction Worker Activity Analysis Informed by Body Posture. *Journal of Computing in Civil Engineering*, 34(4). [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000898](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000898)
- Rodrigues, P. B., Xiao, Y., Fukumura, Y. E., Awada, M., Aryal, A., Becerik-Gerber, B., Lucas, G., & Roll, S. C. (2022). Ergonomic assessment of office worker postures using 3D automated joint angle assessment. *Advanced Engineering Informatics*, 52, 101596-. <https://doi.org/10.1016/j.aei.2022.101596>
- Seo, J., & Lee, S. (2021). Automated postural ergonomic risk assessment using vision-based posture classification. *Automation in Construction*, 128, 103725-. <https://doi.org/10.1016/j.autcon.2021.103725>
- Sharma, D., Kumar, R., & Jain, A. (2022). Breast cancer prediction based on neural networks and extra tree classifier using feature ensemble learning. *Measurement: Sensors*, 24, 100560–. <https://doi.org/10.1016/j.measen.2022.100560>
- Subasi, A. (2013). Classification of EMG signals using PSO optimized SVM for diagnosis of neuromuscular disorders. *Computers in Biology and Medicine*, 43(5), 576–586. <https://doi.org/10.1016/j.combiomed.2013.01.020>
- Ventura, L., Lorenzini, M., Kim, W., & Ajoudani, A. (2021). A Flexible Robotics-Inspired Computational Model of Compressive Loading on the Human Spine. *IEEE Robotics and Automation Letters*, 6(4), 8229–8236. <https://doi.org/10.1109/LRA.2021.3100936>
- Wang, D., Dai, F., & Ning, X. (2015). Risk Assessment of Work-Related Musculoskeletal Disorders in Construction: State-of-the-Art Review. *Journal of Construction Engineering and Management*, 141(6). [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0000979](https://doi.org/10.1061/(ASCE)CO.1943-7862.0000979)
- Yan, X., Li, H., Li, A. R., & Zhang, H. (2017). Wearable IMU-based real-time motion warning system for construction workers' musculoskeletal disorders prevention. *Automation in Construction*, 74, 2–11. <https://doi.org/10.1016/j.autcon.2016.11.007>
- Yu, Y., Li, H., Umer, W., Dong, C., Yang, X., Skitmore, M., & Wong, A. Y. L. (2019a). Automatic Biomechanical Workload Estimation for Construction Workers by Computer Vision and Smart Insoles. *Journal of Computing in Civil Engineering*, 33(3). [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000827](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000827)
- Yu, Y., Li, H., Yang, X., Kong, L., Luo, X., & Wong, A. Y. L. (2019b). An automatic and non-invasive physical fatigue assessment method for construction workers. *Automation in Construction*, 103, 1–12. <https://doi.org/10.1016/j.autcon.2019.02.020>

EFFECTS OF A PASSIVE BACK SUPPORT EXOSKELETON ON BALANCE WHILE WALKING

Masood Khan¹, JoonOh Seo², Maxwell Fordjour Antwi-Afari³, Yuan Zhou⁴

1) Ph.D. Student, Department of Building and Real Estate, Hong Kong Polytechnic University, Hong Kong SAR. Email: masood2.khan@connect.polyu.hk

2) Ph.D., Assoc. Prof., Department of Building and Real Estate, Hong Kong Polytechnic University, Hong Kong SAR. Email: joonoh.seo@polyu.edu.hk

3) Ph.D., Lect., Department of Civil Engineering, College of Engineering and Physical Sciences, Aston University, United Kingdom. Email: m.antwifari@aston.ac.uk

4) Ph.D., Post-Doc Fellow, Department of Building and Real Estate, Hong Kong Polytechnic University, Hong Kong SAR. Email: zhouyuan1117@msn.cn

Abstract: Passive back support exoskeletons (PBSEs) have been promoted as a means of alleviating the physical strain associated with manual tasks in industrial settings. These devices are known to influence the wearer's kinematics, muscle activation, and balance. Slips and trips, which are frequent precursors to falls, often occur during construction tasks. The effects of PBSEs on balance and the ability to recover from slip- and trip-like perturbations during walking have not been thoroughly examined. The present study aimed to investigate the effects of a PBSE on ground reaction forces (GRF) after slip and trip-like perturbations during walking on an instrumented treadmill. Nine male participants walked on an instrumented treadmill under two conditions: without wearing a PBSE (WOE) and after wearing a PBSE (WE). Each participant experienced normal walking, slip, and trip events, presented in a random order. GRFs were recorded using a force plate integrated into the treadmill. Fx (force in the mediolateral direction) was higher ($p = 0.003$) in WE (mean, 166.32 N) than in WOE (mean, 140.52 N) by 18.36 % after slip perturbations. Fy (force in the anteroposterior direction), Fz (force in the vertical direction), and Fr (the resultant force) did not show statistically significant differences between WE and WOE after slip perturbations. Following trip perturbations, a statistically significant increase was observed in Fz ($p < 0.001$) and Fr ($p < 0.001$). Fz and Fr were higher in WE than WOE by 25.24 % and 8.88 %, respectively. Wearing a PBSE may alter the GRF in a mediolateral or vertical direction that may predispose the wearer to fall. Construction workers should be provided with balance training while wearing a PBSE and then exercise caution while walking on a construction site.

Keywords: Fall, Recovery, Slip, Trip, Ground Reaction Forces, Treadmill, Construction.

1. INTRODUCTION

The most common cause of disability among construction workers is work-related musculoskeletal disorders (WMSD) (Millennium 2003, Wang, Dong et al. 2017). The prevalence rates of WMSD were 36%, 68%, 76%, and 41% in Ethiopia (12-month prevalence) (Lette, Ambelu et al. 2018), Taiwan (12-month prevalence) (Leung, Chan and Yu 2012), Malaysia (12-month prevalence) (Deros, Daruis et al. 2014), and Hong Kong (3-month prevalence) (Yi and Chan 2016) respectively. According to a study, more than 77% of American construction workers reported having at least one musculoskeletal ailment in the previous 12 months [7]. WMSDs can cause severe financial hardships and absenteeism in the construction industry in addition to physical suffering (Cheng, Leu et al. 2010). According to Okenwa Emegwa's (2014) research, WMSDs were linked to over 85% of sick leave cases in the Swedish construction sector (Okenwa Emegwa 2014). Workers in construction are subjected to a physical workload that includes heavy lifting, crouching, kneeling, working with hands above shoulder level, and vibration. During construction, adopting non-neutral body positions may raise the chances of acquiring WMSD (Punnett and Wegman 2004, Takala, Pehkonen et al. 2010).

To mitigate WMSDs, the use of an exoskeleton as an additional intervention has gained more attention in recent years (De Looze, Bosch et al. 2016, Antwi-Afari, Li et al. 2021). This may be due to the fact that it is wearable, can support the wearer without requiring modifications to current work environments, and may be used in situations where other approaches are impractical. As a possible occupational intervention to reduce the risk of overexertion

injuries related to manual material handling, passive back-support exoskeletons (PBSEs) are gaining popularity (Nussbaum, Lowe et al. 2019, Kermavnar, de Vries et al. 2021). Because they are less expensive and easier to install in the workplace, PBSEs, as opposed to active ones, are now more developed for application in occupational settings (Nussbaum, Lowe et al. 2019). With a PBSE, the wearer can lower the metabolic expenditure and levels of back muscle activation during a variety of symmetric and asymmetric lifting exercises (Aleml, Geissinger et al. 2019, Baltrusch, Van Dieën et al. 2019, Koopman, Kingma et al. 2020, Anwer, Li et al. 2023).

However, previous research has expressed concern that the usage of a PBSE may have unanticipated (or unwanted) impacts on the wearer (Masood, ANTWI-AFARI et al. 2024). This concern was likely brought on by the external torques that the PBSE created around the hip and back, as a PBSE typically engages when hip flexion occurs (Baltrusch, Van Dieën et al. 2018, Baltrusch, Van Dieën et al. 2019). Exoskeleton parts supporting or encircling the thighs, waist, and chest are inflexible structures present in PBSEs. The masses of commercially available passive PBSEs vary from 2.8 to 4.5 kg. Using a PBSE while doing a holding or lifting action has been linked to an increase in leg muscular activity (Sadler, Graham and Stevenson 2011, Ulrey and Fathallah 2011). The concerns of unwanted effects on wearers have been linked particularly to more rigid movements (Koopman et al., 2019b), reduced ROM (Abdoli-Eramaki, Stevenson et al. 2007), more physical strain in situations that the PBSE is not meant for (von Glinski, Yilmaz et al. 2019), and deviation from the kinematics and anatomy of wearers (Huysamen, Power and O'Sullivan 2018). The inflexible structure and external hip extension torque of the PBSE may impede corrective postural movements, and the device's additional weight may put additional load on the postural control system. Employing PBSEs may be necessary in situations where the wearer must lift, move, carry a load, or ascend or descend stairs. These activities are more likely to cause the wearer to lose their postural balance than stationary PBSE use. The majority of walking-related falls result from outside disruptions that impair equilibrium, such as trips and slips (Berg, Alessio et al. 1997, Heijnen and Rietdyk 2016). Risk factors for trips and falls in work environments include uneven or slick surfaces, dim illumination, and unstable footwear (Afanuh, Anderson and Bell 2012). Given the growing prevalence of PBSE technologies in the workplace (Nussbaum, Lowe et al. 2019) and the fact that slips, trips, and falls remain a major concern in workplaces worldwide, it is imperative to determine whether and to what extent PBSEs negatively impact the postural balance of the wearers. Therefore, this study aimed to assess the effects of a PBSE on ground reaction forces (GRFs) after slip- and trip-like perturbations. The study hypothesized that due to the added mass and movement restriction imposed by PBSE, the GRFs will increase after wearing a PBSE.

2. METHOD

The study was approved by the Human Subjects Ethics Sub-committee (HSESC) of The Hong Kong Polytechnic University (HSESC Reference Number: HSEARS20231101001). A total of 9 participants (mean age: 31.33 years, height: 175.55 cm, weight: 77.11 kg, BMI: 25.05 kg/m²) participated in the study. Only male participants (age 25-45 years) were recruited, as most of the construction workers are males. The recruited participants had no current history of musculoskeletal pain or deformities. Prior to starting the experiment, the risks and benefits of the study were discussed with each participant, and written informed consents were obtained.

2.1. Protocol:

Each participant underwent three conditions:

1. Normal walking: Participants walked at a speed of 3 kilometres per hour (kph).
2. Slip: The treadmill's speed increased from 3 kph to 8 kph.
3. Trip: The treadmill's speed reduced from 3 kph to 0 kph.

An instrumented treadmill integrated with force plates (gaitway® 3D, h/p/cosmos sports & medical, Nussdorf-Traunstein, Germany) was used to induce slip- and trip-like perturbations. Slips and trips were repeated thrice at intervals of 10–30 seconds. First, participants were provided with a description of the experiment. A safety harness was used to prevent participants from falling. The Ottobock back (Ottobock, Germany) exoskeleton was used for the experiment. The appropriate use of this PBSE was demonstrated to each participant. This PBSE had two modes: 'On' (activated) and 'Off' (deactivated). The 'On' mode was used in the experiment. This resulted in two experimental conditions: a. WE (with PBSE): Participants wore the activated PBSE, which means the PBSE supported the wearer's back. b. WOE (without wearing the PBSE): The participant did not wear the PBSE. Figures 1 and 2 show PBSE and the treadmill used in the study. Figure 3 shows the study's complete activity protocol.



Figure 1. The passive back support exoskeleton used in the study.



Figure 2. Participant walking on the treadmill integrated with force plates.

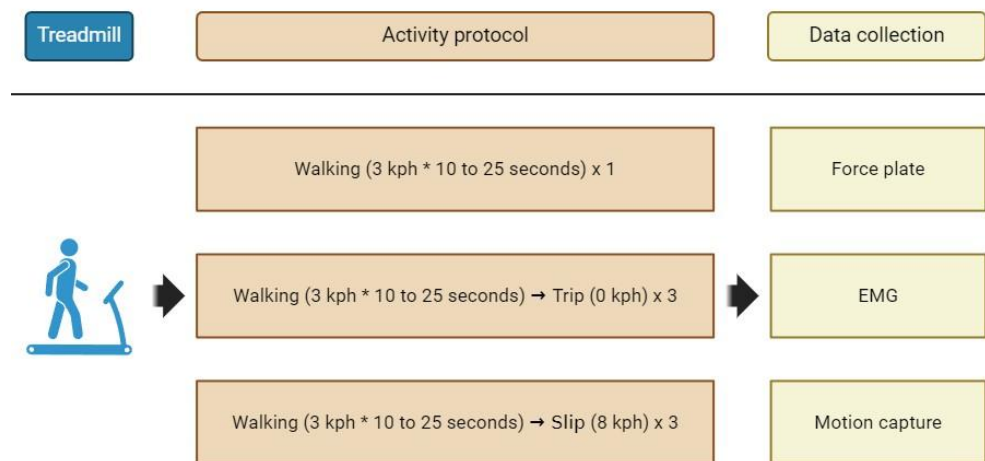


Figure 3. The activity protocol used in the study.

2.2. Statistical analysis

The GRF data were collected at a frequency of 2000 Hz. The raw data was low-pass filtered (4th order zero-phase-shift Butterworth) at 300 Hz using the application MATLAB (Version R2024a, The MathWorks, Inc., US). Normal distribution was assessed using the Shapiro-Wilk test of normality. This test revealed that data was not normally distributed. Therefore, a non-parametric test, i.e. the Mann-Whitney U test, was used to compare WE and WOE conditions using SPSS software (version 29). The results were considered significant for $p \leq 0.05$. The mean of the first three steps after slip and trip events were used for analysis.

The data were divided into two groups: Slip and Trip. In each group, two conditions were created and compared: With the PBSE (represented as WE) and without any PBSE (represented as WOE). A statistically significant difference was considered with $p \leq 0.05$. Three components of GRFs, i.e. F_x , F_y , F_z and F_r were compared between WE and WOE conditions. F_x represented the horizontal force exerted by the ground on a body in the mediolateral direction, F_y represented the horizontal force exerted by the ground on a body in the anteroposterior direction, F_z

represented the vertical force exerted by the ground on a body and Fr represented the resultant force of these three forces.

3. RESULTS

Table 1 includes the GRF descriptive data for WE and WOE conditions after slip and trip events. Table 2 shows z and p-values obtained after performing the Mann-Whitney U test to compare both exoskeleton conditions (WE vs. WOE). Figures 4, 5, 6, and 7 depict the amplitudes of Fx, Fy, Fz, and Fr after slip and trip perturbations.

Table 1. Ground reaction forces (GRF) after slip and trip perturbations.

GRF component	Slip		Trip	
	WE	WOE	WE	WOE
F _x (N)	166.32±69.84	140.52±74.56	123.46±47.29	112.30±44.91
F _y (N)	203.22±79.49	184.04±55.21	185.24±61.03	182.21±57.91
F _z (N)	1327.12±243.69	1262.93±209.44	1219.41±198.81	973.65±376.46
Fr (N)	1355.86±249.64	1286.85±212.14	1241.12±203.88	1139.89±165.97

WE: With Exoskeleton; WOE: Without Exoskeleton; GRF: Ground reaction forces.

Table 2. Comparison of exoskeleton conditions (WE vs. WOE).

GRF component	Slip		Trip	
	Z	p-value	Z	p-value
F _x	-2.979	0.003*	-1.796	0.073
F _y	-.810	0.418	-0.474	0.635
F _z	-1.490	0.136	-4.898	<0.001*
Fr	-1.505	0.132	-3.360	<0.001*

*Significant $p \leq 0.05$

WE: With Exoskeleton; WOE: Without Exoskeleton;

GRF: Ground reaction forces.

A statistically significant difference ($p=0.003$) was observed in F_x after slip perturbations. F_x (i.e. force in the mediolateral direction) was 18.36% higher in WE (mean, 166.32) than in WOE (mean, 140.52). F_y, F_z, and Fr increased in WE after slip perturbation; however, these increments were not statistically significant ($p>0.05$).

A statistically significant increase was observed in F_z ($p<0.001$) and Fr ($p<0.001$) after trip perturbation. F_z was higher in WE (mean, 1219.41) than WOE (mean, 973.65) by 25.24 %. The Fr was also higher in WE (mean, 1241.12) than WOE (mean, 1139.89) by 8.88 %.

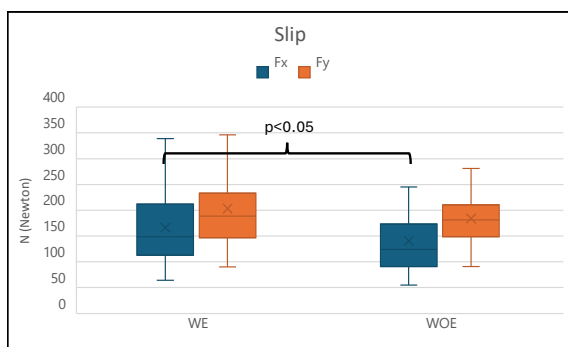


Figure 4. Comparison of exoskeleton conditions after slip perturbation: F_x and F_y. WE: With Exoskeleton; WOE: Without Exoskeleton.

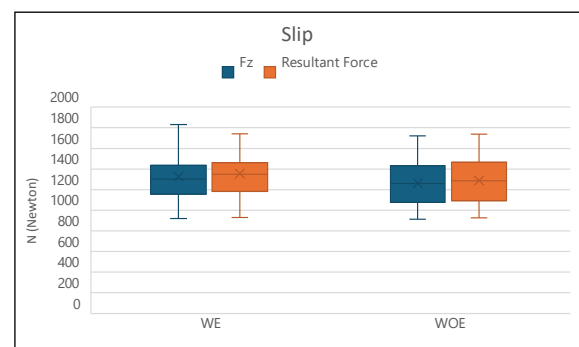


Figure 5. Comparison of exoskeleton conditions after slip perturbation: F_z and resultant force. WE: With Exoskeleton; WOE: Without Exoskeleton.

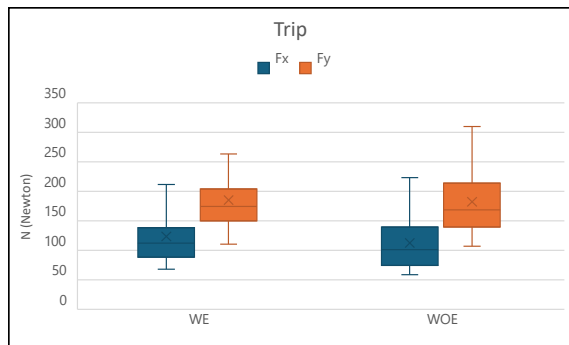


Figure 6. Comparison of exoskeleton conditions after trip perturbation: Fx and Fy.
WE: With Exoskeleton; WOE: Without Exoskeleton.

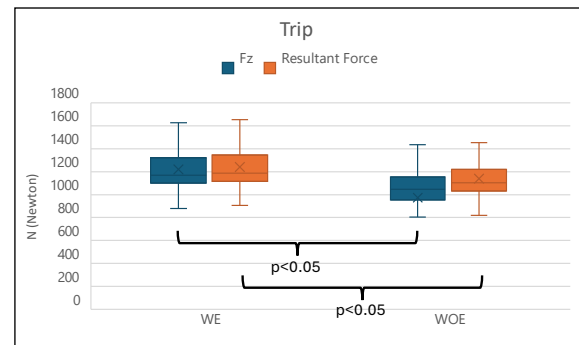


Figure 7. Comparison of exoskeleton conditions after trip perturbation: Fz and resultant force.
WE: With Exoskeleton; WOE: Without Exoskeleton.

4. DISCUSSION

Walking is an important activity performed by the construction workers at the construction sites. A PBSE may raise the risk of falls by negatively impacting gait (Park, Kim et al. 2022) and reactions to significant postural disturbances (Park, Lee et al. 2022, Dooley, Kim et al. 2023). There are concerns that PBSEs will affect the balance of the wearers. Particularly, these concerns have been linked to more rigid movements (Koopman, Kingma et al. 2019), reduced ROM (Abdoli-Eramaki, Stevenson et al. 2007), more physical strain in situations that the PBSE is not meant for (von Gliniski, Yilmaz et al. 2019), and deviation from the kinematics and anatomy of humans (Huysamen, Power and O'Sullivan 2018). This study aimed to investigate the effects of a PBSE on balance after slip and trip-like perturbations. The study hypothesized that PBSE would adversely affect the recovery following slip and trip perturbations. The slip and trip perturbations were induced on an instrumented treadmill, and the balance was measured using GRF data from the force plate. The results of the present study showed changes in the GRF components (Fx, Fy, and Fz) between WE and WOE conditions after slip and trip perturbations. After slip perturbations, Fx was higher in WE compared to WOE; however, Fy, Fz, and Fr were not statistically different from WOE. After trip perturbations, the Fz and the Fr were higher in WE compared to WOE. Fx and Fy were not statistically different from WOE.

A recent study performed by Dooley et al. (Dooley, Kim et al. 2024) to assess the impact of arm and PBSE on changes in reactive balance following perturbations also reported adverse changes in recovery kinematics. However, they reported that these passive exoskeletons do not increase the probability of falls. In the present study, the PBSE increased the mediolateral as well as vertical GRF. The net Fr were also higher after wearing the PBSE than without wearing it. A possible reason for these changes could be that after wearing a PBSE, weight is added to the wearer's body. This added mass may increase the weight transfer time from one leg to another while walking, thereby resulting in increased GRF. The PBSE restricts the movements at the back and hip joints; therefore, the wearer may have to make compensatory movements for stability after perturbations. Due to these compensatory movements, the wearers' bodies may exert more force to navigate the restrictions at these joints. Another possible reason could be that the PBSE's function is to assist the wearer in lifting an object from the ground. Therefore, it applies hip extension force. This hip extension torque counters the hip flexion torque required to recover following the slips and trip perturbations (Park, Lee et al. 2022). Since hip flexion torque is difficult, the wearer's body may shift his/her body weight laterally. There are a few limitations worth acknowledging. First, the results in the present study were obtained from a controlled environment (laboratory); however, the actual construction sites are different from laboratories. At the actual construction sites, there are objects, machines or other workers on the path. Moreover, the participants in the present study did not fully cover their bodies. In addition to these, the construction workers wear PPEs, including hard hats, protective clothing, high visibility clothing, etc., which may affect the functioning of PBSE and thus the response of the workers to slip or trip events after wearing the PBSE. Therefore, the results of the present study should be interpreted in the context of a laboratory.

5. CONCLUSIONS

Passive exoskeletons have been recommended to use to ease the construction tasks. Slips and trips occur frequently at construction sites that may result in falls and can cause injuries. The effect of a PBSE on balance while walking has not been fully investigated. This study aimed to evaluate the effects of a PBSE on balance after slip-and-trip-like perturbations. A total of 9 male participants participated in the study. An instrumented treadmill with integrated force plates was used to induce the slip- and trips and to collect the GRF data. During the slip perturbation, the treadmill speed increased and then became normal without the participants knowing, so they had to take a few quick steps to balance themselves. During the trip perturbation, the treadmill suddenly stopped and then resumed normally without the participant's knowledge. The participants were not told the timings of the perturbations so that their natural responses could be achieved.

The PBSE was found to affect the mediolateral and vertical components of the GRFs. Interestingly, the anteroposterior component was not significantly different between wearing and not wearing a PBSE. The possible reason for no change in anteroposterior GRF after wearing the PBSE could be the movement restrictions imposed in the sagittal plane. The hip flexion was restricted. Therefore, the participants might have put less force in this direction. Due to limitations in the sagittal plane, the participant might have shifted their weight laterally, resulting in increased GRF in the mediolateral direction. Also, since the wearer's weight increased after wearing a PBSE and they had difficulty freely moving their lower extremities, their weight transfer time from one leg to another could have increased, increasing vertical GRF. The findings suggest that balance training should be provided to construction workers while wearing a PBSE and should be trained to walk so that they can adopt precautions and strategies to offset the increased GRFs and avoid falling. It will be advisable to use passive exoskeletons that are lighter and offer fewer movement restrictions at the back and hip joints.

ACKNOWLEDGMENTS

Kan Tong Po International Fellowship 2024 (KTP\R1\241112) for funded project titled "Human-exoskeleton-collaboration (HEC) in construction: An objective assessment of spinal biomechanics and physiological measurements"

REFERENCES:

- Abdoli-Eramaki, M., J. M. Stevenson, S. A. Reid and T. J. Bryant (2007). "Mathematical and empirical proof of principle for an on-body personal lift augmentation device (PLAD)." *Journal of biomechanics* **40**(8): 1694-1700.
- Afanuh, S., V. P. Anderson and J. Bell (2012). "Preventing slips, trips, and falls in wholesale and retail trade establishments."
- Alemi, M. M., J. Geissinger, A. A. Simon, S. E. Chang and A. T. Asbeck (2019). "A passive exoskeleton reduces peak and mean EMG during symmetric and asymmetric lifting." *Journal of Electromyography and Kinesiology* **47**: 25-34.
- Antwi-Afari, M. F., H. Li, S. Anwer, D. Li, Y. Yu, H.-Y. Mi and I. Y. Wuni (2021). "Assessment of a passive exoskeleton system on spinal biomechanics and subjective responses during manual repetitive handling tasks among construction workers." *Safety science* **142**: 105382.
- Anwer, S., H. Li, M. Abdul-Rahman and M. F. Antwi-Afari (2023). *Development and evaluation of a low-cost passive wearable exoskeleton system for improving safety and health performance of construction workers: A pilot study*. ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction, IAARC Publications.
- Baltrusch, S., J. Van Dieën, S. Bruijn, A. Koopman, C. Van Bennekom and H. Houdijk (2019). "The effect of a passive trunk exoskeleton on metabolic costs during lifting and walking." *Ergonomics*.
- Baltrusch, S., J. Van Dieën, C. Van Bennekom and H. Houdijk (2018). "The effect of a passive trunk exoskeleton on functional performance in healthy individuals." *Applied ergonomics* **72**: 94-106.
- Berg, W. P., H. M. Alessio, E. M. Mills and C. Tong (1997). "Circumstances and consequences of falls in independent community-dwelling older adults." *Age and ageing* **26**(4): 261-268.
- Cheng, C.-W., S.-S. Leu, C.-C. Lin and C. Fan (2010). "Characteristic analysis of occupational accidents at small construction enterprises." *Safety science* **48**(6): 698-707.
- De Looze, M. P., T. Bosch, F. Krause, K. S. Stadler and L. W. O'sullivan (2016). "Exoskeletons for industrial application and their potential effects on physical work load." *Ergonomics* **59**(5): 671-681.

- Deros, B. M., D. D. Daruis, N. K. Khamis, D. MOŞHAMAD, S. F. M. Daud, S. M. Amdan, R. Abd Aziz and N. Jamal (2014). "Prevalence of work related musculoskeletal disorders Symptoms among construction workers: a case study in Malaysia." *Iranian Journal of Public Health* **43**(Supple 3): 53-57.
- Dooley, S., S. Kim, M. A. Nussbaum and M. L. Madigan (2023). "A passive leg-support exoskeleton adversely affects reactive balance after simulated slips and trips on a treadmill." *Journal of Biomechanics* **151**: 111533.
- Dooley, S., S. Kim, M. A. Nussbaum and M. L. Madigan (2024). "Occupational arm-support and back-support exoskeletons elicit changes in reactive balance after slip-like and trip-like perturbations on a treadmill." *Applied Ergonomics* **115**: 104178.
- Heijnen, M. J. H. and S. Rietdyk (2016). "Falls in young adults: Perceived causes and environmental factors assessed with a daily online survey." *Human movement science* **46**: 86-95.
- Huysamen, K., V. Power and L. O'Sullivan (2018). "Elongation of the surface of the spine during lifting and lowering, and implications for design of an upper body industrial exoskeleton." *Applied ergonomics* **72**: 10-16.
- Kermavnar, T., A. W. de Vries, M. P. de Looze and L. W. O'Sullivan (2021). "Effects of industrial back-support exoskeletons on body loading and user experience: An updated systematic review." *Ergonomics* **64**(6): 685-711.
- Koopman, A. S., I. Kingma, M. P. de Looze and J. H. van Dieën (2020). "Effects of a passive back exoskeleton on the mechanical loading of the low-back during symmetric lifting." *Journal of biomechanics* **102**: 109486.
- Koopman, A. S., I. Kingma, G. S. Faber, M. P. de Looze and J. H. van Dieën (2019). "Effects of a passive exoskeleton on the mechanical loading of the low back in static holding tasks." *Journal of biomechanics* **83**: 97-103.
- Lette, A., A. Ambelu, T. Getahun and S. Mekonen (2018). "A survey of work-related injuries among building construction workers in southwestern Ethiopia." *International journal of industrial ergonomics* **68**: 57-64.
- Leung, M.-y., I. Y. S. Chan and J. Yu (2012). "Preventing construction worker injury incidents through the management of personal stress and organizational stressors." *Accident Analysis & Prevention* **48**: 156-166.
- Masood, K., M. F. ANTWI-AFARI, S. JoonOh, S. ANWER and K. HEUNG (2024). *Applications, Challenges, and Future Research Directions for Passive Exoskeletons in the Construction Industry: A Critical Review*. International conference on construction engineering and project management, Korea Institute of Construction Engineering and Management.
- Millennium, W. S. G. o. t. B. o. M. C. a. t. S. o. t. N. (2003). "The burden of musculoskeletal conditions at the start of the new millennium." *World Health Organization technical report series* **919**: i.
- Nussbaum, M. A., B. D. Lowe, M. de Looze, C. Harris-Adamson and M. Smets (2019). An introduction to the special issue on occupational exoskeletons, Taylor & Francis. **7**: 153-162.
- Okenwa Emegwa, L. (2014). "Determinants of sick leave duration following occupational injuries among workers in the county of Gävleborg, Sweden." *Occupational medicine & health affairs* **2**(4).
- Park, J.-H., S. Kim, M. A. Nussbaum and D. Srinivasan (2022). "Effects of back-support exoskeleton use on gait performance and stability during level walking." *Gait & Posture* **92**: 181-190.
- Park, J.-H., Y. Lee, M. L. Madigan, S. Kim, M. A. Nussbaum and D. Srinivasan (2022). "Wearing a back-support exoskeleton impairs single-step balance recovery performance following a forward loss of balance—An exploratory study." *Journal of Biomechanics* **144**: 111352.
- Punnett, L. and D. H. Wegman (2004). "Work-related musculoskeletal disorders: the epidemiologic evidence and the debate." *Journal of electromyography and kinesiology* **14**(1): 13-23.
- Sadler, E. M., R. B. Graham and J. M. Stevenson (2011). "The personal lift-assist device and lifting technique: a principal component analysis." *Ergonomics* **54**(4): 392-402.
- Takala, E.-P., I. Pehkonen, M. Forsman, G.-Å. Hansson, S. E. Mathiassen, W. P. Neumann, G. Sjøgaard, K. B. Veiersted, R. H. Westgaard and J. Winkel (2010). "Systematic evaluation of observational methods assessing biomechanical exposures at work." *Scandinavian journal of work, environment & health*: 3-24.
- Ulrey, B. L. and F. A. Fathallah (2011). *Biomechanical effects of a personal weight transfer device in the stooped posture*. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, SAGE Publications Sage CA: Los Angeles, CA.
- von Glinski, A., E. Yilmaz, S. Mrotzek, E. Marek, B. Jettkant, A. Brinkemper, C. Fisahn, T. A. Schildhauer and J. Geßmann (2019). "Effectiveness of an on-body lifting aid (HAL® for care support) to reduce lower back muscle activity during repetitive lifting tasks." *Journal of Clinical Neuroscience* **63**: 249-255.
- Wang, X., X. S. Dong, S. D. Choi and J. Dement (2017). "Work-related musculoskeletal disorders among construction workers in the United States from 1992 to 2014." *Occupational and environmental medicine* **74**(5): 374-380.
- Yi, W. and A. Chan (2016). "Health profile of construction workers in Hong Kong." *International journal of environmental research and public health* **13**(12): 1232.

BACK EXOSKELETONS FOR CONSTRUCTION WORKERS: INVESTIGATION OF HUMAN-EXOSKELETON INTERACTION FORCES

Ting Lei¹, Kelvin Holam Heung², and Joon Oh Seo³

1) Ph.D. Student, Department of Building and Real Estate, The Hong Kong Polytechnic University, Hong Kong. Email: tinga.lei@connect.polyu.hk

2) Ph.D., Research Assistant Prof., Department of Building and Real Estate, The Hong Kong Polytechnic University, Hong Kong. Email: kelvin.heung@polyu.edu.hk

3) Ph.D., Assoc. Prof., Department of Building and Real Estate, The Hong Kong Polytechnic University, Hong Kong. Email: joonoh.seo@polyu.edu.hk

Abstract: Low back pain (LBP) is a prevalent issue among construction workers, with prevalence rates ranging from 27% to 52%. This significant burden not only affects worker health but also incurs substantial economic costs, exemplified by Brazil's annual expenditure of approximately USD \$500 per worker and Spain's EUR €8945.6 million in related costs. To mitigate LBP, back exoskeletons have emerged as promising solutions, designed to reduce low back load during repetitive lifting tasks. This study compares the human-exoskeleton interaction forces in three back exoskeletons - SV Exosuit (a soft active exosuit integrated with a safety vest), Laevo V2 (a rigid passive exoskeleton), and MATE-XB (also rigid and passive) - through experimental tasks involving bending and squatting. Three healthy male participants performed these tasks while wearing each exoskeleton, during which human-exoskeleton interaction forces at the thigh and shoulder/chest were measured. Results demonstrated that SV Exosuit produced higher contact forces at both body regions, attributed to its smaller contact area and less cushioning material. Conversely, Laevo V2 exhibited two peaks in contact forces during a motion cycle due to its torque generation mechanism, highlighting the influence of supporting torque design on user comfort. This research underscores the critical need for optimizing exoskeleton designs to enhance comfort and usability in construction settings. Future studies should investigate a larger sample size and additional body regions to comprehensively assess the relationship between supportive torque, contact forces, joint angle, and user comfort.

Keywords: Back Exoskeleton, Interaction Force, User Discomfort, Construction Health, Low Back Pain, Wearable Robots

1. INTRODUCTION

1.1 Prevalence Of Low Back Pain

Construction workers usually perform long period and high physical demanding tasks, causing a high incidence of low back pain (LBP) (Adhikari et al., 2021; Vasiwala et al., 2021). Safety records indicate that the prevalence of low back injuries, LBP, and low back disorders among construction workers ranges from 27% to 52% (Adhikari et al., 2021; Kashif et al., 2022; Wang et al., 2017). The financial burden on both workers and managers is significant, encompassing medical expenses and losses due to early retirement and absenteeism. For instance, Brazil spends approximately USD \$500 annually on LBP treatment (Carregaro et al., 2020), while Spain incurs EUR €8945.6 million in annual costs (Alonso-García & Sarria-Santamera, 2020).

Research highlights a strong association between LBP and heavy workloads (Burdorf & Jansen, 2006; Miranda et al., 2008; Serranheira et al., 2020). Therefore, reducing low back load is crucial for decreasing LBP incidence. However, replacing workers with automated machines poses challenges due to the complexity of construction tasks, which require a high degree of physical dexterity in dynamic and confined environments (Wu et al., 2022). Consequently, physical demands remain unavoidable for construction workers.

1.2 Back Exoskeletons To Reduce Low Back Load

Back-support exoskeletons, or known as back exoskeletons, are wearable devices designed to reduce low back load during repetitive lifting tasks (Koopman et al., 2020). These devices support industrial workers by compensating for the voluntary efforts of low back muscles through providing external support (Kranenborg et al., 2023). Back exoskeletons can be classified as active (utilizing electric motors or pneumatic actuators) or passive (using springs and elastic straps) based on their actuation mechanism (Matthew et al., 2015). They can also be categorized as rigid (made from metal or carbon) or soft (constructed from fabrics and textiles) according to their structural materials (Schwartz et al., 2021). In general, back exoskeletons can provide assistive force of up to 500 N or assistive torque of up to 60 Nm (Koopman et al., 2020; Luger et al., 2021). Previous studies confirm that back exoskeletons can reduce back muscle load by 10% – 40% (Poliero et al., 2020; Walter et al., 2023), potentially lowering the risk of low back disorders by approximately 20% (Zelik et al., 2022). Some studies also indicate that back exoskeletons may enhance productivity. For example, Bennett et al. (2023) found that when performing construction tasks (specifically, pushing and emptying a construction gondola and installing and removing wooden blocks between metal studs), the workers completed 5.3% faster on average with Herowear Apex (HeroWear,

Nashville, USA).

Despite the potential benefits, limitations such as discomfort and motion impediment can hinder the broader application of back exoskeletons (Golabchi et al., 2023; Luger et al., 2021). For example, subjects in the study of Golabchi et al. (2023) reported increased discomfort in the chest, arms, knees, and upper legs when using a passive exoskeleton. Similarly, Kim et al. (2020) noted moderate discomfort in the chest, waist, and thighs with two passive exoskeletons. Discomfort levels can become unacceptable after two hours of use (Hensel & Keil, 2019). Therefore, mitigating discomfort is essential for increasing the adoption of back exoskeletons in construction. Research on human-exoskeleton interaction forces is crucial for improving user comfort and effectiveness. While some studies have explored interaction forces in lower limb exoskeletons (Serrancoli et al., 2019; Shushtari & Arami, 2023; Wang et al., 2020), discomfort in back exoskeletons has primarily been assessed through subjective feedback (Hensel & Keil, 2019; Kim et al., 2020; Luger et al., 2021). Thus, investigating the interaction forces between users and back exoskeletons is necessary for wider application in construction. This study aims to preliminarily investigate interaction forces at the thighs and shoulders/chest of three representative back exoskeletons, comparing the relationships between these forces, supportive torque, and contact materials.

2. METHOD

2.1 Subjects And Exoskeletons

A convenience sample of three healthy young male participants was recruited from university students, with a mean (standard deviation, SD) age, body mass, and height of 27 (0.8) years, 79.7 (9.8) kg, and 177.3 (6.0) cm. Participants were not trained for labor-intensive work and were unfamiliar with back-support devices. Selection criteria included: (a) no history of pathological low back pain, and (b) no injuries or disorders affecting trunk lifting tasks. These criteria were set as the compared devices aim to prevent low back injury rather than assist with rehabilitation, and the injuries or medical histories might affect the task performance. The participants provided written informed consent prior to experiments.

Three representative back exoskeletons were compared in this study. Exoskeleton 1, referred to as the SV Exosuit [Fig 1(a)], is a soft active back exosuit that is integrated with a safety vest. It is based on the design previously proposed by our team (Lei et al., 2024) and provides supportive force through eight pneumatic artificial muscles (PAMs) located in the back region. Upon detecting trunk bending, the integrated motion sensor triggers the system to provide compressed gas to the PAMs, causing them to contract. The contraction force is transmitted through shoulder and thigh straps to the wearer. Exoskeleton 2, Laevo V2 (Laevo, Rijswijk, the Netherlands) [Fig 1(b)], is a rigid passive back exoskeleton that generates supportive force through two spring-based torque generators positioned next to the hips. When the wearer bends the trunk, these torque generators are compressed, producing elastic forces that are transferred via a chest pad and two thigh pads to the wearer. Exoskeleton 3, MATE-XB (Comau, Turin, Italy) [Fig 1(c)], is also a rigid passive back exoskeleton that operates similarly to Laevo V2 but transfers the load through shoulder straps and thigh pads. These three exoskeletons were selected for their different designs: SV Exosuit is a soft and active exosuit that transfers load with shoulder straps and thigh straps, MATE-XB is a rigid and passive exoskeleton that transfers load with shoulder straps and thigh pads, while Laevo V2 is a rigid and passive that transfers load with chest pad and thigh pads. The variety allows for a thorough examination of the relationships between the human-exoskeleton interaction forces and the exoskeleton design.

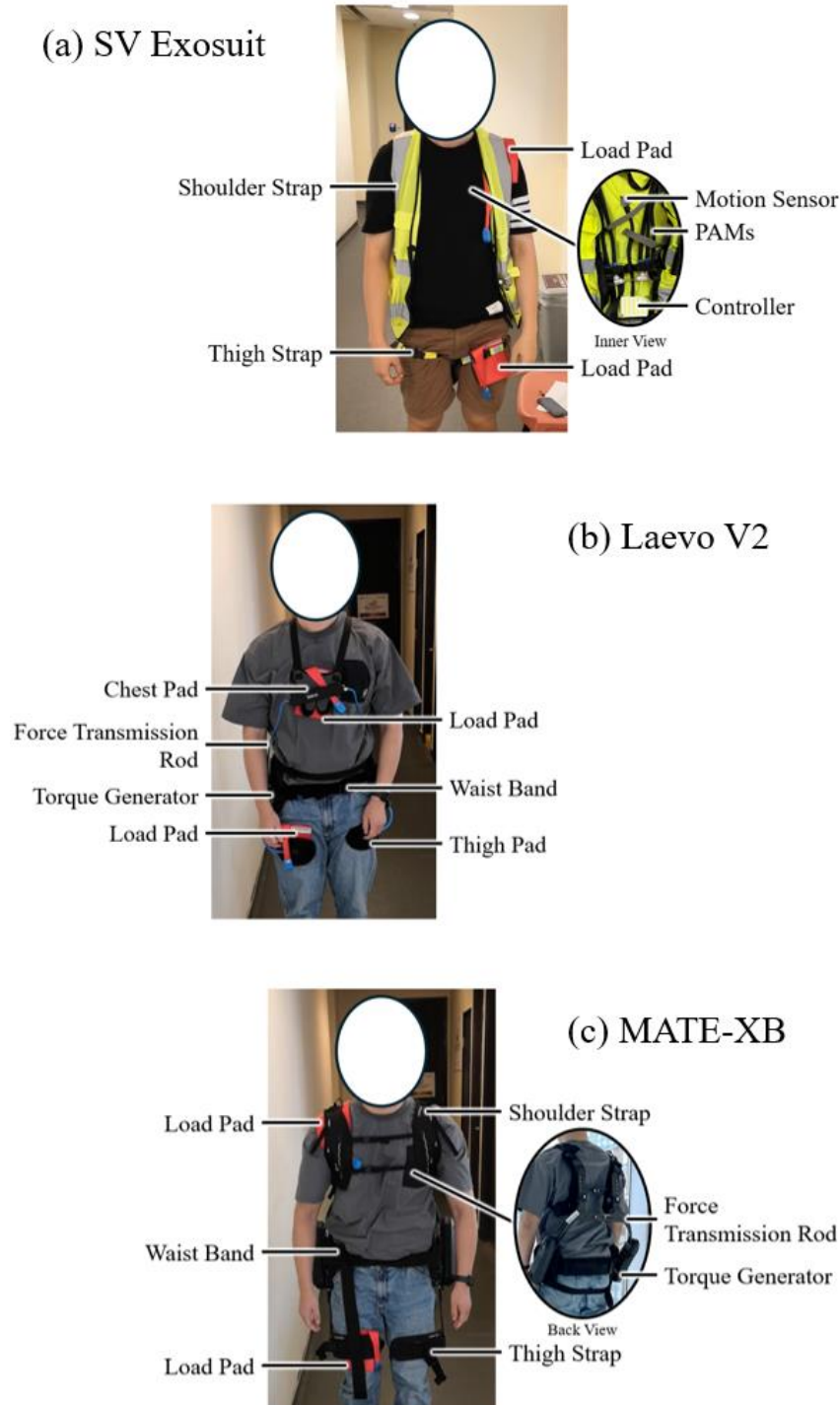


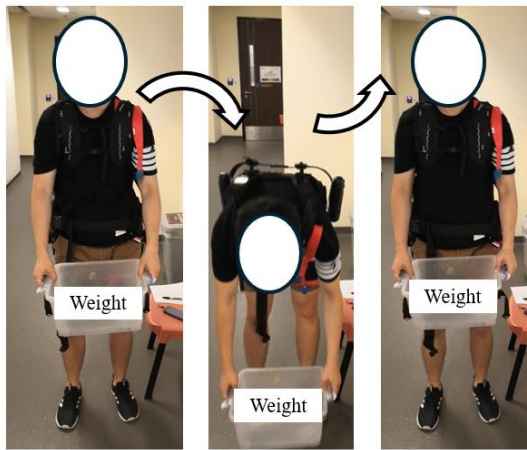
Fig 1. Main structures of the three compared exoskeletons and the installation positions of the load pads.

2.2 Experimental Design

The experiment includes two tasks: bending task and squatting task. In the bending task, as shown in Fig 2(a), participants held a weight while standing, then bent to place the weight on the floor without bending their knees, and finally lifted the weight back to a standing position. In the squatting task, as shown in Fig 2(b), participants held a weight while standing, squatted to place the weight on the floor, and then performed a squat-lift with the weight to return to a standing position. Each exoskeleton condition was tested with both tasks, repeated five times with a 5 kg weight and five times with a 15 kg weight. The sequence of these four sessions within one exoskeleton condition was randomized. The order of these sessions was randomized. Each of the three exoskeletons was tested in a randomized sequence, and participants were allowed three to five minutes to acclimate

to each exoskeleton before performing the tasks.

(a) Bending Task



(b) Squatting Task



Fig 2. Illustrative graphs of (a) bending task and (b) squatting task.

2.3 Data Collection And Analysis

The human-exoskeleton interaction forces were measured at two body parts, the thigh and the shoulder (for SV Exosuit and MATE-XB) or the chest (for Laevo V2), as these areas are commonly reported to experience the highest discomfort when using back exoskeletons (Alemi et al., 2020; Kim et al., 2020; Poliero et al., 2022). Interaction forces were collected using flexible and stretchable load pads (novel GmbH, Munich, Germany) at a sampling rate of 100 Hz. The load pads were positioned between the target body regions (i.e. shoulder/chest and thigh) and the corresponding parts of the exoskeleton (Fig 1). Given that both bending and squatting tasks were symmetrically performed, and no significant differences were observed in muscle loads during symmetric tasks with back exoskeletons (Alemi et al., 2019), load pads were placed on only one side of each participant in this study. The collected contact force data were normalized from time frame to motion cycle frame for comparison, with one motion cycle representing one bending or squatting task. Outliers were removed from the analysis.

3. RESULTS

The compared contact forces are presented in Fig 3. Generally, the contact forces at both the thigh and shoulder/chest regions were higher with SV Exosuit compared to the other two devices, likely due to its contact design. The shoulder straps of SV Exosuit are approximately 7 cm wide, narrower than the 10 cm width of the MATE-XB shoulder straps and the 12 cm width of the Laevo V2 chest pad. Also, the thigh straps of SV Exosuit (about 3 cm width) have smaller contact areas than the thigh pads of the other two devices (about 12 cm width). A smaller contact area and less cushioning material can significantly increase the human-exoskeleton interaction force (Kozinc et al., 2021), indicating that adequate contact area and cushioning is essential in exoskeleton design. Further research is needed to explore the relationship between user comfort and the contact materials used in exoskeletons for the optimal material selection. The design of the clothing components also influences the contact force. Compared to SV exosuit, the shoulder straps of MATE-XB are connected with two straps across the chest region, allowing the load partly distributed to the chest. Also, there is no waist band in the design of SV Exosuit, while both MATE-XB and Laevo V2 incorporate a waist band. The waist band potentially distribute part of the load at the shoulders/chest and the thighs to the waist region. This suggests that effective load distribution through design may be beneficial. Moreover, the contact force is related to the supportive force provided by the exoskeleton. For instance, it is estimated that SV Exosuit can generate up to about 54 Nm assistive torque (Lei et al., 2024), surpassing the 30 Nm maximum of Laevo V2 (Van Harmelen et al., 2022), resulting in greater applied forces on the body. Further investigation is necessary to analyze the relationship between the human-exoskeleton interaction forces and the assistive torque provided by the back exoskeleton. Further, in most fractions of the motion cycle, the contact force at the thigh exceeded that at the shoulder/chest, likely due to the moment arm from the waist to the shoulder/chest being nearly twice as long as that to the thigh. In Laevo V2, the shorter moment arm from the chest pad to the waist resulted in a smaller difference in contact forces between the chest and thigh.

In the bending task, the contact forces at the thigh remained consistent regardless of whether participants held 5 kg or 15 kg, while shoulder/chest forces slightly increased with the heavier weight. This indicates that in the bending task, as load increases, the supportive forces of these three back exoskeletons tend to exert more influence on the upper body rather than the lower limbs. Notably, both SV Exosuit and MATE-XB displayed a single peak in contact forces during the task, whereas Laevo V2 exhibited two peaks. This discrepancy arises because the supportive torques of SV Exosuit and MATE-XB correlate positively with the bending angle, while Laevo V2 reaches maximum torque at a bending angle of $40^{\circ} - 50^{\circ}$ (Van Harmelen et al., 2022). Distributing peak contact forces throughout the motion cycle may impact perceived discomfort, although further study is required to understand the extent and nature of these effects. Comparing the contact force curves of SV Exosuit and MATE-XB, the peak force for SV Exosuit (60% – 70% of motion cycle) occurred later than that for MATE-XB (40% – 60 % of motion cycle). This difference is attributed to how each device supports torque: MATE-XB generates torque based on spring compression in direct response to bending angles, while SV Exosuit provides torque by actively injecting compressed gas following a specific bending angle, resulting in a delayed peak contact force. The contact force variation of these three back exoskeletons suggest that the design of the supporting torque affects the human-exoskeleton interaction forces, which may also have impact on user comfort.

In the squatting task, as in the bending task, contact forces at the shoulder/chest region increased with heavier weights for all three exoskeletons. However, unlike the bending task, contact forces at the thigh increased with heavier weights for the SV Exosuit and Laevo V2. This variation is primarily due to differences in joint angles during these activities. The waist joint angle varies minimally during both tasks ($0^{\circ} - 70^{\circ}$ for the bending task and $0^{\circ} - 60^{\circ}$ for the squatting task), while the hip joint angle varies significantly ($-5^{\circ} - 50^{\circ}$ for the bending task and $-5^{\circ} - 90^{\circ}$ for the squatting task) (Hwang et al., 2009). This explains the relatively higher contact forces at the thigh in the squatting task compared to in the bending task. Further study is needed to confirm the relationship between joint angles and contact forces. Notably, with MATE-XB, contact forces at the thigh decreased when participants held heavier weights during the squatting task. This reduction is likely due to the displacement of the thigh pads. As MATE-XB's force transmission components are rigid and the thigh pads are not fixed, increased pressure may lead to the pads shifting toward the knees, resulting in reduced contact forces at the thigh pad due to an increased moment arm. Previous studies have also observed displacement of back exoskeletons during activities, attributed to relative motion and misalignment between the exoskeleton and human anatomy (Näf et al., 2018; Park et al., 2022; Sarkisian et al., 2021).

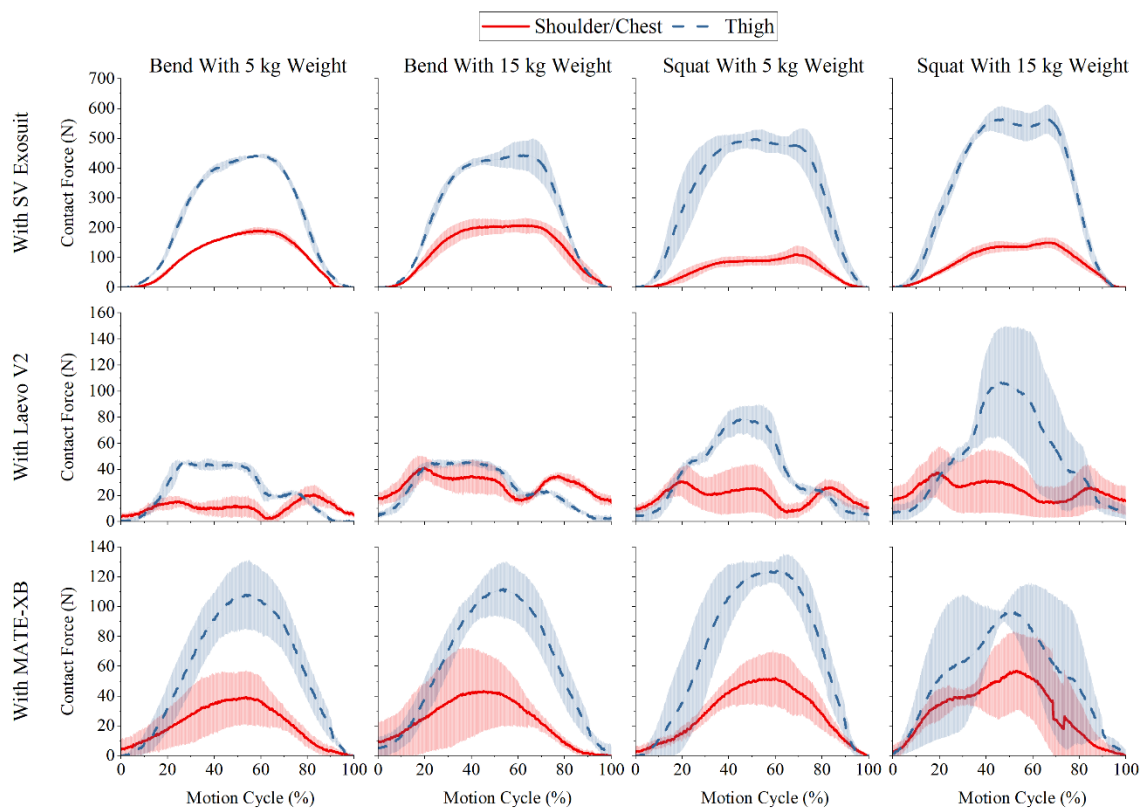


Fig 3. Contact forces in bending and squatting tasks with 5 and 15 kg weight under three back exoskeleton conditions.

4. DISCUSSION

Back exoskeletons are engineered to alleviate the low back load on wearers and reduce the risk of low back pain (LBP). However, the human-exoskeleton interaction force remains a significant source of discomfort, contributing to the low acceptance of these devices in the construction industry (Hensel & Keil, 2019; Levesque et al., 2017). Consequently, investigating the human-exoskeleton interaction force during the activities is needed for improving the exoskeleton design. This study addresses the pressing need to investigate these interaction forces to enhance back exoskeleton design.

The analysis revealed that contact forces at the shoulder/chest and thigh regions were notably higher with the SV Exosuit compared to the other two exoskeletons. This difference may be attributed to the smaller contact area and the use of less cushioning materials (Kozinc et al., 2021). This finding underscores the importance of considering both the shape and material of contact areas in exoskeleton designs, suggesting that improved cushioning could reduce pressure on the wearer. Moreover, design elements such as connecting shoulder straps across the chest might help distribute loads more effectively, although further research is needed to validate these strategies. Optimizing the lengths of the moment arms from the waist to the shoulder/chest and thigh is also crucial. Short moment arms can increase contact forces, while longer moment arms might hinder natural movements like walking and squatting (Park et al., 2022). Thus, achieving a balance in moment arm lengths is essential for effective exoskeleton functionality.

The contact forces exhibited by Laevo V2 varied from the other two devices, primarily due to differences in the relationship between supportive torque and bending angle. Variations in torque strategies between SV Exosuit and MATE-XB also resulted in different peak contact force moments. Results suggest that the variation of supporting torque affects the contact forces. This highlights the need to consider how the distribution of contact forces and peak moments across the motion cycle relate to wearer comfort, a topic that warrants further exploration.

Additionally, the study observed displacement of the thigh pad in the MATE-XB during squatting tasks, indicating potential misalignment between the wearer and the exoskeleton. Previous studies have noted that the relative motion and misalignment between the wearer and the exoskeleton can cause undesirable interaction forces and torques (Jarrasse & Morel, 2012; Sarkisian et al., 2021), yet these factors are inevitable (Sarkisian et al., 2020). Consequently, it is necessary to optimize the exoskeleton design to minimize these side effects. The results also suggest that the joint angle can affect the contact force at the corresponding region, and further study is needed to confirm the relationship. This relationship might depend not only the supporting extent but also the impediment of natural movement, reinforcing the necessity for design optimization to mitigate these effects.

Several limitations of this study should be acknowledged. The small sample size of three participants limits statistical power. Furthermore, the investigation focused on only two body regions (shoulder/chest and thigh), leaving out other important areas like the waist and hip that could influence overall comfort. The absence of empirical data on supporting torques also restricted the analysis of their relationship with contact forces. Additionally, as relating user comfort with human-exoskeleton interaction force is a further topic, current study did not well investigate user comfort for now. Future studies should aim for a larger, statistically robust sample size and include additional body regions. Investigating the relationship between supportive torque curves and contact forces would enhance understanding of how support is distributed across the body and inform better load distribution designs. Moreover, exploring the link between user comfort and contact forces will shed light on pressure tolerance and improve the design of pressure distribution features in back exoskeletons. Lastly, further examination of how joint angles affect contact forces is crucial for optimizing exoskeleton functionality and user comfort.

5. CONCLUSIONS

The investigation into the interaction forces between users and back exoskeletons revealed significant insights into their design and functionality for reducing low back load in construction workers. SV Exosuit, characterized by its active support through pneumatic artificial muscles, demonstrated notably higher interaction forces at both the thigh and shoulder/chest regions compared to Laevo V2 and MATE-XB. This increase can be largely attributed to the smaller contact area and lack of sufficient cushioning in the SV Exosuit, suggesting that while its supportive capabilities are advanced, user comfort remains a critical challenge. Laevo V2's design, which employed a rigid passive mechanism, exhibited distinct contact force behaviour, including two peaks across the motion cycle, influenced by its torque generation strategy that reaches its maximum supportive torque at the bending angle between 40° – 50° . This indicates a need for careful consideration of how torque is distributed throughout the motion cycle to minimize discomfort. Furthermore, the MATE-XB's observed displacement of the thigh pad during squatting tasks points to potential misalignment issues that could affect user experience and exoskeleton effectiveness. These findings highlight the necessity for further research focused on optimizing exoskeleton designs, particularly in terms of cushioning materials and load distribution strategies, to enhance user comfort and promote wider adoption in the construction industry. Future studies should also aim to include a diverse participant pool and investigate additional anatomical regions, ultimately informing better design practices and improving the efficacy of back exoskeletons in preventing LBP among construction workers.

ACKNOWLEDGMENTS

The authors acknowledged the first authorship of Mr. Ting Lei and the co-correspondences of Dr. Joon Oh Seo and Dr. Kelvin Heung of this manuscript.

REFERENCES

- Adhikari, B., Ghimire, A., Jha, N., Karkee, R., Shrestha, A., Dhakal, R., Niraula, A., Majhi, S., Pandit, A. K., & Bhandari, N. (2021). Factors associated with low back pain among construction workers in Nepal: A cross-sectional study. *PLoS One*, 16(6), e0252564. <https://doi.org/10.1371/journal.pone.0252564>
- Alemi, M. M., Geissinger, J., Simon, A. A., Chang, S. E., & Asbeck, A. T. (2019). A passive exoskeleton reduces peak and mean EMG during symmetric and asymmetric lifting. *Journal of Electromyography and Kinesiology*, 47, 25-34. <https://doi.org/https://doi.org/10.1016/j.jelekin.2019.05.003>
- Alemi, M. M., Madinei, S., Kim, S., Srinivasan, D., & Nussbaum, M. A. (2020). Effects of Two Passive Back-Support Exoskeletons on Muscle Activity, Energy Expenditure, and Subjective Assessments During Repetitive Lifting. *Human Factors*, 62(3), 458-474. <https://doi.org/10.1177/0018720819897669>
- Alonso-García, M., & Sarriá-Santamera, A. (2020). The Economic and Social Burden of Low Back Pain in Spain: A National Assessment of the Economic and Social Impact of Low Back Pain in Spain. *Spine*, 45(16). https://journals.lww.com/spinejournal/fulltext/2020/08150/the_economic_and_social_burden_of_low_back_pain_in.15.aspx
- Bennett, S., Han, W., Mahmud, U. S. M. D., Adamczyk, P., Dai, F., Wehner, M., Veeramani, D., & Zhu, Z. (2023). Usability and Biomechanical Testing of Passive Exoskeletons for Construction Workers: A Field-Based Pilot Study. *Buildings*, 13, 822. <https://doi.org/10.3390/buildings13030822>
- Burdorf, A., & Jansen, J. P. (2006). Predicting the long term course of low back pain and its consequences for sickness absence and associated work disability. *Occupational and Environmental Medicine*, 63(8), 522. <https://doi.org/10.1136/oem.2005.019745>
- Carregaro, R. L., Tottoli, C. R., Rodrigues, D. d. S., Bosmans, J. E., da Silva, E. N., & van Tulder, M. (2020). Low back pain should be considered a health and research priority in Brazil: Lost productivity and healthcare costs between 2012 to 2016. *PLoS One*, 15(4), e0230902. <https://doi.org/10.1371/journal.pone.0230902>
- Golabchi, A., Jasimi Zindashti, N., Miller, L., Rouhani, H., & Tavakoli, M. (2023). Performance and effectiveness of a passive back-support exoskeleton in manual material handling tasks in the construction industry. *Construction Robotics*, 7(1), 77-88. <https://doi.org/10.1007/s41693-023-00097-4>
- Hensel, R., & Keil, M. (2019). Subjective Evaluation of a Passive Industrial Exoskeleton for Lower-back Support: A Field Study in the Automotive Sector. *IIEE Transactions on Occupational Ergonomics and Human Factors*, 7(3-4), 213-221. <https://doi.org/10.1080/24725838.2019.1573770>
- Hwang, S., Kim, Y., & Kim, Y. (2009). Lower extremity joint kinetics and lumbar curvature during squat and stoop lifting. *BMC Musculoskeletal Disorders*, 10(1), 15. <https://doi.org/10.1186/1471-2474-10-15>
- Jarrasse, N., & Morel, G. (2012). Connecting a Human Limb to an Exoskeleton. *IEEE Transactions on Robotics*, 28(3), 697-709. <https://doi.org/10.1109/TRO.2011.2178151>
- Kashif, M., Albalwi, A., Raqib, A., Farooq, M., Ullah, R., Sakoor, M., & Kamran, Z. (2022). Work-related musculoskeletal disorders among Pakistani construction workers: Prevalence, characteristics, and associated risk factors. *Work*, 72, 119-126. <https://doi.org/10.3233/WOR-205009>
- Kim, S., Madinei, S., Alemi, M. M., Srinivasan, D., & Nussbaum, M. A. (2020). Assessing the potential for “undesired” effects of passive back-support exoskeleton use during a simulated manual assembly task: Muscle activity, posture, balance, discomfort, and usability. *Applied Ergonomics*, 89, 103194. <https://doi.org/https://doi.org/10.1016/j.apergo.2020.103194>
- Koopman, A. S., Kingma, I., de Looze, M. P., & van Dieën, J. H. (2020). Effects of a passive back exoskeleton on the mechanical loading of the low-back during symmetric lifting. *Journal of Biomechanics*, 102, 109486. <https://doi.org/https://doi.org/10.1016/j.jbiomech.2019.109486>
- Kozinc, Ž., Babič, J., & Šarabon, N. (2021). Human pressure tolerance and effects of different padding materials with implications for development of exoskeletons and similar devices. *Applied Ergonomics*, 93, 103379. <https://doi.org/https://doi.org/10.1016/j.apergo.2021.103379>
- Kranenborg, S. E., Greve, C., Reneman, M. F., & Roossien, C. C. (2023). Side-effects and adverse events of a shoulder- and back-support exoskeleton in workers: A systematic review. *Appl Ergon*, 111, 104042. <https://doi.org/10.1016/j.apergo.2023.104042>
- Lei, T., Seo, J., Liang, K., Xu, J., Li, H., Zhou, Y., Khan, M., & Heung Kelvin, H. (2024). Lightweight Active

- Soft Back Exosuit for Construction Workers in Lifting Tasks. *Journal of Construction Engineering and Management*, 150(7), 04024073. <https://doi.org/10.1061/JCEMD4.COENG-14490>
- Levesque, L., Pardoel, S., Lovrenovic, Z., & Doumit, M. (2017, 5-7 Oct. 2017). Experimental comfort assessment of an active exoskeleton interface. 2017 IEEE International Symposium on Robotics and Intelligent Sensors (IRIS),
- Luger, T., Bär, M., Seibt, R., Rieger, M. A., & Steinhilber, B. (2021). Using a Back Exoskeleton During Industrial and Functional Tasks-Effects on Muscle Activity, Posture, Performance, Usability, and Wearer Discomfort in a Laboratory Trial. *Hum Factors*, 65(1), 5-21. <https://doi.org/10.1177/00187208211007267>
- Matthew, R. P., Mica, E. J., Meinhold, W., Loeza, J. A., Tomizuka, M., & Bajcsy, R. (2015, 28 Sept.-2 Oct. 2015). Introduction and initial exploration of an Active/Passive Exoskeleton framework for portable assistance. 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS),
- Miranda, H., Viikari-Juntura, E., Punnett, L., & Riihimäki, H. (2008). Occupational loading, health behavior and sleep disturbance as predictors of low-back pain. *Scandinavian Journal of Work, Environment & Health*, 34(6), 411-419. <http://www.jstor.org/stable/40967738>
- Näf, M. B., Koopman, A. S., Baltrusch, S., Rodriguez-Guerrero, C., Vanderborght, B., & Lefeber, D. (2018). Passive Back Support Exoskeleton Improves Range of Motion Using Flexible Beams [Original Research]. *Frontiers in Robotics and AI*, 5. <https://doi.org/10.3389/frobt.2018.00072>
- Park, J.-H., Kim, S., Nussbaum, M. A., & Srinivasan, D. (2022). Effects of back-support exoskeleton use on gait performance and stability during level walking. *Gait & Posture*, 92, 181-190. <https://doi.org/https://doi.org/10.1016/j.gaitpost.2021.11.028>
- Poliero, T., Fanti, V., Sposito, M., Caldwell, D. G., & Natali, C. D. (2022). Active and Passive Back-Support Exoskeletons: A Comparison in Static and Dynamic Tasks. *IEEE Robotics and Automation Letters*, 7(3), 8463-8470. <https://doi.org/10.1109/LRA.2022.3188439>
- Poliero, T., Lazzaroni, M., Toxiri, S., Di Natali, C., Caldwell, D. G., & Ortiz, J. (2020). Applicability of an Active Back-Support Exoskeleton to Carrying Activities. *Front Robot AI*, 7, 579963. <https://doi.org/10.3389/frobt.2020.579963>
- Sarkisian, S. V., Ishmael, M. K., Hunt, G. R., & Lenzi, T. (2020). Design, Development, and Validation of a Self-Aligning Mechanism for High-Torque Powered Knee Exoskeletons. *IEEE Transactions on Medical Robotics and Bionics*, 2(2), 248-259. <https://doi.org/10.1109/TMRB.2020.2981951>
- Sarkisian, S. V., Ishmael, M. K., & Lenzi, T. (2021). Self-Aligning Mechanism Improves Comfort and Performance With a Powered Knee Exoskeleton. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 29, 629-640. <https://doi.org/10.1109/TNSRE.2021.3064463>
- Schwartz, M., Theurel, J., & Desbrosses, K. (2021). Effectiveness of Soft versus Rigid Back-Support Exoskeletons during a Lifting Task. *International Journal of Environmental Research and Public Health*, 18(15).
- Serrancolí, G., Falisse, A., Dembia, C., Vantilt, J., Tanghe, K., Lefeber, D., Jonkers, I., Schutter, J. D., & Groote, F. D. (2019). Subject-Exoskeleton Contact Model Calibration Leads to Accurate Interaction Force Predictions. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(8), 1597-1605. <https://doi.org/10.1109/TNSRE.2019.2924536>
- Serranheira, F., Sousa-Uva, M., Heranz, F., Kovacs, F., & Sousa-Uva, A. (2020). Low Back Pain (LBP), work and absenteeism. *Work*, 65, 463-469. <https://doi.org/10.3233/WOR-203073>
- Shushtari, M., & Arami, A. (2023). Human-Exoskeleton Interaction Force Estimation in Indego Exoskeleton. *Robotics*, 12(3).
- Van Harmelen, V., Schnieders, J., & Wagemaker, S. (2022). Measuring the amount of support of lower back exoskeletons. *Laevo White Paper*.
- Vasiwala, R. A., Singh, S., Wong, H. Y., Su, J. Y., Wong, Z. J., & Tong, J. Y. (2021). Prevalence and Severity of Low Back Pain in Construction Workers. *Asia Pacific Environmental and Occupational Health Journal*, Vol 7 (1): 1-6. <https://doi.org/10.24049/1743/10289>
- Walter, T., Stutzig, N., & Siebert, T. (2023). Active exoskeleton reduces erector spinae muscle activity during lifting. *Front Bioeng Biotechnol*, 11, 1143926. <https://doi.org/10.3389/fbioe.2023.1143926>
- Wang, X., Dong, X. S., Choi, S. D., & Dement, J. (2017). Work-related musculoskeletal disorders among construction workers in the United States from 1992 to 2014. *Occupational and Environmental Medicine*, 74(5), 374. <https://doi.org/10.1136/oemed-2016-103943>
- Wang, Y., Qiu, J., Cheng, H., & Zheng, X. (2020). Analysis of Human-Exoskeleton System Interaction for Ergonomic Design. *Human Factors*, 65(5), 909-922. <https://doi.org/10.1177/0018720820913789>
- Wu, H., Li, H., Fang, X., & Luo, X. (2022). A survey on teaching workplace skills to construction robots. *Expert Systems with Applications*, 205, 117658. <https://doi.org/https://doi.org/10.1016/j.eswa.2022.117658>

Zelik, K. E., Nurse, C. A., Schall, M. C., Sesek, R. F., Marino, M. C., & Gallagher, S. (2022). An ergonomic assessment tool for evaluating the effect of back exoskeletons on injury risk. *Applied Ergonomics*, 99, 103619. <https://doi.org/https://doi.org/10.1016/j.apergo.2021.103619>

EEG TECHNOLOGY IN CONSTRUCTION SAFETY RESEARCH: CURRENT TRENDS, CHALLENGES, AND FUTURE DIRECTIONS

Tanghan Jiang¹, Jiayu Chen², and Yihai Fang³

1) Ph.D. Candidate, Department of Civil Engineering, Monash University, Melbourne, Australia. Email: Tanghan.Jiang1@monash.edu

2) Ph.D., Assoc. Prof., Department of Civil Engineering, Tsinghua University, Beijing, China. Email: jiayuchen@tsinghua.edu.cn

3) Ph.D., Senior Lecturer., Department of Civil Engineering, Monash University, Melbourne, Australia. Email: Yihai.Fang@monash.edu

Abstract: The construction industry faces persistent challenges in reducing accidents caused by human factors. Traditional safety management measures, while valuable, often fail to capture the intricate cognitive and behavioral dynamics of workers in hazardous environments. Electroencephalogram (EEG) technology has emerged as a promising tool for understanding these complexities by analyzing brain activity in real-time. This paper synthesizes the latest studies on the application of EEG in construction safety, focusing on its role in identifying cognitive states, enhancing safety protocols, and integrating with advanced technologies such as Virtual Reality (VR) and machine learning. By systematically reviewing key findings, this study identifies critical gaps in the current research and explores opportunities for developing data-driven, adaptive safety management strategies. The insights presented aim to inspire future innovations and refine approaches to hazard perception and prevention in the construction sector.

Keywords: EEG, Construction safety, Electroencephalogram, Safety management

1. INTRODUCTION

Safety enhancements are critically sought after by the construction industry worldwide, as they directly impact workers' lives and overall construction productivity. In fact, over 70% of construction accidents are attributed to human factors (Winge et al., 2019), and this trend is on the rise. To analyse and prevent construction accidents, traditional safety management strategies revolved around conducting on-site safety inspections and delivering training programs regularly. While proved effective, these methods are insufficient for fundamentally preventing accidents. This shortfall is mainly due to their limited ability to comprehend the intricate interplay between workers' cognitive processes and behavioural responses when faced with potential hazards (Gao et al., 2019). Thus, additional efforts, including applying self-administered questionnaires and conducting interviews after the construction sessions, serve as supplementary methods to gain deeper insights into the workers' mental states. However, these methods are often criticised for being too subjective to reveal psychological factors (Lohani et al., 2019). There is an urgent need to conduct an in-depth analysis of construction workers' recognition mechanisms to prevent accidents caused by human factors effectively. In essence, an appropriate research tool plays a crucial role.

Electroencephalogram (EEG) could meet such demand, as it could record the brain's electrical activity through non-invasive electrodes placed on the scalp (Teplan, 2002). EEG captures the electrical fluctuations from ionic current flows within the neurons in the brain. These signals are typically categorised into distinct frequency bands: delta (0.5–4 Hz), theta (4–7 Hz), alpha (8–13 Hz), beta (14–30 Hz), and gamma waves (>30 Hz) (Buzsaki, 2006), and each frequency band correlates with different mental states and activities. This trending technique also marks one step further to the neural underpinnings of how workers perceive, identify, and react to various risks and could offer exciting insights into enhancing traditional safety management approaches (Holm et al., 2009).

Current attempts to use EEG in the construction safety field mainly focus on examining workers' physical responses to various hazards and their cognitive states, including attention, vigilance, and mental workload (Wang et al., 2017a). This focus is driven by an emerging understanding that workers' interactions with hazardous scenarios are profoundly influenced by their experiences, cognitive states, and personal characteristics (Gao et al., 2020). Preliminary studies in this direction have begun to unravel these influences, revealing how individual differences affect safety-related behaviours and decisions. However, research in this area remains fragmented, with studies exploring various aspects in isolation. This fragmentation hinders the development of a systematic understanding of how these cognitive and behavioural factors interplay in risky construction environments.

Recognising this gap, a targeted review of workers' behaviours and brain activities under risky scenarios becomes crucial. Such a review is crucial and time-intensive considering the rapidly growing EEG-involved construction safety studies and its potential to reshape training and risk management strategies. This review could build connections between existing theoretical understanding and practical safety protocols by systematically organising and analysing these relevant studies. It could also serve as a start that guides EEG data collection and analysis targeting construction tasks and offers insights into developing tailored training programs and adaptive safety measures.

2. METHOD

To thoroughly search for studies that identify the boundary of knowledge and prospective research gaps, this paper employed a three-step review approach: (1) searching and screening the literature; (2) summarising and analysing the existing attempts; (3) discussing research gaps and future directions. The main search platforms were the Web of Science and Scopus with the following keywords: “construction worker* OR construction industry” AND “safety* OR safety training* OR risk* OR hazard*OR accidents*” AND “ EEG* OR Electroencephalogram* OR brain activities* OR biometric*”.

The inclusion criteria for this review were: (1) studies published between 2017 and 2023 to capture recent advancements; (2) papers using EEG technology in construction-related safety applications; and (3) research with detailed methodology and quantitative results. Exclusion criteria included papers without peer review, studies outside the construction sector, and those lacking explicit experimental protocols. Each study was screened first by abstract, followed by a full-text review by two independent reviewers. Discrepancies were resolved by a third reviewer. In total, 22 papers were selected as the final literature database based on their close relevance to the topic (Figure 1).

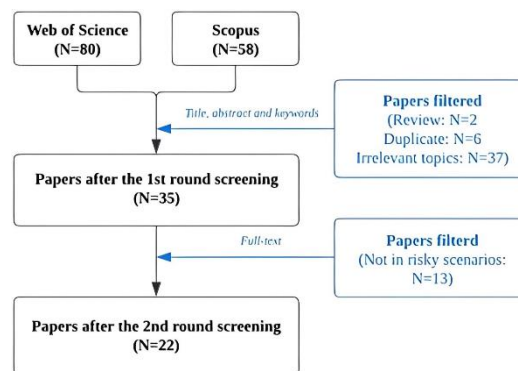


Figure 1: Review methodology

3. RESULTS

3.1 Descriptive and bibliographic analysis

The selected studies are from a diverse range of journals and time. Notably, the “Automation in Construction” journal leads with about 32% of the total publications. Other prominent journals include “Journal of Construction Engineering and Management” (14%) and “Safety Science” (9%). On the other hand, the publishing time of these studies ranged from 2017 to 2023. The most productive years are 2022 and 2023, with 38% and 33% of the studies, respectively. This indicates a growing interest and development in this research area in recent years.

The evolution of key research topics was visualised according to the keywords used in the reviewed studies (Figure 2). The size of each node is proportional to the keyword’s frequency of mention across the literature, indicating its significance. The legend of the timeline in this network diagram represents the range of publication years, and the colour transitions from deep blue to light yellow reflect the chronological order in which the keywords first appeared in the literature. The interconnecting arcs signify the links between pairs of keywords, with the thickness of each arc correlating to the strength of the connection between them.

This diagram illustrates the interconnected nature of construction workers’ behaviours and brain activities within hazardous contexts, as evidenced by the clustering and linkage of keywords. The “electroencephalogram (EEG)” is central to this network, which emerges as a pivotal node, indicating its significant role in measuring brain activities. Its strong connections to keywords like “safety,” “vigilance,” and “perception” underscore the critical integration of cognitive status monitoring and safety management attempts. Notably, the dense interconnectivity around “construction safety” highlights an overarching concern for construction workers’ safety, merging research on situational awareness with cognitive responses to hazards. The presence of newer, lighter-coloured nodes related to “performance” and “attention” points to a growing research focus on real-time assessment of workers’ physical and mental statuses.

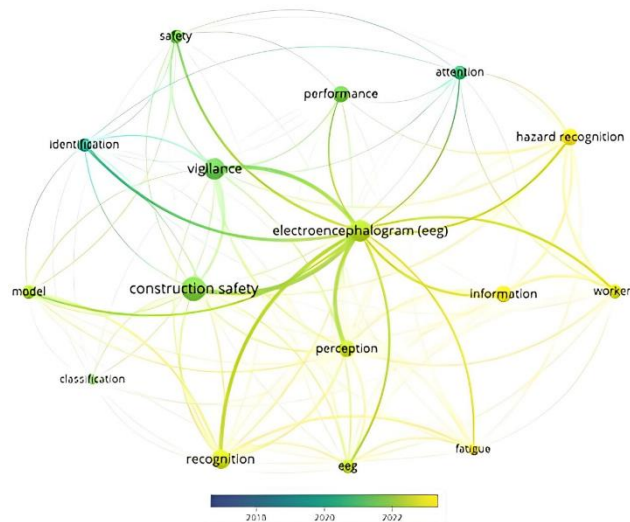


Figure 1. General structure of reports, papers, and essays

3.2 Main research directions

Theoretical frameworks

The theoretical frameworks in construction safety research are the foundation for understanding workers' cognitive and neurophysiological underpinnings in hazardous situations. This research direction is instrumental in transforming complex EEG data into meaningful insights about worker safety behaviours. The studies in this direction can be divided into two main categories: neural mechanisms during hazard perception and the psychological influences on hazard identification.

In the first category, studies analysed the activity and interconnectivity of different brain regions during hazard perception and identification. For instance, one representative study (Liao et al., 2023) investigated construction workers' brain connectivity during hazard identification tasks, employing EEG data to simulate directed and time-variant information flow across the brain. This research highlighted a top-down attention modulation from the dorsal attention network, illustrating a reorientation mechanism for processing threatening stimuli. Another typical study (Zhou et al., 2022) investigated the psychological mechanisms of the impact of construction experience on hazard identification performance. Compared to experienced workers, novices presented more evident gamma-band activity due to their heightened sensitivity, suggesting more efficient working memory and attentional control during hazard identification.

The second category explored how dispositional factors or individual differences, such as risk propensity and work experience, influence the neural processes in hazard identification. Liu et al. (2021) examined the mediating effect of brain activities between dispositional factors and hazard identification performance. Utilising a multi-level logistic regression approach, this research uncovered a positive correlation between risk propensity and stimulus-locked difference waveforms in brain areas associated with hazard perception. Concurrently, it noted an inverse relationship between injury exposure experience and response-locked waveforms, pointing to distinct neural processing patterns. Zhou and Liao (2023) extended the findings on dispositional factors by examining how specific personality characteristics (e.g., risk tolerance and stress resilience) influence the neural mechanisms underpinning hazard identification. It used EEG to track brain responses during simulated construction scenarios. This research revealed that workers with higher risk tolerance exhibit distinct neural patterns, particularly in risk assessment and decision-making regions. With all the studies in this research category, a more comprehensive understanding of workers' cognitive processes was achieved, particularly highlighting the interplay between individual psychological characteristics, prior experiences, and neural responses in construction safety.

Practical applications

This direction aims to translate theoretical insights into proactive real-world safety applications. Studies falling in this direction can be categorised into two types: the real-time monitoring and assessment of workers' mental states and the examination of individual and collective responses to hazards and safety training.

Significant advancements have been achieved in the domain of real-time monitoring and assessment. For example, a study guided the development of EEG-based systems for monitoring attention and vigilance (Wang et al., 2017b). This study is pivotal as it could detect changing psychological states and mark those that might lead to risks or accidents. Complementary to this, Shayesteh et al. (2023) delved into evaluating mental workload to prevent cognitive overload, a significant factor in workplace accidents. This research underscores balancing task demands with workers' cognitive capacities. Furthermore, integrating hybrid kinematic-EEG signals for detecting workers' vigilance was explored (Cheng et al., 2022), demonstrating the efficacy of multi-modal approaches in

safety management. These initiatives were further supplemented by findings from the study measuring workers' emotional valence in different working conditions, thereby offering insights into how emotional states can impact safety on construction sites (Wang et al., 2023).

The second type within this direction emphasises the importance of personalised responses to hazards and the optimisation of safety training. For example, a study investigated how different training methods impact workers' hazard identification skills and ability to respond to hazardous situations (Jeon and Cai, 2023). It suggested that interactive and immersive training methodologies significantly enhance training effectiveness and pointed out the need for optimising current training methodologies. Similarly, the relationship between physical and mental fatigue in construction workers was thoroughly examined (Xing et al., 2020). This study found a compounding effect where physical fatigue can exacerbate mental fatigue, leading to diminished cognitive function and increased risk of accidents. This insight is instrumental in developing comprehensive fatigue management strategies that address both physical and mental aspects, enhancing construction workplace safety. Findings from all these studies enriched the understanding of the psychological and cognitive dimensions of construction work in risky scenarios but also guided the development of more effective, adaptive, and personalised safety management practices.

Technical platforms and tools

Technical platforms and tools in this research area mark a significant advancement in synthesising theoretical knowledge and practical applications into safety training and hazard management assisted by emerging technologies.

The centre of this category is integrating EEG data with VR environments. This was evidenced by the research leveraging an EEG classifier within a VR setting to identify crucial EEG signal characteristics indicative of hazard perception (Jeon and Cai, 2021). Such integration increased the realism of hazard identification training and provided a new approach to understanding how cognitive states correlate with simulations mimicking real-world scenarios. Building upon this, Jeon and Cai (2022) refined cognitive state assessment by improving hazard classification accuracy. Utilising wearable EEG devices, this research employed multi-class classification techniques to distinguish between various construction hazards. Applying machine learning algorithms further demonstrated the potential of EEG data to enhance hazard identification in dynamic and complex construction environments.

Another innovative approach in this field is using machine vision technologies for hazard inspection. A study integrated human knowledge extracted from EEG data into machine vision systems, employing transfer learning techniques to enhance prediction accuracy for hazard classification (Chen et al., 2022). Such an approach is constructive when scenario data sets are limited, building a close connection between collected EEG data, cognitive models and machine learning algorithms.

Further advancing the field, a study integrating deep EEG-net and other physiological indicators into a VR environment (Huang et al., 2022) marked a significant advancement in developing training platforms. This new platform could offer comprehensive and real-time user performance monitoring and assessment. Compared to traditional training platforms, this system provided a more immersive and interactive experience, potentially increasing the effectiveness and acceptance of VR-based safety training. By leveraging theoretical knowledge and state-of-the-art technologies, all works in this direction significantly enhanced the capabilities of professional training and risk management, paving the way for safer and more efficient construction environments.

3.3 Participant and apparatus selection

Participant types

The exploration of construction workers' behaviour and brain activity under risky scenarios often necessitates a sample population that closely mirrors the real-world demographic of the construction workforce. 68% of studies recruited real workers as participants, while 14% recruited students and 18% remained unclear.

EEG devices and channel numbers

The EEG devices used in the studies varied in their models and channel numbers, reflecting different research needs. Devices with higher channel counts, such as the 32-channel systems from Neuroscan Systems and Brain Products (Chen et al., 2022; Zhou and Liao, 2023), offer a significant advantage in detailed brain activity mapping. The increased number of channels allows for more precise localisation of brain activity, which is particularly beneficial for in-depth cognitive studies and complex tasks like identifying and classifying multiple hazard types. This level of detail enables researchers to understand better the specific neural mechanisms involved in cognitive processes and hazard identification. However, these high-channel-count systems also have challenges, including higher costs, more complex setup and calibration requirements, and potentially longer preparation times. The complexity of data interpretation also increases with the number of channels, requiring more complex and robust analysis techniques and may even require expert guidance and cross-validations of other physiological indicators.

On the other hand, devices with fewer channels, such as the 14-channel Emotiv Epoc+ (Jeon and Cai, 2021; Wang et al., 2019), are generally utilised for general cognitive assessments. These devices are more user-

friendly and more accessible to deploy, making them suitable for studies that require quick setup or where portability is a priority. While limiting the resolution of brain activity mapping, the lower channel count still provides valuable insights into general cognitive states and processes. This can be particularly useful in applied settings where a detailed understanding of specific brain regions may be less critical. The major drawbacks of these lower-channel-count devices include reduced spatial resolution, which may inaccurately detect or overlook subtle changes in cognitive or psychological status.

Electrode types and deployment

The studies chose different electrode types based on the experimental environment and the required EEG signal accuracy. Wet electrodes require conductive gel and are usually selected for their high signal accuracy, especially in studies targeting detailed brain activity monitoring (Liao et al., 2023). Dry electrodes are preferred for their ease of use in dynamic settings but may offer slightly less accurate data (Choi et al., 2023). Though not selected in the reviewed studies, semi-dry electrodes balance the high signal quality of wet electrodes and the user-friendly nature of dry electrodes, making them suitable for studies requiring both ease of application and reliable data collection.

3.3 Data processing procedure

EEG data processing is a multi-step process crucial for extracting meaningful insights from raw neural signals (Figure 3). The first step is using scalp-based electrodes to collect brain function-related electrical activities. Since the collected data is often noisy and prone to artifacts, it then demands a systematic processing pipeline to ensure reliability and validity, which includes artifact removal, data labelling, and ultimate feature extraction and analysis.

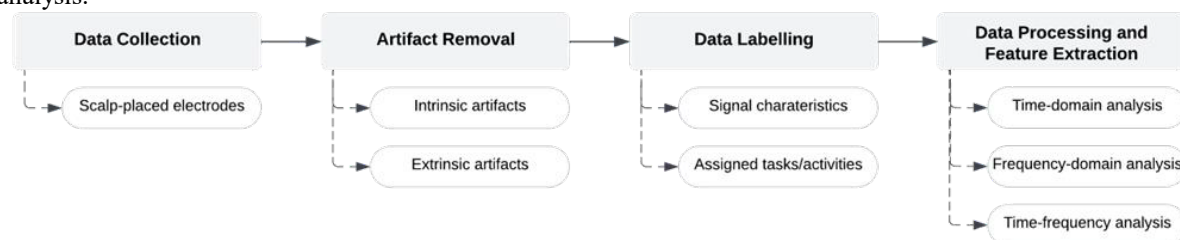


Figure 3: EEG signal processing flowchart

Artifact sources and removal methods

The primary sources of detected EEG signal artifacts can be divided into two main categories: extrinsic artifacts and intrinsic artifacts (Hwang et al., 2018). Extrinsic artifacts usually arise from external factors, such as environmental noise (e.g., ambient noise from equipment, power lines, and wireless devices) and technical issues (e.g., electrode popping, poor electrode contact, or errors with the EEG recording device). In contrast, intrinsic artifacts originate from within the individual, such as physiological activities (e.g., eye movements, blinks, muscle activities) and other brain activities that are unrelated to the specific cognitive or behavioural aspect under investigation (e.g., drowsiness, distractions, and research-unrelated thought processes) (Jiang et al., 2019). If not adequately addressed, these artifacts can lead to errors in the brain signal analysis and inaccurate interpretations of neural activities, particularly in high-cognitive-demanding scenarios, such as hazard identifications and reactions (Wang et al., 2022; Zhang et al., 2023).

To address extrinsic artifacts effectively, the studies under review have commonly employed frequency filters, such as Finite Impulse Response (FIR) bandpass filters (Jeon and Cai, 2021). These filters effectively reduce signal drifts and isolate the EEG signal's relevant frequency bands. Notch filters are also frequently used to eliminate electromagnetic interference, specifically targeting the 50/60 Hz frequency range caused by power line noise.

In the case of intrinsic artifacts, complex statistical methods such as Independent Component Analysis (ICA) using algorithms like Infomax have been extensively used. ICA plays a critical role in isolating and removing artifacts, such as muscle movements, by decomposing the EEG signal into independent components (Wang et al., 2019). Furthermore, spatial filtering techniques, including using plugins like ICLabel (Jeon and Cai, 2023), categorise these components according to their sources (e.g., eye movement or muscle activity). This categorisation process often involves manual inspection of time courses and scalp topography maps to ensure the validity of the artifact removal process, resulting in a cleaner EEG signal.

Data labelling and integration with experimental settings

Data labelling is pivotal in understanding participants' cognitive states and responses triggered by the specific event, such as attention, stress level, and hazard identification. Labelling usually combines EEG signal characteristics with the assigned task or activity. For example, particular EEG patterns might be labelled as 'high attention' levels during tasks requiring vigilance (Jeon and Cai, 2022).

Labelling is also crucial for integrating the EEG data with the experimental settings, including environmental and task-specific contexts. This integration is evident in studies where researchers correlate EEG

data with the specific tasks performed by construction workers, or the hazards encountered during data collection. For example, EEG data recorded during high-risk activities like working at heights are analysed differently than those recorded during more routine tasks, reflecting the varying cognitive demands of these activities (Liao et al., 2023).

There are five main approaches for general EEG data labelling. (1) Event-Related Labelling (ERL) involves assigning event markers as labels related to specific events or stimuli during the recording process. (2) Task-Related Labelling (TRL) depends on the participant's performance in various tasks. (3) Automated Labelling (AL) employs algorithms or machine learning models to label EEG data. (4) Manual Labelling (ML) requires experts to scrutinise EEG recordings for segmenting and labelling, either visually or with other methods. (5) Subjective Labelling (SL) assigns labels based on the participant's self-report on their perceptions, feelings, and experiences (Siuly et al., 2016).

Although various approaches have been developed, challenges remain at current labelling attempts. For example, labelling often involves a degree of subjectivity, especially in interpreting complex brainwave patterns. Therefore, ensuring consistency and accuracy in labelling across different datasets and researchers is a significant challenge. The resolution and sensitivity of current available EEG monitoring devices can also limit the accuracy of data labelling. Additionally, aligning and integrating these data sources for accurate labelling can be more complex in studies combining EEG data with other data types, such as behavioural metrics or environmental measurements.

Moreover, it is also important to note that not all EEG studies employ data labelling. Certain analyses may focus on broader signal patterns or aggregate data trends rather than labelling specific cognitive states. This suggests that although labelling is a powerful tool for specific types of analyses, particularly those that require a precise understanding of context-driven cognitive states, it is not the only pathway for EEG data interpretation.

Data processing and feature extraction

The processing of cleaned EEG data usually includes three main types of analysis: time-domain analysis, frequency-domain analysis, and time-frequency analysis. Time-domain analysis focuses on the amplitude and variability of the EEG signal over time. This approach is useful for understanding the temporal dynamics of cognitive processes, such as the amplitude and duration of signal components. For example, a study explored the impact of construction experience on hazard identification performance, utilising time-domain analysis to investigate gamma-band activity and its relationship with cognitive functions like working memory and attention control (Zhou et al., 2022). Frequency-domain analysis examines power distribution across various frequency bands, which can indicate different cognitive states. For example, increased power in the theta band might be associated with high cognitive workload or attentional processes in hazard perception tasks (Shayesteh et al., 2023). Time-frequency analysis provides a more comprehensive view by considering both the temporal and spectral aspects of the EEG signal. It is not merely a combination of time-domain and frequency-domain analyses but a unique method that maps the signal onto a time-frequency space. This analysis could offer insights into how cognitive states change in response to stimuli or tasks, particularly in scenarios involving human-robot teaming in construction (Wang and Liao, 2023).

Feature extraction is another crucial step in this phase. This process involves distilling the vast amount of EEG data into meaningful features that can be used for further analysis or predictive modelling. These features might include specific patterns or markers identified in the EEG signal, such as peak amplitudes, power spectral densities, or connectivity measures between different brain regions. Moreover, specific research questions or objectives often guide the selection of extracted features. For instance, in studies focusing on hazard identification, the features related to attentional engagement or stress response are usually considered most relevant (Chen et al., 2022).

Integrating machine learning algorithms in this phase has also become increasingly common. These algorithms can help classify different cognitive states or predict behavioural outcomes based on the extracted EEG features. For example, machine learning models can be trained to differentiate between high and low cognitive workload states or to predict a worker's response to an imminent hazard based on their EEG patterns (Wang et al., 2022).

4. RECOMMENDATIONS FOR FUTURE WORK (DISCUSSION)

The evolving landscape of EEG applications in construction safety research presents numerous pathways for future exploration. The following directions, though not exhaustive, could be beneficial for advancements:

Enhanced artifact management in dynamic environments: Artifacts in EEG data, such as those caused by physical movements or environmental noise, can significantly affect data accuracy. Future research should explore advanced removal techniques or new sensor designs that filter out non-brainwave signals. For example, developing algorithms that differentiate moving participants' brain signals and muscle movements would greatly enhance EEG data quality in active construction settings.

Integration of EEG with other biometric data: Combining EEG data with other physiological indicators like heart rate, sweat response, pupil dilation or electrodermal activity (EDA) could provide a more comprehensive

view of a worker's physical and mental status. For instance, a study could correlate increased heart rate and specific EEG patterns to identify hazard perceptions or other decision-making moments. More than one physiological source could cross-validate and improve the data collection and analysis validity, thereby accelerating the understanding of workers' behaviours and cognitive mechanisms.

Real-time EEG monitoring systems: Developing systems that can analyse EEG data in real-time and provide immediate feedback to workers or supervisors on the construction site could revolutionise safety management protocols. For example, a system that alerts workers or supervisors when signs of extreme stress or cognitive overload are detected could help schedule tasks and set proper rest sessions to prevent accidents.

EEG data analysis with machine learning: Machine learning algorithms can handle complex EEG data more effectively than traditional analytical methods. Research could focus on training models based on EEG patterns to predict hazardous situations or unsafe behaviours. For instance, a machine learning model could predict when a worker tends to be exposed to risky conditions due to cognitive fatigue or distractions.

Improving EEG equipment: Current EEG equipment is mostly lab-based and does not ideally suit dynamic and harsh construction sites. Future research should aim to develop durable, wireless, and less intrusive EEG headsets that workers can easily and comfortably wear without affecting their scheduled work. Although some studies have already started integrating EEG sensors into safety helmets (Li et al., 2014), the accuracy of data collection is still worth improving and validating. Things are similar regarding combining EEG with VR environments. Some electrodes are under the pressure of VR headsets, which could decrease the data collection quality. Wearing two devices on the head could also make participants uncomfortable, resulting in the forced shortening of experiment durations. Therefore, there is an intense demand for optimising existing EEG equipment to be more accurate in data collection and have a more user-friendly wearing experience.

It is also important to note that this review is not without limitations. First, the analysis was limited to papers indexed in Web of Science and Scopus, potentially excluding relevant studies from other databases. Second, the selection criteria may introduce biases by focusing only on English-language publications. Future research should consider expanding the scope to include gray literature and studies from diverse geographic regions to provide a more comprehensive understanding. Additionally, real-time integration of EEG with other biometric systems and advancements in artifact management remain key areas for future exploration. Practical applications, such as on-site real-time EEG monitoring, hold the potential to revolutionize safety practices in the construction industry.

5. CONCLUSIONS

This review comprehensively examines the use of EEG in understanding the cognitive processes of construction workers in high-risk environments. Notably, the integration of EEG with emerging technologies, such as VR and machine learning shows great promise in enhancing safety training. The study also covers various experimental settings, from screen-based simulations to VR, to understand workers' hazard perception and response, highlighting the influence of experience on cognitive responses. Additionally, it addresses the complexities in selecting and deploying EEG apparatus, the challenges of different EEG device types, and the importance of the EEG data processing steps, including artifact removal, data labelling, and feature extraction. Overall, this review underscores the advancements in using EEG to understand worker safety and behaviour in construction, pointing to the need for further research in experiment optimisation and technology integration for real-time monitoring, thereby laying the groundwork for future developments in construction site safety.

ACKNOWLEDGMENTS

This research is supported by Building 4.0 CRC. The support of the Commonwealth of Australia through the Cooperative Research Centre Programme is acknowledged.

REFERENCES

- Chen, J., Xu, Q., Fang, D., Zhang, D., Liao, P.C., 2022. Perceptual decision-making 'in the wild': How risk propensity and injury exposure experience influence the neural signatures of occupational hazard recognition. *International Journal of Psychophysiology*, 177, 92–102.
- Cheng, B., Fan, C., Fu, H., Huang, J., Chen, H., Luo, X., 2022. Measuring and Computing Cognitive Statuses of Construction Workers Based on Electroencephalogram: A Critical Review. *IEEE Transaction of Computer Social System*, 9, 1644–1659.
- Choi, D., Seo, S., Park, H., Hong, T., Koo, C., 2023. Forecasting personal learning performance in virtual reality-based construction safety training using biometric responses. *Automation Construction* 156.
- Gao, Y., Gonzalez, V.A., Yiu, T.W., 2019. The effectiveness of traditional tools and computer-aided technologies for health and safety training in the construction sector: A systematic review. *Computers & Education* 138, 101–115.
- Gao, Y., González, V.A., Yiu, T.W., 2020. Exploring the relationship between construction workers' personality

- traits and safety behavior. *Journal of Construction Engineering Management* 146, 04019111.
- Holm, A., Lukander, K., Korpela, J., Sallinen, M., Müller, K.M.I., 2009. Estimating brain load from the EEG. *The Scientific World Journal* 9, 639–651.
- Huang, D., Wang, X., Liu, J., Li, J., Tang, W., 2022. Virtual reality safety training using deep EEG-net and physiology data. *Visual Computer* 38, 1195–1207.
- Hwang, S., Jebelli, H., Choi, B., Choi, M., Lee, S., 2018. Measuring workers' emotional state during construction tasks using wearable EEG. *Journal of Construction Engineering Management* 144, 04018050.
- Jeon, J.H., Cai, H., 2021. Classification of construction hazard-related perceptions using: Wearable electroencephalogram and virtual reality. *Automation Construction* 132.
- Jeon, J.H., Cai, H., 2022. Multi-class classification of construction hazards via cognitive states assessment using wearable EEG. *Advanced Engineering Informatics* 53.
- Jeon, J.H., Cai, H., 2023. Wearable EEG-based construction hazard identification in virtual and real environments: A comparative study. *Safety Science* 165.
- Jiang, X., Bian, G.-B., Tian, Z., 2019. Removal of artifacts from EEG signals: a review. *Sensors* 19, 987.
- Li, P., Meziane, R., Otis, M.J.-D., Ezzaidi, H., Cardou, P., 2014. A Smart Safety Helmet using IMU and EEG sensors for worker fatigue detection. In: 2014 IEEE International Symposium on Robotic and Sensors Environments (ROSE) Proceedings. IEEE, pp. 55–60.
- Liao, P.C., Zhou, X., Chong, H.Y., Hu, Y., Zhang, D., 2023. Exploring construction workers' brain connectivity during hazard recognition: a cognitive psychology perspective. *International Journal of Occupational Safety and Ergonomics* 29, 207–215.
- Liu, Y., Asce, S.M., Habibnezhad, M., Asce, M., Shayesteh, S., Jebelli, H., Asce, A.M., Lee, S., 2021. Paving the Way for Future EEG Studies in Construction: Dependent Component Analysis for Automatic Ocular Artifact Removal from Brainwave Signals.
- Lohani, M., Payne, B.R., Strayer, D.L., 2019. A review of psychophysiological measures to assess cognitive states in real-world driving. *Frontier in Human Neuroscience* 13.
- Shayesteh, S., Ojha, A., Liu, Y., Jebelli, H., 2023. Human-robot teaming in construction: Evaluative safety training through the integration of immersive technologies and wearable physiological sensing. *Safety Science* 159.
- Siuly, S., Li, Y., Zhang, Y., 2016. EEG signal analysis and classification. *IEEE Transaction on Neural System and Rehabilitation Engineering* 11, 141–144.
- Teplan, M., 2002. Fundamentals of EEG measurement. *Measurement science review* 2, 1–11.
- Wang, D., Chen, J., Zhao, D., Dai, F., Zheng, C., Wu, X., 2017a. Monitoring workers' attention and vigilance in construction activities through a wireless and wearable electroencephalography system. *Automation Construction* 82, 122–137.
- Wang, D., Chen, J., Zhao, D., Dai, F., Zheng, C., Wu, X., 2017b. Monitoring workers' attention and vigilance in construction activities through a wireless and wearable electroencephalography system. *Automation Construction* 82, 122–137.
- Wang, D., Li, H., Chen, J., 2019. Detecting and measuring construction workers' vigilance through hybrid kinematic-EEG signals. *Automation Construction* 100, 11–23.
- Wang, J., Su, Y.S., Liao, P.C., 2023. Re-investigation of the Mediating Effect of Brain Activities between Dispositional Factors and Hazard Recognition: A Multilevel Logistic Regression Approach. *KSCE Journal of Civil Engineering* 27, 3646–3658.
- Wang, M., Liao, P.C., 2023. Personality Assessment Based on Electroencephalography Signals during Hazard Recognition. *Sustainability* 15.
- Wang, M., Zhao, Y., Liao, P.C., 2022. EEG-based work experience prediction using hazard recognition. *Automation Construction* 136.
- Winge, S., Albrechtsen, E., Mostue, B.A., 2019. Causal factors and connections in construction accidents. *Safety Science* 112, 130–141.
- Xing, X., Zhong, B., Luo, H., Rose, T., Li, J., Antwi-Afari, M.F., 2020. Effects of physical fatigue on the induction of mental fatigue of construction workers: A pilot study based on a neurophysiological approach. *Automation Construction* 120.
- Zhang, Q., Liang, M., Chan, A.P.C., Liao, P.C., 2023. Visual attention and cognitive process in construction hazard recognition: Study of fixation-related potential. *Automation Construction* 148.
- Zhou, X., Liao, P.C., 2023a. EEG-Based Performance-Driven Adaptive Automated Hazard Alerting System in Security Surveillance Support. *Sustainability* 15.
- Zhou, X., Liao, P.C., Xu, Q., 2022. Reinvestigation of the Psychological Mechanisms of Construction Experience on Hazard Recognition Performance. *Human Factors* , 66(1), pp.221-233.

ENHANCING TRAFFIC SAFETY IN SMART CITIES THROUGH PERCEIVED RISK ANALYSIS FOR VULNERABLE ROAD USERS

Yuting Zhang¹, Xinyu Pan², Jiayu Chen³

1) Ph.D. Candidate, Department of Civil Engineering, Tsinghua University, Beijing, China. Email: yuting-z22@mails.tsinghua.edu.cn

2) Ph.D. Candidate, Department of Civil Engineering, Tsinghua University, Beijing, China. Email: panxinyu24@mails.tsinghua.edu.cn

3) D. Eng., Assoc. Prof., Department of Civil Engineering, Tsinghua University, Beijing, China. Email: jiaiyuchen@tsinghua.edu.cn

Abstract: The rapid growth of urbanization and traffic complexity in smart cities has led to heightened safety risks, particularly for vulnerable road users (VRUs) like pedestrians and cyclists. This study aims to enhance VRU safety by investigating their perceived risk and decision-making processes in various traffic scenarios. Traditional methods, such as surveys and observational studies, have limitations in capturing VRUs' cognitive processes and real-time behavior. This research proposes a multimodal approach combining electroencephalography (EEG) and eye-tracking technology to analyze VRUs' cognitive and physiological responses in simulated urban environments. By integrating data on attention, vigilance, and emotional states, this approach provides a comprehensive assessment of VRUs' risk perception mechanisms. Utilizing the CARLA simulator, a platform for pedestrian-vehicle interaction is developed to simulate realistic urban scenarios, while physiological and behavioral data are collected to establish a model for predicting VRU behavior. The study contributes to theoretical insights on risk perception and practical implications for designing safer urban infrastructure in smart cities, ultimately aiming to reduce VRU accidents and enhance traffic safety.

Keywords: Traffic safety, Risk perception, Electroencephalograph (EEG), Eye tracking

1. INTRODUCTION

In recent years, the rapid urbanization and motorization have led to a surge in traffic complexity, escalating the risk of accidents, especially for vulnerable road users (VRUs) [1]. Compared with vehicles, VRUs can change the directions and speeds at any time with higher randomness and uncertainty [2]. Despite advancements in road infrastructure, the safety of VRUs remains a critical challenge. In pedestrian detection system, one of the most important concepts is risk perception, which refers to an individual's assessment of the potential dangers they may encounter [3]. For pedestrians, this perception significantly influences their behavior, including decisions to cross streets, wait for traffic signals, or navigate crowded areas. Many studies have shown that the VRU's risk perception provides important information for identifying potential collision risks [4]. For example, pedestrians with a heightened sense of risk are more likely to engage in cautious behaviors, whereas those with a lower risk perception may take unnecessary risks [5]. However, the theoretical mechanisms of risk identification and decision-making for pedestrians are not yet clear, which is not conducive to improving the accuracy of predicting the actions that pedestrians are about to take [6]. Traditional methods for assessing risk perception typically include surveys, questionnaires, and observational studies. While these approaches can provide valuable insights, they often have limitations. Surveys may be subject to biases, such as social desirability,

where respondents provide reports which they believe are more acceptable rather than their true feelings [7]. Observational studies, while useful, can fail to capture the internal cognitive processes influencing decision-making. Consequently, these methods may not fully account for the complexities of pedestrian behavior in real-world scenarios. In the past, predictive models of pedestrian behavior often relied on statistical analyses and historical data. While these models can offer insights into general trends, they often overlook the individual cognitive and emotional factors that shape pedestrian decisions [8]. Moreover, many predictive frameworks are based on assumptions that may not hold true in diverse urban contexts. This gap highlights the need for more nuanced methodologies that can capture the dynamic nature of pedestrian behavior. To bridge this gap, this research aims to develop a physiological measurement method for perceptual risk and further construct the risk identification and behavioral decision model. This can integrate human factors into road safety management of the smart city framework.

Electroencephalogram (EEG) is a tool that reflects brain neural activity by measuring physiological electrical signals on the scalp surface [9]. EEG can be used in clinical diagnostics to determine the type and location of brain disorders, such as epilepsy and brain injuries. In addition, scholars in the field of human factors have utilized EEG extensively for research related to emotions, stress, cognitive load, and other psychological and behavioral sciences [10–12]. Today, portable and non-invasive EEG devices have become important research tools for studying cognitive processes and states in non-medical settings. There are already several examples of EEG being applied in safety behavior research for drivers, pilots, and others [13]. Eye-tracking technology, by measuring eye position and movement information, can be used to analyze visual features such as fixations and saccades [14]. These features have been shown to be related to human information processing and decision-making processes. Non-invasive eye-tracking technology has been widely applied in visual-related psychological and behavioral research. Portable screen-based eye trackers can capture and record various data points, such as the subject's eye position, pupil size, eyelid closure, blinking behavior, and fixation patterns, to help understand how individuals visually perceive stimuli and make decisions [15]. The multimodal integration of EEG and eye-tracking combines the advantages of both technologies, providing a more comprehensive analysis of cognitive processes. Applying a brain-eye fusion measurement scheme to experiments in virtual simulation platforms for road traffic scenarios can safely induce various behavioral decisions from subjects in different traffic scenarios. By leveraging multimodal data from brain-eye fusion, researchers can calculate physiological indicators such as attention and vigilance, which aids in understanding risk perception and behavioral mechanisms.

2. Methodology

In real-world contexts, experiments investigating the interaction between pedestrians and vehicles are constrained by multiple factors, including regulations, ethical considerations, and safety concerns, which complicate their implementation. Consequently, conducting experiments within simulated environments provides a viable alternative. Therefore, this study provides a brain-eye fusion measurement system for traffic participants' risk perception based on the autonomous driving simulation platform CARLA. The system ultimately results in a risk perception measurement model for road users in simulated scenarios based on a physiological sensing system. The research is divided into three phases and shown in Fig.1:

- (1) Risk perception analysis: This process focuses on a comprehensive examination of the risk identification and perception mechanisms of vulnerable road users (VRUs), aiming to identify

various factors that influence unsafe behaviors. By investigating how VRUs perceive potential risks in their surrounding environment, we can reveal the psychological dynamics and decision-making processes individuals experience when faced with danger. This understanding will ultimately shed light on why certain behaviors lead to unsafe outcomes.

- (2) Physiological sensing system establishment: A comprehensive physiological sensing system will be developed to measure various indicators of risk perception levels among VRUs. This system will integrate multiple sensor technologies to continuously monitor physiological responses, such as heart rate and skin conductance, providing objective data that will aid in assessing the user's ability to perceive risk and react in specific situations.
- (3) Risk identification and behavioral decision models: Utilizing the data collected from the physiological sensing system, the perceived risk among VRUs will be quantified. Based on this quantification, a behavioral decision model will be proposed that takes into account the perceived risks. This model aims to combine physiological data with behavioral characteristics to predict how VRUs are likely to behave when confronted with different levels of risk.

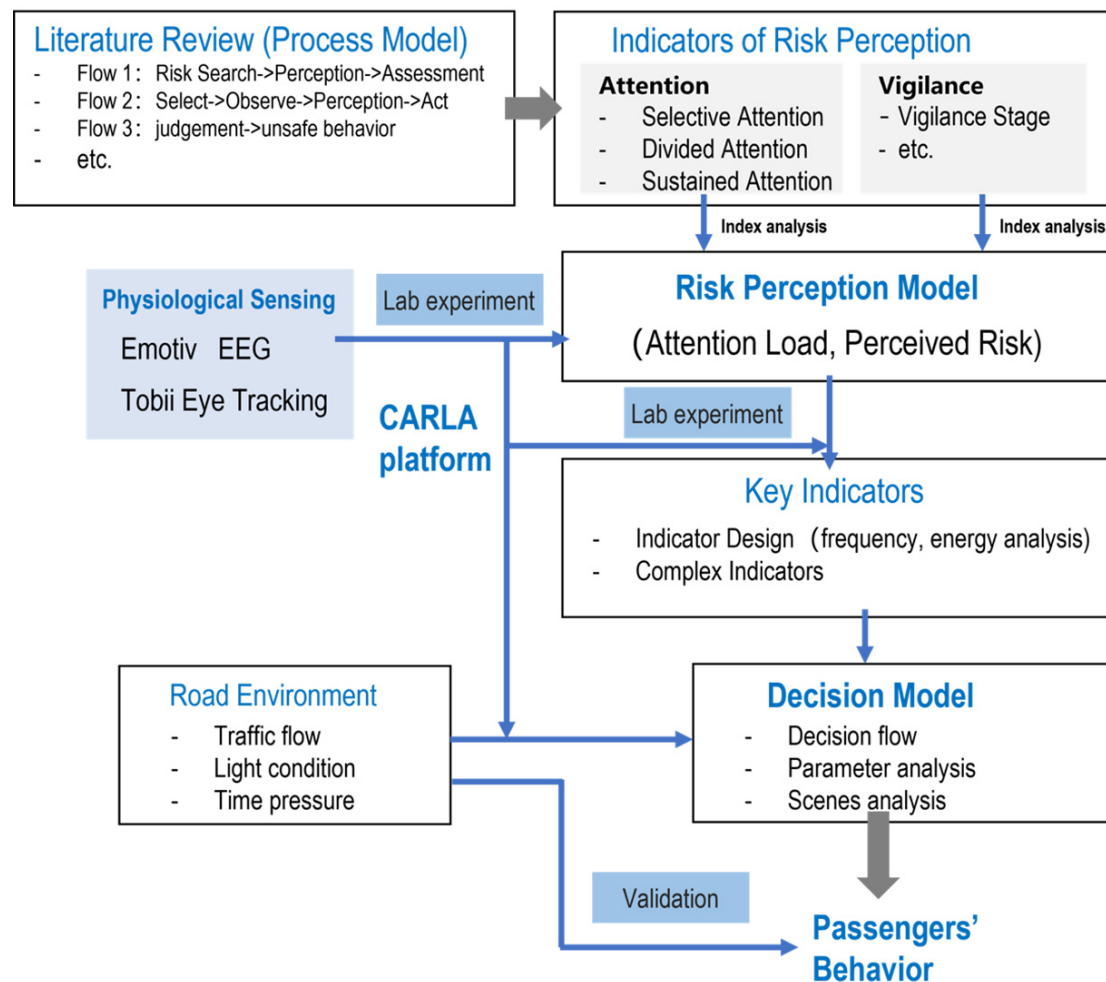


Figure 1. Workflow of the study

2.1 Pedestrian-vehicle Interaction Platform

In this research, we utilize the Python API authorized by CARLA Simulator to develop various intersection traffic scenarios, with visual representation based on Unreal Engine 4. These scenarios are

built based on the digital resource of CARLA, including buildings, vehicles, and other infrastructure that closely resemble the real-world conditions expected in research. The implementation of the scenarios is divided into the server (Service Module) and the client (Agent Module). The sever, built on the Ureal Engine, is responsible for presenting the simulation, including the streets, buildings, traffic light changes, sensor computation principles, and more, thus providing a complete simulated world. The client is responsible for inputting Python commands into the Application Programming Interface (API) to update the world provided by the server, including changes in weather, vehicle speeds, and the number of pedestrians. To provide interaction functionalities for the experimenters, the invention calls the Pygame toolkit within the API, designing movement control components for pedestrians and vehicles. Before the simulated scenario is finally presented on the screen, the experimental subjects first receive verbal task instructions through listening, then observe the scene visually, and further use their own judgment and decision-making to determine the necessary actions. They then use the keyboard to perform basic operations such as dragging to adjust the view, controlling the direction and speed of movement, and emergency braking to complete the tasks.

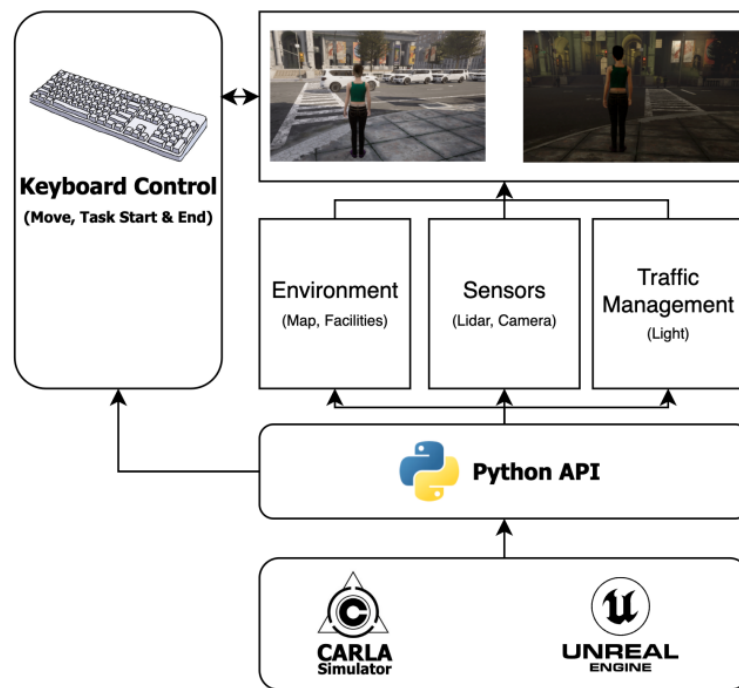


Figure 2. General structure of the interaction platform

2.2 Physiological Data Collection

For EEG collection module, a portable EEG device, the Emotiv Flex Saline with 32 channels, is used with a sampling rate of 120Hz. These data are then transmitted to EmotivPro via Bluetooth in real-time. For eye-tracking collection module, the Tobii Pro Fusion screen-based eye tracker is used, with a sampling rate of 120 Hz. The Tobii Pro Lab is utilized to record corneal infrared reflection information during the subject's screen observation, labeled as visual tracking type. In addition, a 1080P resolution, 30Hz frame rate camera is placed above the display monitor, directly facing the subject's face, and records the subject's facial expressions in real-time during the experiment. To ensure the completeness

of the behavior data, a Python logging tool is deployed to record keyboard operations during the simulation scene at a 30Hz sampling rate. In addition, a 1080P resolution, 30Hz frame rate camera is placed above the display monitor, directly facing the subject's face, and records the subject's facial expressions in real-time during the experiment. Before the experimental tasks, the basic information of the subjects is collected, such as traffic experience, personality, and risk tendency. For each task, the subject's cognitive load and the immediate and retrospective safety self-assessment are assessed by questionnaires.

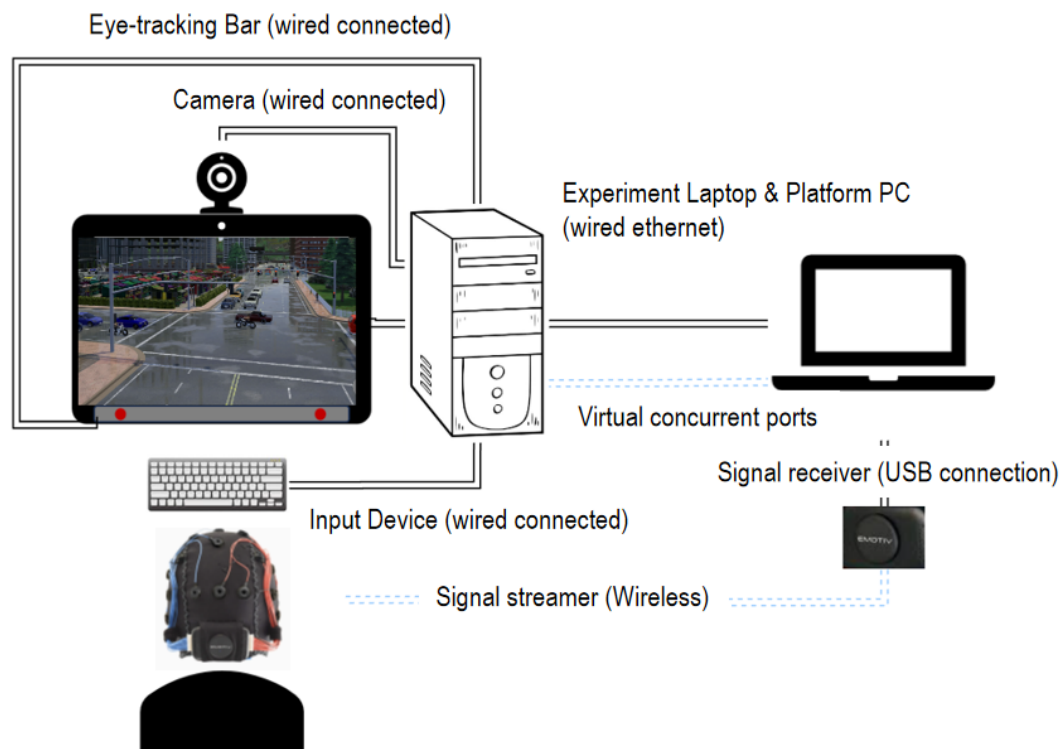


Figure 3. Setup for data collection

2.3 Data Analysis

The final establishment of the risk perception measurement model in this study employs EEG, eye movement, and emotional indicators as features. The specific feature extraction steps include:

- (1) EEG indicators: the artifact removal and feature extraction are performed based on band-pass filtering, re-referencing, independent component analysis (ICA) and time-frequency transformation through wavelet packet transform (WPT). The indicators value then can be calculated by the frequency domain energy during each task period. These indicators are related one's attention, vigilance, mental fatigue and other aspects associated with cognitive state.
- (2) Eye movement indicators: In this procedure, the area of interest (AOI) and time of interest (TOI) method is used to assess how participants allocate their attention and what specific elements draw their focus within the visual field during their crossing. Based on different types and characteristics, the eye movement indicators mainly include fixation duration, fixation point coordinated, saccade velocity, saccade amplitude, and blink frequency.
- (3) Emotional indicators: The frame extraction is performed, followed by detecting and aligning faces using computer vision algorithms to standardize facial position and posture, and removing

background and lighting noise. Facial key points (e.g., corners of eyes, eyebrows, corners of mouth) are extracted frame by frame to analyze facial muscle movements, and Local Binary Pattern (LBP) is used to monitor skin texture and subtle expression changes. Based on the Facial action coding system (FACS), action units related to specific emotions are extracted and analyzed, such as eyebrow raising, eyelid contraction, and mouth corner elevation. Ultimately, these features are used to calculate the basic emotional valence activation level, duration, and frequency during task periods.

3. Discussion

The experimental data includes a total of 90.75 hours of EEG signals, eye tracking, facial recordings, keyboard events, and 1,155 completed surveys across various task segments. Upon reviewing the gaze sequences and durations of all participants after the experiment, we found that some individuals consistently focused their visual attention on themselves within the scene during the interaction, despite the fact that this gaze habit was highly inefficient for acquiring environmental information. Preliminary findings indicate that both subjective and objective factors can impact VRU's risk perception, including personal traits and environmental conditions. Individuals with a high level of risk aversion tend to be more sensitive to potential dangers and may avoid risky situations. Past experiences, such as having been involved in a traffic accident, significantly shape one's risk perception, often leading to heightened vigilance. Additionally, personality traits such as anxiety and decision-making styles—whether intuitive or analytical—also play crucial roles in how risks are evaluated. On the environmental side, road design elements such as width, lighting, signage, and traffic flow can affect pedestrians' risk perception, with poorly designed roads likely increasing perceived danger. The surrounding environment, whether urban or rural, further influences this perception, as urban areas typically present more noise and traffic, resulting in heightened awareness of risk. Social norms and the behavior of others, along with weather conditions such as rain, snow, or fog, also contribute to how pedestrians assess potential hazards, emphasizing the need for a comprehensive understanding of these intertwined factors in the context of pedestrian safety. However, under real-world conditions, environmental variables are often random. Individuals exhibit significant variability when faced with this randomness. This means that predicting individual behavior, especially, requires careful consideration of how past experiences have shaped their personality and thought processes. In that condition, integrating eye-tracking and EEG can provide a more comprehensive view of risk perception and cognitive engagement. For instance, researchers can analyze eye movement patterns while simultaneously recording brain activity, allowing for a deeper understanding of how visual attention correlates with cognitive responses to risks. This combined approach can be particularly valuable where understanding the interplay between visual attention and cognitive processing is crucial for optimizing safety and effectiveness.

4. Conclusion

This study underscores the critical role of understanding risk perception among vulnerable road users (VRUs) in improving traffic safety in smart cities. Through a multimodal approach integrating electroencephalography (EEG) and eye-tracking technologies, we developed a comprehensive framework for assessing the cognitive and physiological factors that influence VRUs' perception of risk in various traffic scenarios. The integration of physiological data into traffic safety analysis allows for a more nuanced understanding of VRUs' decision-making processes. By analyzing the interplay between visual attention and cognitive responses, our framework enhances predictive capabilities for VRU actions, especially in complex urban environments. This research not only contributes to the theoretical

understanding of risk perception mechanisms but also offers practical implications for designing safer urban infrastructure that considers human factors. Future studies should aim to expand on this work by exploring real-world applications and validating the proposed model in diverse urban settings.

Acknowledgment

This work was financially supported by Mercedes-Benz Group China Ltd. -Tsinghua Joint Research Institute (#04191400124/20223910001). Any opinions, findings, conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of Mercedes-Benz Group China Ltd.

- [1] H. Chen, Y. Liu, C. Hu, X. Zhang, Vulnerable Road User Trajectory Prediction for Autonomous Driving Using a Data-Driven Integrated Approach, *IEEE Trans. Intell. Transp. Syst.* 24 (2023) 7306–7317. <https://doi.org/10.1109/TITS.2023.3254809>.
- [2] H. Chen, Y. Liu, C. Hu, X. Zhang, Vulnerable Road User Trajectory Prediction for Autonomous Driving Using a Data-Driven Integrated Approach, *IEEE Trans. Intell. Transp. Syst.* 24 (2023) 7306–7317. <https://doi.org/10.1109/TITS.2023.3254809>.
- [3] D.D. Dinh, N.H. Vu, R.C. Mcllroy, K.A. Plant, N.A. Stanton, Effect of attitudes towards traffic safety and risk perceptions on pedestrian behaviours in Vietnam, *IATSS Res.* 44 (2020) 238–247. <https://doi.org/10.1016/j.iatssr.2020.01.002>.
- [4] S. Rankavat, G. Tiwari, Pedestrians risk perception of traffic crash and built environment features - Delhi, India, *Saf. Sci.* 87 (2016) 1–7. <https://doi.org/10.1016/j.ssci.2016.03.009>.
- [5] S. Rankavat, G. Tiwari, Influence of actual and perceived risks in selecting crossing facilities by pedestrians, *TRAVEL Behav. Soc.* 21 (2020) 1–9. <https://doi.org/10.1016/j.tbs.2020.05.003>.
- [6] F. Luque, V. Armada, L. Piovano, R. Jurado-Barba, A. Santamaria, Understanding Pedestrian Cognition Workload in Traffic Environments Using Virtual Reality and Electroencephalography, *ELECTRONICS* 13 (2024) 1453. <https://doi.org/10.3390/electronics13081453>.
- [7] B.C.K. Choi, A.W.P. Pak, A catalog of biases in questionnaires., *Prev. Chronic. Dis.* 2 (2005) A13.
- [8] Cognitive Internet of Vulnerable Road Users in Traffic: Predictive Neural Modulations of Road Crossing Intention-All Databases, (n.d.). <https://webofscience.clarivate.cn/wos/allldb/full-record/INSPEC:25140046> (accessed October 10, 2024).
- [9] Y. Ma, X. Yan, Changes of EEG signals in vehicle-pedestrian conflicting events with Front Collision Warning on side-parking road, in: G. Tan, F. Cen (Eds.), *Int. Conf. Intell. TRAFFIC Syst. SMART CITY ITSSC 2021*, Spie-Int Soc Optical Engineering, Bellingham, 2022: p. 121651A. <https://doi.org/10.1117/12.2627944>.
- [10] D. Wang, H. Li, J. Chen, Detecting and measuring construction workers' vigilance through hybrid kinematic-EEG signals, *Autom. Constr.* 100 (2019) 11–23. <https://doi.org/10.1016/j.autcon.2018.12.018>.
- [11] Y. Wang, Y. Huang, B. Gu, S. Cao, D. Fang, Identifying mental fatigue of construction workers using EEG and deep learning, *Autom. Constr.* 151 (2023) 104887. <https://doi.org/10.1016/j.autcon.2023.104887>.
- [12] H. Jebelli, S. Hwang, S. Lee, EEG-based workers' stress recognition at construction sites, *Autom. Constr.* 93 (2018) 315–324. <https://doi.org/10.1016/j.autcon.2018.05.027>.
- [13] Y. Peng, Q. Xu, S. Lin, X. Wang, G. Xiang, S. Huang, H. Zhang, C. Fan, The Application of Electroencephalogram in Driving Safety: Current Status and Future Prospects, *Front. Psychol.* 13 (2022)

919695. <https://doi.org/10.3389/fpsyg.2022.919695>.

[14] D. Babic, D. Babic, M. Fiolic, M. Ferko, Factors affecting pedestrian conspicuity at night: Analysis based on driver eye tracking, *Saf. Sci.* 139 (2021) 105257. <https://doi.org/10.1016/j.ssci.2021.105257>.

[15] J. de Winter, P. Bazilinskyy, D. Wesdorp, V. de Vlam, B. Hopmans, J. Visscher, D. Dodou, How do pedestrians distribute their visual attention when walking through a parking garage? An eye-tracking study, *ERGONOMICS* 64 (2021) 793–805. <https://doi.org/10.1080/00140139.2020.1862310>.

ENHANCING THE PERFORMANCE OF MULTI-OBJECT TRACKING IN TRAFFIC STREAM VIDEOS THROUGH INITIAL VELOCITY AND FRAME-SKIPPING STRATEGIES

Chia-Chun Lin¹, Albert Y. Chen²

1) Graduate Research Assistant, Department of Civil Engineering, National Taiwan University, Taiwan. Email: r12521522@ntu.edu.tw

2) Professor, Department of Civil Engineering, National Taiwan University, Taiwan. Email: albertchen@ntu.edu.tw

Abstract: Multi-object tracking in videos is an important task in various domains, such as traffic engineering and construction management. This paper proposes two methods, Grid Mean State and InCo-Skip, to improve multi-object tracking performance, particularly under frame-skipping scenarios. The study focuses on traffic flow counting, using YOLOv8 for vehicle tracking. Initial tests show that while car tracking remains accurate, motorcycles suffer a significant accuracy degradation when homogeneous frame skipping is applied. Grid Mean State addresses the issue by utilizing velocity vectors from earlier frames, and InCo-Skip provides an alternative skipping strategy to balance computational efficiency and accuracy. The combined methods show a substantial enhancement in counting accuracy, achieving up to 28.2% improvement for motorcycles under challenging conditions.

Keywords: Object tracking, Kalman filter, Initial velocity, Frame skip

1. INTRODUCTION

Object tracking is a critical task in various fields, including traffic analysis (Chen et al., 2020), Construction Safety Management (Lin et al., 2021), infrastructure inspection (Wang et al., 2021), and Mixed Reality (Kinoshita et al., 2022). Accurately tracking objects, especially in real-time, presents challenges in achieving a balance between computational resources and accuracy.

In traditional Multi-Object Tracking (MOT) algorithms, such as ByteTrack (Zhang et al., 2022), the Kalman filter is widely used for motion prediction, where the state vector includes parameters such as position, aspect ratio, and velocity. However, the assumption of zero initial velocity during the first few frames can lead to unsuccessful tracking, particularly under Homogenous frame-skipping scenarios. This issue is amplified when tracking smaller objects like motorcycles, where motion dynamics change rapidly.

As a result, this study aims to address the limitations of the zero initial velocity assumption by proposing a method that improves the Kalman filter's initial velocity estimation. At the same time, this study hopes to introduce better frame-skipping strategies to mitigate the amplification of initial velocity inaccuracies during frame skipping.

2. LITERATURE REVIEW

2.1 Introduction to Kalman Filter

The Kalman Filter (Kalman, 1960), first introduced in the 1960s, was originally applied in spacecraft navigation and signal processing to estimate system states with noise and uncertainty. Its principle lies in modeling the motion of a target through a state-space model, which helps to predict future states while reducing noise interference by continuously updating the Kalman Filter through measurements. This method is most effective under the assumption that both the system model and the noise follow Gaussian distributions.

The Kalman filter is based on two primary equations: the prediction and the update.

(1) Prediction

$$\hat{x}_k = F_k x_{k-1} + B_k u_k \quad (1)$$

$$P_k = F_k P_{k-1} F_k^T + Q_k \quad (2)$$

Equation (1) predict the state \hat{x}_k at the next time step based on the previous state \hat{x}_{k-1} and control inputs u_k . In equation (2), P_k is the covariance matrix that reflects the uncertainty in the state prediction, with F representing the state transition matrix and Q_k representing the process noise.

(2) Update

$$K_k = P_k H_k^T (H_k P_k H_k^T + R_k)^{-1} \quad (3)$$

$$\hat{x}_k = \hat{x}_k + K_k(z_k - H_k\hat{x}_k) \quad (4)$$

$$P_k = (I - K_k H_k)P_k \quad (5)$$

From Equation (3) and (4), K_k is the Kalman gain, which determines how much the predictions should be adjusted based on the new measurements z_k . H represents the measurement matrix, and R accounts for measurement noise. At last, Equation (5) updates the covariance matrix P_k by applying the K_k .

2.2 Kalman Filter in Multi-Object Tracking (MOT) and ByteTrack

The Kalman Filter, as introduced earlier, has evolved significantly since its initial application in aerospace and signal processing. By the 2010s, it became a cornerstone in MOT algorithms due to its ability to estimate and predict the states of objects moving through space (Li et al., 2010; Pathan et al., 2009; Weng et al., 2006). In MOT, the Kalman Filter models each object's motion using a state vector, which typically represents the object's position, velocity, and sometimes shape attributes. A common form of the state vector in MOT is:

$$\text{state vector} = [x, y, a, h, \dot{x}, \dot{y}, \dot{a}, \dot{h}]$$

Where:

- x and y represent the center coordinates of the bounding box.
- a denotes the aspect ratio of the bounding box.
- h is the height of the bounding box.
- \dot{x} and \dot{y} are the velocities in the x and y directions, respectively.
- \dot{a} is the rate of change of the aspect ratio.
- \dot{h} is the rate of change of the height.

This allows the Kalman Filter to predict both the object's future position and its motion behavior. However, a key challenge in these algorithms is the estimation of initial velocity values, which significantly affect tracking performance. Many algorithms often default to a zero initial velocity, which may cause either failure of tracking, or tracking delays and inaccuracies until the velocity is correctly estimated after several frames.

Recent advances in MOT, such as ByteTrack, combine the Kalman Filter with IoU (Intersection over Union) (Rezatofighi et al., 2019) metrics for object association. As shown in Figure 1, IoU is used to compare bounding boxes across frames to link detections, while the Kalman Filter predicts motion between frames. Although this combination enhances tracking performance, the issue of poor initial velocity estimation remains, especially under conditions where frame skipping is required to reduce computational load. Frame skipping reduces the number of frames available for updates, making accurate initial velocity estimation even more critical for maintaining tracking accuracy.

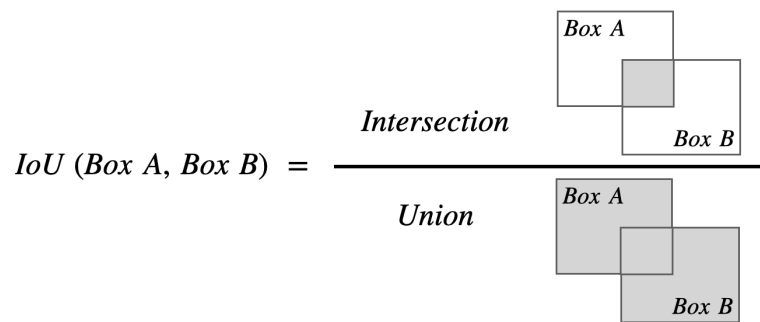


Figure 1. IoU Calculation Principle

Numerous studies (Linderroth et al., 2011; Weiner, 1981; Zhao & Huang, 2017) have highlighted the importance of accurate initialization of the Kalman Filter. While these studies focus on general initialization strategies for the Kalman filter, they do not specifically address initial velocity assumptions. Additionally, research on frame-skipping strategies (Park et al., 2020), introduces algorithms to determine when to skip frames, which adds computational complexity rather than simplifying it.

In response to these issues, our study focuses on developing methods to improve initial velocity assumption and testing new frame-skipping strategies to maintain accuracy while minimizing additional computational overhead across various tracking scenarios.

3. METHODOLOGY

In this section, we describe two key methods employed in our tracking system: Grid Mean State and InCo-Skip Method. These methods serve to enhance accuracy of initial trajectory estimation and efficiency in multi-object tracking systems, ensuring that velocity estimations are initialized correctly and frame-skipping strategies are optimized.

3.1 Grid Mean State

The Grid Mean State method is used to initialize the velocity vectors in our tracking system. This method is divided into two primary phases, namely the Initial Tracking Phase and Full Video Tracking Phase. The process is illustrated in Figure 2. In Initial Tracking Phase, we calculate Grid Mean State by leveraging the state vectors of objects being successfully tracked. Then we use Grid Mean State to replace the initial zero velocity assumption in subsequent Full Video Tracking Phase.

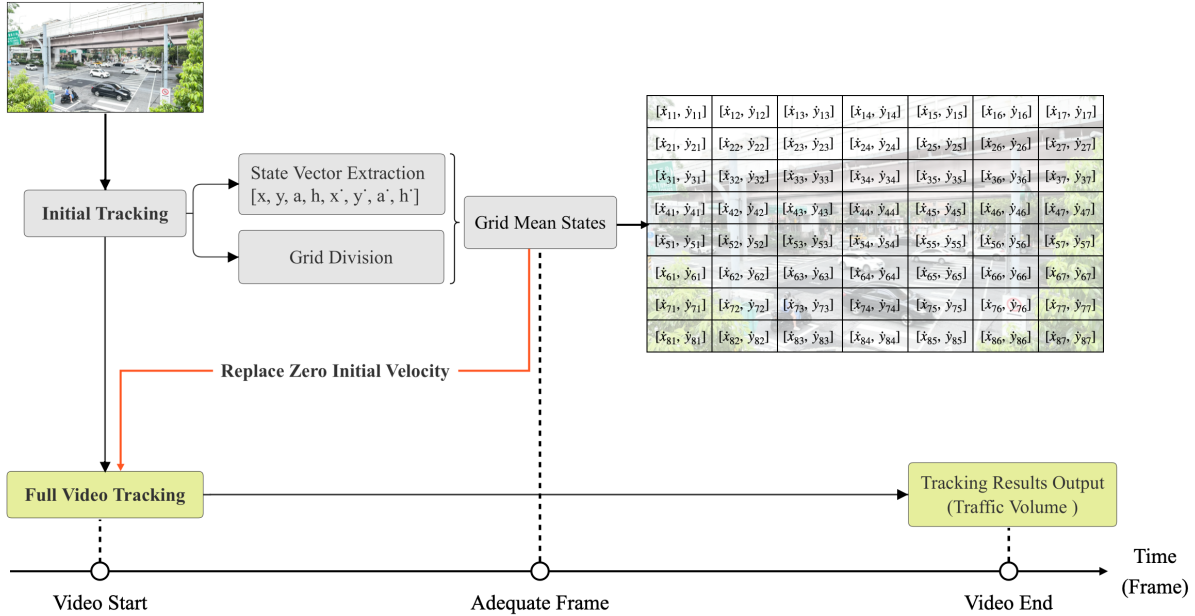


Figure 2. Grid Mean State Calculation Process

The procedure can be broken down into the following steps:

(1) State vector extraction

In the Initial Tracking Phase, YOLOv8 is used to detect and track objects in the video frame. For each tracked object, the state vector at a given frame t is defined as follows:

$$\text{state vector}(t) = [x, y, a, h, \dot{x}, \dot{y}, \dot{a}, \dot{h}]$$

(2) Grid division and object assignment

The entire frame is divided into a grid of predefined size, such as 16x9 (for a 16:9 aspect ratio video), where each grid cell corresponds to a specific section of the frame. Each grid cell is represented by its coordinates (i, j) , where i is the horizontal index, and j is the vertical index.

Once the frame is divided into grids, the position (x, y) of each successfully tracked object is used to determine which grid cell the object belongs to at each time step. Specifically, for an object located at coordinates (x, y) , we can determine the grid it belongs to by Equation (6).

$$i = \left\lfloor \frac{x}{\text{grid width}} \right\rfloor, \quad j = \left\lfloor \frac{y}{\text{grid height}} \right\rfloor \quad (6)$$

(3) Grid Mean State computation

For each grid cell (i, j) , we compute the Grid Mean State based on the state vectors of all objects that appear in that grid during the Initial Tracking Phase. The mean state is calculated for the velocity components $[\dot{x}, \dot{y}, \dot{a}, \dot{h}]$. If no objects are tracked in a particular grid, the mean state for that grid is set to zero.



Figure 3. Visualization of Grid Mean State with Velocity Arrows

(4) Usage of Grid Mean States for full video tracking

The calculated Grid Mean State is then used to replace the initial zero velocity assumption in subsequent Full Video Tracking Phase. This ensures that the initial trajectory estimation is more accurate, particularly for small or fast-moving objects. By improving the accuracy of the first few frames of tracking, we allow the Kalman filter's updates to take over and maintain accurate object trajectories in later frames.

3.2 InCo-Skip (Inhomogeneous and Counterintuitive Frame Skip)

In video tracking, reducing computational load is crucial, especially when dealing with real-time processing or large datasets. A common approach is the Homogeneous Skip method, where alternate frames (1, 3, 5, etc.) are processed, skipping every other frame. While this method effectively cuts down computation by 50%, it introduces a major drawback related to initial tracking accuracy. Specifically, with homogeneous frame skipping, the first frame's position must be used to predict the object's position in the third frame. Due to the zero-initial velocity assumption in the Kalman filter, the predicted position in the third frame aligns with the position in the first frame, and the comparison for object matching is made between the first and third frames rather than the first and second frames. This mismatch significantly reduces the success rate of initial tracking, particularly for small or fast-moving objects, which might not be tracked at all.

In response to this issue, InCo-Skip is introduced. Instead of uniformly skipping every other frame, this method processes consecutive pairs of frames (e.g., frames 1 and 2), then skips the next two frames (3 and 4), and continues by processing frames 5 and 6, skipping frames 7 and 8, and so on. This ensures that the tracking system has at least two consecutive frames for comparison which is similar to not skipping any frames at all during initial tracking. Once initial tracking is successful, skipping frame pairs is less detrimental to accuracy because the Kalman filter updates the object state using its learned velocity and direction from the previous frames. Even after skipping frame pairs, the filter can make stable predictions and maintain tracking continuity.

We named this method InCo-Skip, which stands for Inhomogeneous and Counterintuitive Frame Skip. The name reflects both the non-uniform skipping pattern and the method's unconventional approach of skipping consecutive frames while maintaining tracking stability. The Figure 4 shows the comparison between Homogeneous Skip and InCo-Skip.

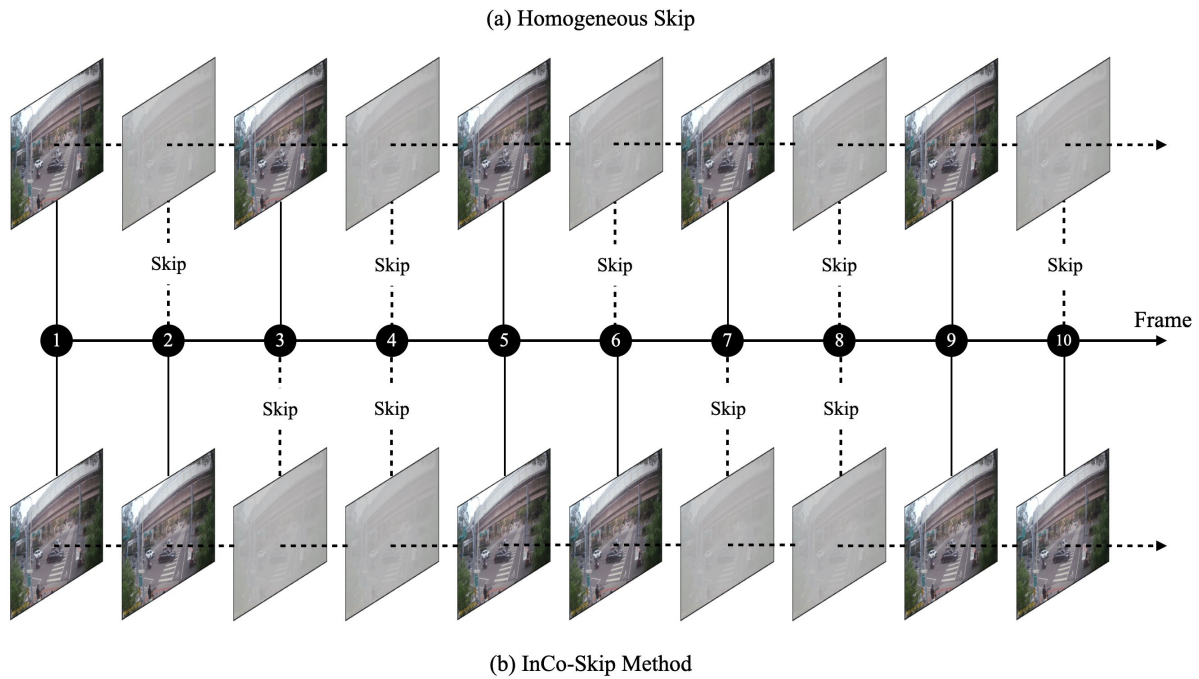


Figure 4. (a) Homogeneous Skip and (b) InCo-Skip Method

4. RESULTS AND DISCUSSIONS

We evaluate the effectiveness of the two methods introduced in Section 3, Grid Mean State and InCo-Skip, in improving tracking accuracy, particularly for motorcycles under Homogenous frame-skipping scenarios. The primary metric we are concerned with is Counting Accuracy, which refers to how accurately the tracking system, based on YOLOv8, can count vehicles compared to manual counts.

Tracking accuracy is also influenced by the IoU threshold, which is a crucial parameter in the tracking process. A well-chosen IoU threshold can enhance tracking performance and improve the overall accuracy of vehicle counts. Therefore, in the upcoming tests, we will compare the effects of different IoU values on tracking accuracy to understand their influence on our evaluation of traffic flow.

This study conducts initial testing on a five-minute video at 30 frames-per-second (fps), capturing traffic flow as shown in Table 1. The accuracy for cars remains consistent even under Homogeneous Skip. However, accuracy of motorcycles drops significantly, likely due to smaller object sizes and faster movement speeds. This observation motivates the need for more robust methods to ensure that frame skipping, which is essential for real-time processing or large datasets, does not lead to substantial accuracy loss. Thus, we test and evaluate the Grid Mean State and InCo-Skip methods to see how they mitigate this issue.

Table 1. Accuracy comparison of car and motorcycle in No Skip and Homogeneous Skip

Vehicle Type	*IoU threshold	Accuracy (%)		Difference
		No Skip	Homogeneous Skip	
Car	0.20 ~ 0.01	98.6	98.6	0
	0.20	98.5	54.9	↓ 43.6
Motorcycle	0.10	98.5	63.4	↓ 35.1
	0.05	98.5	66.2	↓ 32.3
	0.01	98.5	77.5	↓ 21.0

* In YOLOv8, the value of 1 - IoU is used as the match threshold.

4.1 Effect of Grid Mean State on motorcycle accuracy under frame-skipping scenarios

This section compares motorcycle counting accuracy using a zero-initial velocity state versus the Grid Mean State method under frame-skipping scenarios. As shown in Table 4.2, applying the Grid Mean State improves motorcycle accuracy under Homogenous frame-skipping conditions, accuracy increases from 54.9% to 60.6% at IoU = 0.20. However, the effect of the Grid Mean State diminishes at lower IoU thresholds.

Conversely, in the InCo-Skip scenario, the Grid Mean State method shows clear improvements across all IoU thresholds, with an increase in accuracy of up to 18.3% at IoU = 0.20. However, as the IoU value decreases, the increase in accuracy becomes less significant. In summary, we observe that regardless of whether in the Homogeneous Skip or InCo-Skip scenario, the Grid Mean State has a more significant impact at higher IoU values.

Table 2. Motorcycle accuracy using Grid Mean State vs. Zero Initial Velocity

Skip Method	IoU threshold	Accuracy (%)		Difference
		Zero Initial Velocity	Grid Mean State	
Homogenous	0.20	54.9	60.6	↑ 5.7
	0.10	63.4	63.4	0
	0.05	66.2	64.8	↓ 1.4
	0.01	77.5	77.5	0
InCo-Skip	0.20	39.4	57.7	↑ 18.3
	0.10	78.9	87.3	↑ 8.4
	0.05	88.7	94.4	↑ 5.7
	0.01	95.8	95.8	0

4.2 Effect of InCo-Skip vs. Homogeneous Skip on motorcycle accuracy

In this section, this study compares the performance of the InCo-Skip method against Homogeneous Skip in terms of motorcycle counting accuracy, focusing on the impact of different IoU thresholds and the initial velocity assumption (zero velocity or Grid Mean State). As shown in Table 3, the results highlight some key patterns.

The InCo-Skip method underperforms compared to Homogeneous Skip at a higher IoU threshold (0.20), with a 15.5% and 2.9% decrease in accuracy. However, as the IoU threshold decreases, InCo-Skip shows a remarkable improvement, outperforming Homogeneous Skip by up to 29.6% at IoU=0.05. This suggests that InCo-Skip is particularly effective in lower IoU value.

Table 3. Motorcycle accuracy using Homogeneous Skip vs. InCo-Skip

Initial Velocity	IoU threshold	Accuracy (%)		Difference
		Homogeneous Skip	InCo-Skip	
Zero	0.20	54.9	39.4	↓ 15.5
	0.10	63.4	78.9	↑ 15.5
	0.05	66.2	88.7	↑ 22.5
	0.01	77.5	95.8	↑ 18.3
Grid Mean State	0.20	60.6	57.7	↓ 2.9
	0.10	63.4	87.3	↑ 23.9
	0.05	64.8	94.4	↑ 29.6
	0.01	77.5	95.8	↑ 18.3

4.3 Combined performance of Grid Mean State and InCo-Skip methods

In this section, we evaluate the combined effect of using both the Grid Mean State and InCo-Skip methods on motorcycle counting accuracy. As observed in previous sections, the Grid Mean State method shows better accuracy improvements at higher IoU thresholds, while the InCo-Skip method performs better at lower IoU thresholds. Therefore, the combination of these two methods aims to leverage the strengths of both, mitigating each other's limitations.

Table 4 demonstrates that when using both methods, the motorcycle counting accuracy improves across all IoU thresholds compared to using only the Zero Initial Velocity with Homogeneous Skip. At higher IoU thresholds, the improvement is modest, but at lower IoU thresholds, the accuracy gains are more pronounced, showing notable improvements in counting accuracy of up to 28.2%."

Table 4. Motorcycle accuracy with both methods vs. Zero Initial Velocity + Homogeneous Skip

IoU threshold	Accuracy (%)		Difference
	Baseline	Both Methods	
0.20	54.9	57.7	↑ 2.8
0.10	63.4	87.3	↑ 23.9
0.05	66.2	94.4	↑ 28.2
0.01	77.5	95.8	↑ 18.3

To provide a clear recommendation for practical use, our results indicate that setting the IoU threshold between 0.05 and 0.10 yields the most consistent improvement in tracking accuracy when both the Grid Mean State and InCo-Skip methods are applied. By combining these two methods, it effectively balances the strengths of each, reducing the shortcomings observed when using either method in isolation.

5. CONCLUSIONS

In this paper, this study proposed two methods, Grid Mean State and InCo-Skip, to address the challenge of maintaining tracking accuracy for motorcycles in frame-skipping scenarios. Our results showed that while car accuracy remained high, motorcycle accuracy significantly dropped during homogeneous skipping. The Grid Mean State method was more effective at higher IoU thresholds, while InCo-Skip performed better at lower IoU thresholds.

By combining both methods, we achieved improved accuracy across all conditions, highlighting the complementary strengths of each approach. These findings suggest that our methods provide a practical solution for maintaining high tracking accuracy while accommodating frame skipping demands, which is particularly beneficial in real-time applications and when handling large datasets."

6. FUTURE WORK

In this study, the primary focus was on evaluating the effectiveness of the proposed methods based on counting accuracy, using a single test video for performance assessment. While this approach provided initial insights into the methods' potential, there are several areas for further exploration and improvement.

Future work should involve testing on a wider range of publicly available datasets like MOT17 and BDD100k to validate the methods' robustness across diverse scenarios. Additionally, while counting accuracy was the main metric, future research should incorporate other multi-object tracking (MOT) performance metrics, such as ID switch, MOTA (Multiple Object Tracking Accuracy), MOTP (Multiple Object Tracking Precision), and Track Fragmentation. These metrics will provide a deeper understanding of the proposed methods' strengths and limitations, assessing not only accuracy but also stability in more challenging environments.

ACKNOWLEDGMENTS

The authors thank the support from the National Science and Technology Council of Taiwan for grant NSTC 113-2628-E-002-026-MY3.

REFERENCES

- Chen, A. Y., Chiu, Y.-L., Hsieh, M.-H., Lin, P.-W., & Angah, O. (2020). Conflict analytics through the vehicle safety space in mixed traffic flows using UAV image sequences. *Transportation Research Part C: Emerging Technologies*, 119, 102744.
- Kalman, R. E. (1960). *A new approach to linear filtering and prediction problems*.
- Kinoshita, A., Fukuda, T., & Yabuki, N. (2022). Enhanced Tracking Method with Object Detection for Mixed Reality in Outdoor Large Space. *Legal Depot D/2022/14982/02*, 457.
- Li, X., Wang, K., Wang, W., & Li, Y. (2010). A multiple object tracking method using Kalman filter. *The 2010 IEEE International Conference on Information and Automation*, 1862–1866.
- Lin, Z.-H., Chen, A. Y., & Hsieh, S.-H. (2021). Temporal image analytics for abnormal construction activity identification. *Automation in Construction*, 124, 103572.
- Linderöth, M., Soltész, K., Robertsson, A., & Johansson, R. (2011). Initialization of the Kalman filter without assumptions on the initial state. *2011 IEEE International Conference on Robotics and Automation*, 4992–4997.
- Park, J. W., Kim, J., & Lee, H.-J. (2020). Fast Object Detection Using a Frame Skip Method. *2020 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia)*, 1–2.
- Pathan, S. S., Al-Hamadi, A., & Michaelis, B. (2009). Intelligent feature-guided multi-object tracking using Kalman filter. *2009 2nd International Conference on Computer, Control and Communication*, 1–6.
- Rezatofghi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., & Savarese, S. (2019). Generalized intersection over union: A metric and a loss for bounding box regression. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 658–666.
- Wang, M., Kumar, S. S., & Cheng, J. C. P. (2021). Automated sewer pipe defect tracking in CCTV videos based on defect detection and metric learning. *Automation in Construction*, 121, 103438.
- Weiner, L. B. (1981). Kalman Filter Initialization With Large Initial Uncertainty And Strong Measurement Nonlinearity. *Conference Proceedings Southeastcon '81*, 150–151. <https://doi.org/10.1109/SECON.1981.673417>
- Weng, S.-K., Kuo, C.-M., & Tu, S.-K. (2006). Video object tracking using adaptive Kalman filter. *Journal of Visual Communication and Image Representation*, 17(6), 1190–1208.
- Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W., & Wang, X. (2022). Bytetrack: Multi-object tracking by associating every detection box. *European Conference on Computer Vision*, 1–21.
- Zhao, S., & Huang, B. (2017). On initialization of the Kalman filter. *2017 6th International Symposium on Advanced Control of Industrial Processes (AdCONIP)*, 565–570.

AN IMPROVED SAFETY MANAGEMENT METHOD FOR WELDING OPERATIONS BASED ON GAZE ESTIMATION AND OBJECT DETECTION

Ying Liang¹, Yi Tan^{2*} and Dong Jia³

1) Master's Candidate, Sino-Australia Joint Research Center in BIM and Smart Construction, Shenzhen University, Shenzhen, China. Email: 2200474012@email.szu.edu.cn

2) Ph.D., Assoc. Prof., Sino-Australia Joint Research Center in BIM and Smart Construction, Shenzhen University, Shenzhen, China. Email: tanyi@szu.edu.cn

3) Engineer, China Construction Fifth Engineering Division COPR.LTD, China. Email: 973911589@qq.com

Abstract: In the field of construction engineering, particularly in steel structure construction, welding operations are of great importance. However, this activity carries significant safety risks, such as severe accidents like eye burns. Therefore, safety management and protective measures for these high-risk activities are especially crucial. Although computer vision-based object detection algorithms have already been applied in the construction sector, these algorithms generally lack the capability to process high-level semantic information. They can only detect objects but cannot understand the state of the objects. To address this issue, this paper proposes an improved welding safety inspection method based on gaze estimation and object detection. First, from a human-machine interaction perspective, the method combines gaze estimation and object detection to determine the worker's operational status. Second, it performs real-time checks on the personal protective equipment (PPE) used by the welders. Experiments demonstrate the feasibility of this approach, enhancing its ability to respond to complex scenarios and contributing to improved safety levels in construction projects.

Keywords: Gaze estimation, Object detection, Safety management, Welding operations

1. INTRODUCTION

steel structure assembly, prefabricated component connections, and bridge engineering. It primarily serves to connect metals or other materials, ensuring structural integrity and stability. However, the hazardous environment of welding operations presents numerous safety challenges to the construction industry. According to data from the U.S. Bureau of Labor Statistics (BLS), welding accidents account for approximately 25% of fatal occupational accidents and include about one-third of non-fatal amputation incidents. Each year, more than 560,000 workers are injured in welding-related accidents. The U.S. Occupational Safety and Health Administration (OSHA) estimates that there are about 2,500 to 3,000 cases of severe arc flash burns annually, most of which occur on construction sites, particularly during welding on roofs and steel structures. Welding accidents are not only frequent but also extremely severe, highlighting the urgent need to strengthen safety measures and regulations in the construction industry.

Computer vision technology has already been capable of real-time safety detection on construction sites (Weili et al., 2020), including verifying whether workers are properly wearing basic safety equipment such as helmets and reflective vests. However, in certain high-risk construction scenarios, such as welding operations, relying solely on traditional object detection is insufficient to capture richer semantic information, which is necessary for conducting in-depth safety inspections. Gaze estimation technology, as a method that reveals an individual's intent and focus of attention, provides a new dimension for understanding behavior in complex work environments. By analyzing the worker's gaze points, this technology can infer whether their attention is focused on critical tasks, thus identifying potential safety hazards during specific activities.

Therefore, this study integrates object detection and gaze estimation technologies to bridge the "semantic gap" in scene understanding present in traditional object detection. The specific work includes the following: First, an analysis of the operational space limits for welding activities is conducted based on ergonomics. Next, object detection and gaze estimation are combined to assess the worker's operational status, determining whether the worker is engaged in welding or not. Finally, the compliance of the worker's PPE usage is further inspected.

2. RELATED WORK

2.1 Construction Safety Management Based on Computer Vision

On construction sites, workers often fail to wear personal protective equipment (PPE) as required due to factors such as a lack of safety awareness and site constraints (Tom Ka Man et al., 2020). Helmets and reflective vests are essential PPE that must be worn when entering a construction site, and their detection through object recognition has broad applicability. (Huang et al., 2021) used an improved Yolo v3 algorithm to detect whether workers were wearing helmets in compliance with safety standards.

In specific work scenarios, workers are required to wear particular personal protective equipment (PPE). For instance, when working at heights, workers must also wear safety harnesses. (Weili et al., 2018) first used Faster R-CNN to detect the presence of workers and then took the pixel data from the identified bounding boxes

as input for a CNN to recognize whether workers were wearing safety harnesses. (Jiaqi et al., 2022) combined object detection and pose estimation methods to recognize whether workers were wearing helmets and securing safety harnesses correctly. (Lieyun et al., 2018) integrated convolutional neural networks (CNNs) and long short-term memory (LSTM) networks to automatically detect three unsafe behaviors when workers were climbing mobile scaffolding ladders. (Chen et al., 2023) implemented object detection of welding masks using YOLOv5S.

Simple object detection is no longer sufficient for managing safety on construction sites. In certain construction scenarios, it is necessary to determine the worker's operational context and then conduct safety inspections by combining scene information with regulatory requirements. (Numan et al., 2021) used Mask R-CNN and the spatial relationships between objects to infer that a worker was performing scaffold-moving tasks, thereby enabling relevant safety checks. (Weili et al., 2019) used Mask R-CNN to detect and segment workers, steel structure supports, and concrete supports, using pixel overlap to examine the relationship between workers and supports to determine if the worker was on the support structures.

2.1 Related Work on Gaze Estimation

Gaze estimation, as an important method for inferring human behavior, has been widely applied across various fields. In driver assistance systems, it is used to detect the driver's attention state, reducing the risk of accidents caused by distraction or fatigue (X. Fan et al., 2021). In market analysis, gaze estimation helps optimize product displays and advertisement layouts (Bermejo et al., 2020). In education and training, it monitors students' gaze behavior to assess engagement and learning outcomes (Tafasca et al., 2023). In psychology and behavioral research, gaze tracking is used to study social interactions and emotional states (L. F. Fan et al., 2018). Additionally, augmented reality (AR) and virtual reality (VR) systems use gaze prediction to dynamically adjust content, enhancing immersion (Rahman et al., 2019). In public safety and privacy, gaze patterns assist in identifying abnormal behavior, bolstering security measures (Katsini et al., 2020). These applications demonstrate how gaze estimation advances intelligent interactions across different contexts, improving the efficiency of human interactions with devices and environments.

If gaze estimation is applied to the AEC (Architecture, Engineering, and Construction) industry, it could significantly enhance the intelligent management of construction sites. By analyzing workers' gaze directions, their work intentions—such as whether they are focused on the current task, materials, or equipment—can be inferred, ensuring accuracy and efficiency in their work. Additionally, predicting workers' operational status can help prevent potential safety hazards, thereby improving safety management on-site.

3. METHOD

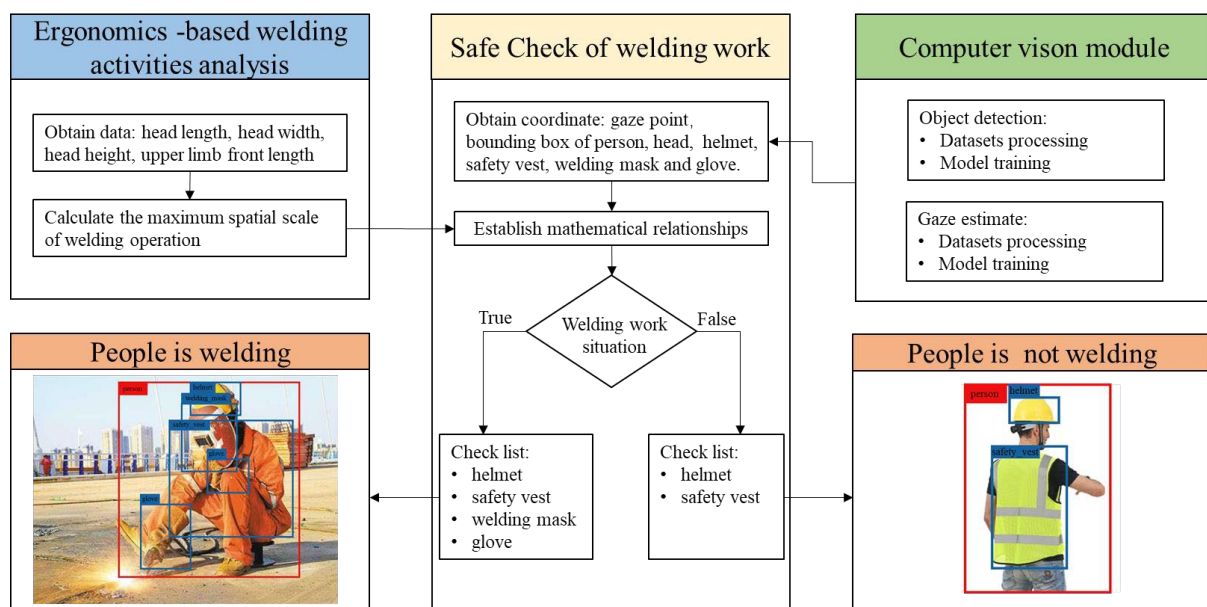


Figure 1. Method of this paper.

This study proposes an improved safety management method for welding work operations based on gaze estimation and object detection. The detailed methodology is illustrated in Figure 1.

- 1) First, coordinate information is obtained based on object detection and gaze estimation, including the gaze point and the bounding boxes for the person, head, helmet, safety vest, welding mask, and gloves.
- 2) Second, an analysis of the welding activity is conducted based on ergonomics. This involves calculating the

worker's maximum operational workspace by acquiring data such as head length, head width, head height, and upper limb front length.

3) Finally, using the theoretical analysis and coordinate relationships mentioned above, mathematical relationships are established to determine whether the worker is engaged in welding. Additionally, checks are performed to ensure that the worker is correctly wearing the helmet, safety vest, welding mask, and gloves.

3.1 Ergonomics -based Welding Activities Analysis

Human body dimensions determine the geometric space occupied by the body and the range of human activities. They are the primary basis for analyzing the spatial range of welding activities. Based on the data from the "Human dimensions of Chinese adults" and the posture analysis of welding activities, the maximum range of welding space is calculated.

During welding operations, workers need to hold the welding torch to perform the task. Therefore, the maximum distance between the eyes and the welding point should be calculated by subtracting the head length from the extended upper limb length, then adding the length of the welding torch, as shown in Figure 2, with the specific calculation provided in Equation (1).

$$\text{spatial scale of welding operation}_{\max} = \text{upper limb front length} - \text{head length} + \text{welding torch length} \quad (1)$$

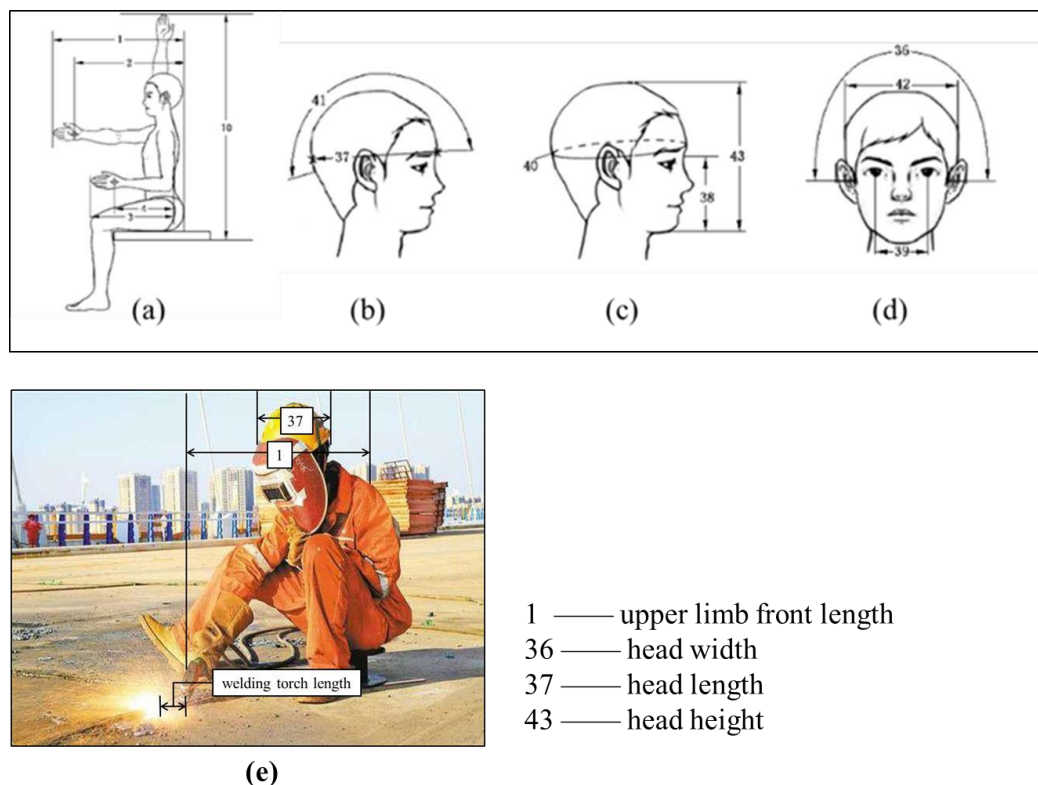


Figure 2. Ergonomics-Based Analysis of Welding Workspace Dimensions

The "Human Dimensions of Chinese Adults" provides data categorized into seven percentiles: the 1st, 5th, 10th, 50th, 90th, 95th, and 99th percentiles. Typically, the 5th, 50th, and 95th percentiles are used to represent body dimensions for individuals with short, average, and tall statures, respectively. According to the "Regulations on Safety Technical Training and Assessment for Special Operation Personnel", special operation personnel must be at least 18 years old and not exceed the statutory retirement age (60 years for men and 50 years for women). The "Big Data Analysis Report on Labor Employment in the Construction Industry (2021)" shows that: By gender, male workers dominate the construction industry, accounting for 89%, while females make up 11%. By age distribution, 38.78% of construction workers are over 50 years old, 26.94% are aged 40-49, 22.98% are aged 30-39, 7.41% are aged 25-29, and 3.89% are under 24. Based on this data, the relevant body dimensions were selected, as shown in Table 1. The maximum value of the d/h ratio, i.e., 2.84, was chosen for further analysis.

Table 1. Human Dimensions of Chinese Adults Data Table

Measureme nt Item	18-25 Years Old Male Body Dimensions			26-35 Years Old Male Body Dimensions			36-60 Years Old Male Body Dimensions		
	Percentile			Percentile			Percentile		
	P5	P50	P95	P5	P50	P95	P5	P50	P95
Upper Limb									
Front Length	760	823	892	760	821	886	759	822	886
Head Length	175	187	200	175	187	200	176	188	200
Head Width	150	160	172	149	160	172	146	157	169
Head Height(h)	217	235	253	213	235	253	209	231	253
Welding Workspace									
Scale(d)	595	646	702	595	644	696	593	644	696
d/h	2.74	2.75	2.77	2.79	2.74	2.75	2.84	2.79	2.75
Measureme nt Item	18-25 Years Old Female Body Dimensions			26-35 Years Old Female Body Dimensions			36-60 Years Old Female Body Dimensions		
	Percentile			Percentile			Percentile		
	P5	P50	P95	P5	P50	P95	P5	P50	P95
Upper Limb									
Front Length	678	749	815	688	750	813	696	757	822
Head Length	166	177	188	166	177	188	167	179	190
Head Width	144	154	164	142	152	163	141	150	161
Head Height(d)	209	228	246	209	228	246	206	227	246
Welding Workspace									
Scale(h)	522	582	637	532	583	635	539	588	642
d/h	2.50	2.55	2.59	2.55	2.56	2.58	2.62	2.59	2.61

3.2 Identification of Welding Activities

Through observing welding activities, it was found that the welder's gaze is typically focused on the welding point during the operation. Therefore, the results from both the gaze estimation model and the object detection model—namely, the gaze point and the welding area—are used to determine whether the worker is engaged in welding, as shown in Figure 3(a). When the coordinate relationships between the gaze point and the welding area simultaneously satisfy Equations (2.1) and (2.2), that is, when the x and y coordinates of the gaze point are within the bounding box of the welding area, it can be concluded that the welder's gaze is focused on the welding point. During welding operations, it is inevitable that supervisors may also be watching the welding point, as depicted in Figure 3(b). Based on the ergonomics analysis of the welding workspace, when the coordinate relationships of the head, eye, and gaze point simultaneously satisfy Equations (2.3) to (2.5), that is, when the head height in the image is calculated using Equation (2.3), the worker's working distance is calculated using Equation (2.4), and then the ratio range between the Welding Workspace Scale and head height, as analyzed in Table 1, is used to determine whether the worker is welding.

$$\text{welding_area_x}_{\min} < x < \text{welding_area_x}_{\max} \quad (2.1)$$

$$\text{welding_area_y}_{\min} < y < \text{welding_area_y}_{\max} \quad (2.2)$$

$$\text{head height} = \text{head_y}_{\max} - \text{head_y}_{\min} \quad (2.3)$$

$$\text{working distance} = \sqrt{(\text{gaze}_x - \text{eye}_x)^2 + (\text{gaze}_y - \text{eye}_y)^2} \quad (2.4)$$

$$\text{working distance} / \text{head_height} \leq 2.84 \quad (2.5)$$

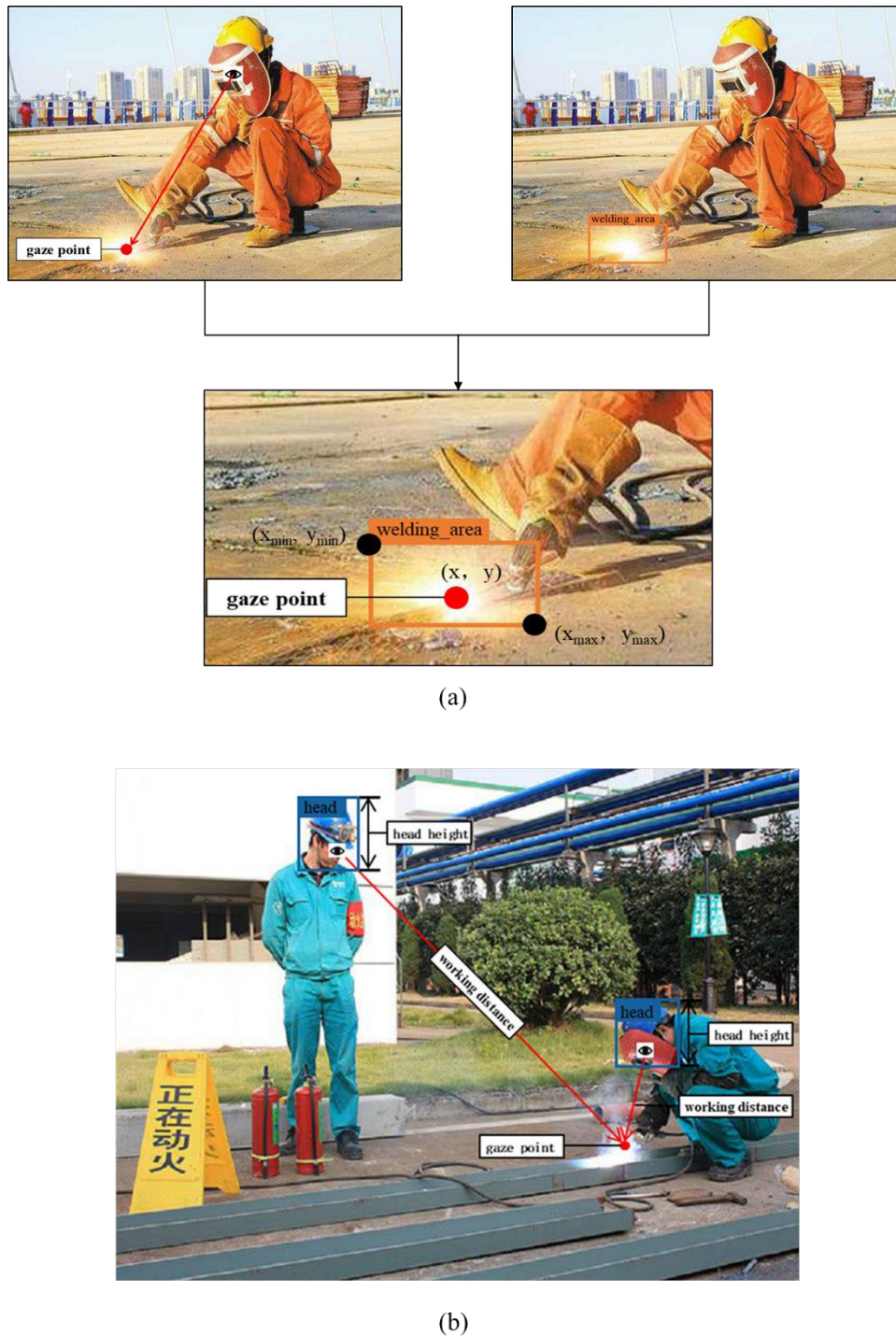


Figure 3. Identification of welding activities

3.3 Identification of PPE

Once it has been determined that the worker is engaged in welding, a specific safety check is required to ensure compliance with PPE requirements. This involves verifying whether the worker is wearing the appropriate PPE, including a helmet, safety vest, welding mask, and gloves as required.

(1) Identification of Helmet

The analysis for identifying workers who are wearing helmets correctly is as follows: The typical height of a safety helmet is about 14 cm. Based on Table 1, the head height of Chinese adults is approximately twice the height of the helmet. If the widths of the bounding boxes for the head and the helmet are the same, the Intersection over Union (IOU) is used to determine whether the worker is wearing a helmet.

When the IOU is greater than or equal to 0.5, the worker is considered to be wearing a helmet. If the IOU is less than 0.5, the worker is not wearing a helmet. See Figure 4 for details.

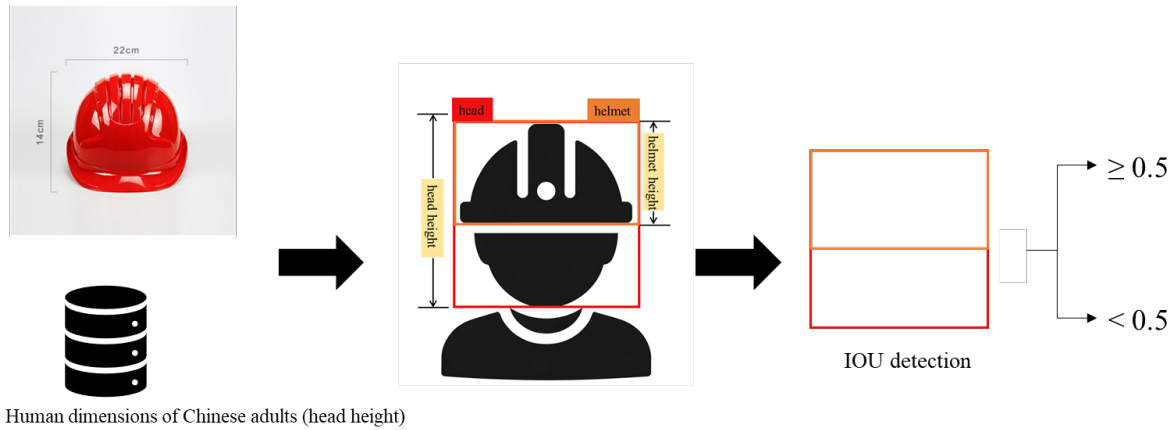


Figure 4. Identification of helmet

(2) Identification of Safety Vest

The analysis for identifying workers correctly wearing a reflective vest is as follows: The bounding box for the safety vest is enclosed within the bounding box for the person. The spatial relationship between these bounding boxes can be expressed by Equations (3.1) and (3.2). See Figure 5 for details.

$$\text{person_x}_{\min} < \text{safety_vest_x}_{\min}, \text{person_y}_{\min} < \text{safety_vest_y}_{\min} \quad (3.1)$$

$$\text{safety_vest_x}_{\max} < \text{person_x}_{\max}, \text{safety_vest_y}_{\max} < \text{person_y}_{\max} \quad (3.2)$$

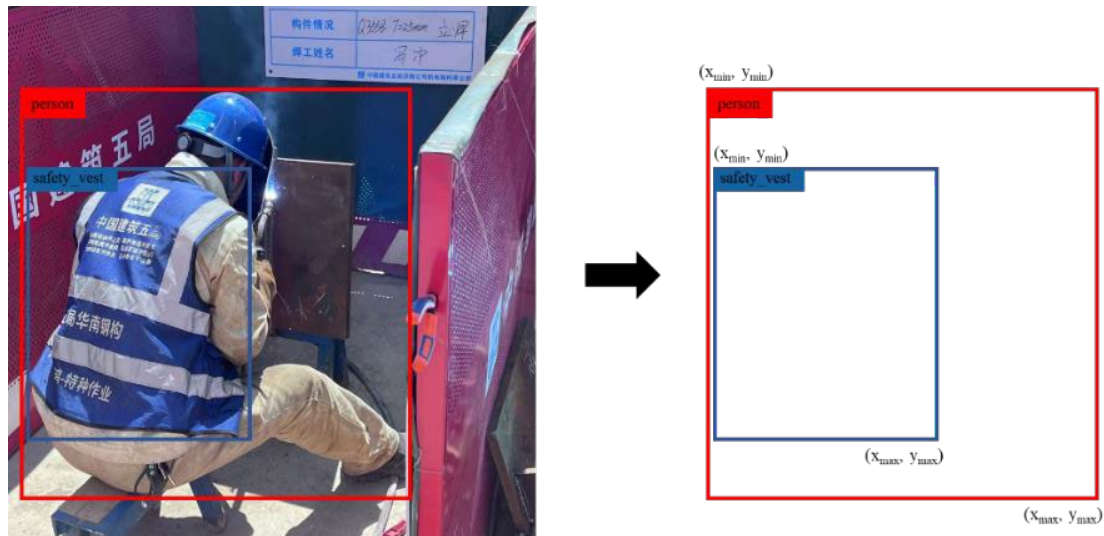


Figure 5. Identification of safety vest

(3) Identification of Weld Mask

The analysis for identifying workers correctly wearing a welding mask is as follows: When a welder is using a welding mask, the gaze arrow will intersect with the bounding box of the mask at two points. The spatial relationship can be represented by Equation (4.1) to Equation (4.5). If the line L intersects with any of the edges L1, L2, L3, or L4 of the mask's bounding box, the worker is considered to be wearing the welding mask. See Figure 6 for details.

$$L: (y - \text{gaze}_y)(\text{gaze}_x - \text{eye}_x) = (\text{gaze}_y - \text{eye}_y)(x - \text{eye}_x) \quad (4.1)$$

$$L1: y = x_{\min} \quad (x_{\min} < x < x_{\max}) \quad (4.2)$$

$$L2: y = x_{\max} \quad (x_{\min} < x < x_{\max}) \quad (4.3)$$

$$L3: x = y_{\min} \quad (y_{\min} < y < y_{\max}) \quad (4.4)$$

$$L4: x = y_{\max} \quad (y_{\min} < y < y_{\max}) \quad (4.5)$$

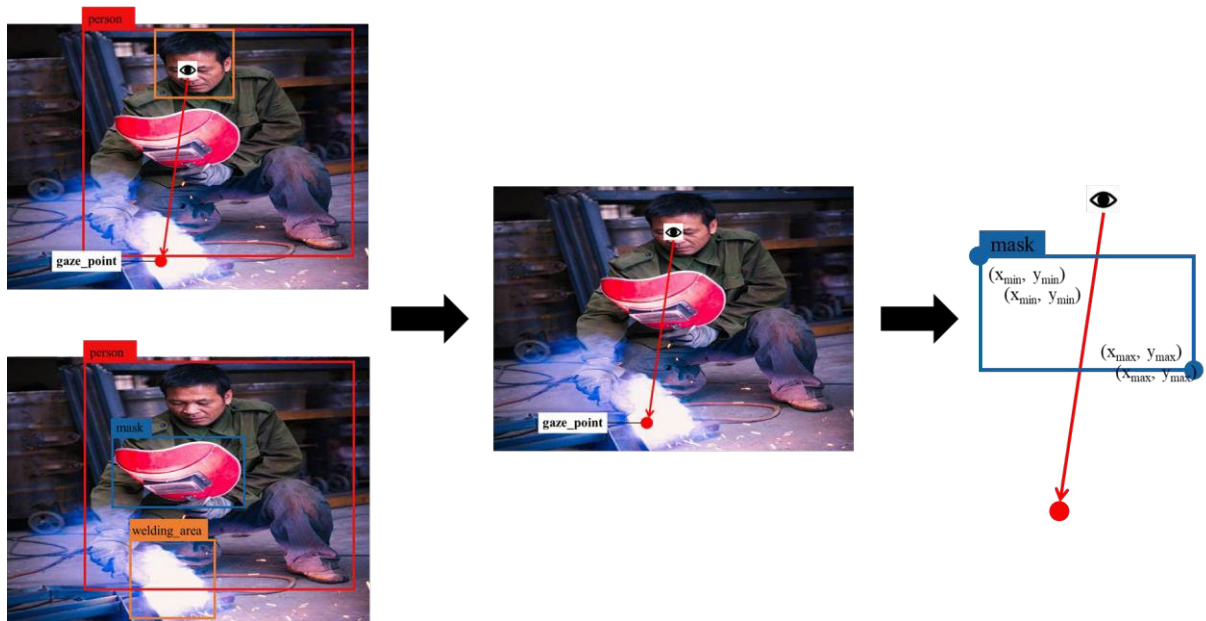


Figure 6. Identification of weld mask

(4) Identification of Glove

The analysis for identifying workers correctly wearing gloves is as follows: The bounding box for the gloves is enclosed within the bounding box for the person. The spatial relationship between these bounding boxes can be expressed by Equations (5.1) and (5.2). See Figure 7 for details.

$$\text{person_x}_{\min} < \text{glove_x}_{\min}, \text{person_y}_{\min} < \text{glove_y}_{\min} \quad (5.1)$$

$$\text{glove_x}_{\max} < \text{person_x}_{\max}, \text{glove_y}_{\max} < \text{person_y}_{\max} \quad (5.2)$$

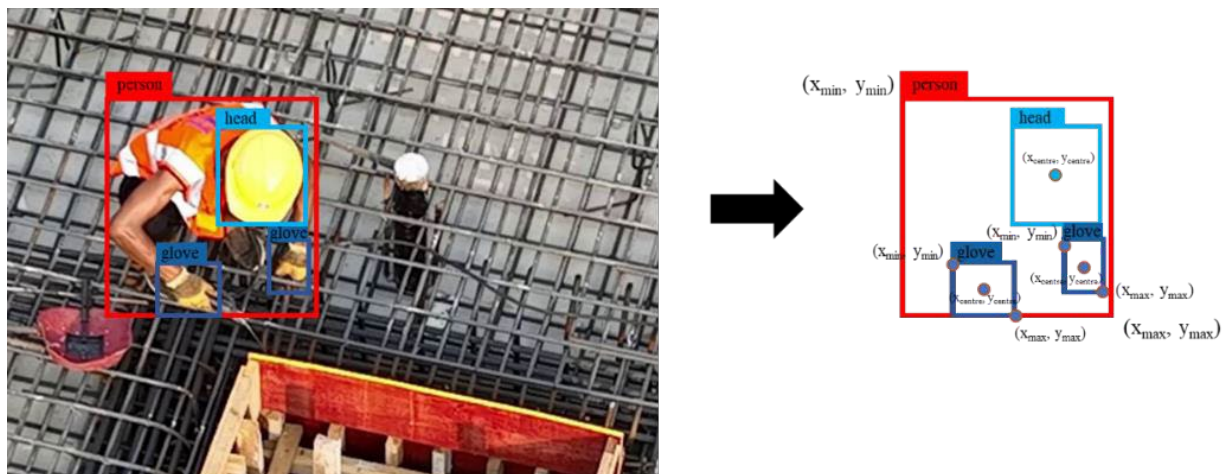


Figure 7. Identification of glove

4. EXPERIMENT

4.1 Data Collection and Model Training

This study utilizes Object-aware Gaze Target Detection (Tonini et al., 2023) for gaze estimation and YOLOv8 for object detection.

Gaze Estimation Dataset: The dataset was collected through web scraping and augmented, resulting in a total of 2304 images. 80% of the images were selected as the training set, while 20% were used as the test set. The results of the model training are shown in Table 2, where P represents precision, R represents recall, and mAP50 represents the mean Average Precision when the IOU (Intersection over Union) is 0.5.

Object Detection Dataset: This dataset was collected through both web scraping and videos taken on construction sites of welding workers performing operations, resulting in a total of 3208 images. 80% of the images were used for training, and 20% for testing. The results of the model training are presented in Table 3, where AUC represents the confidence level of the predicted gaze heatmap relative to the actual gaze points, Distance refers to the distance error between the predicted and actual gaze points, Avg. denotes the average distance error, and Min. distance

represents the minimum distance error.

Table 2. Results of object detection

Class	P ↑	R ↑	mAP50 ↑
all	0.928	0.876	0.907
head	0.97	0.94	0.978
weld_mask	0.919	0.897	0.94
helmet	0.981	0.934	0.972
welding_area	0.827	0.722	0.736
person	0.968	0.977	0.987
safety_vest	0.963	0.972	0.989
glove	0.84	0.578	0.673
extinguisher	0.956	0.988	0.726

Table 3. Results of gaze estimate

AUC ↑	Distance ↓	
	Avg.	Min.
0.866	0.298	0.253

4.2 Control Experiment

Two sets of image datasets were used for the experiment. Group A consists of 20 images taken from actual welding operations on-site. Group B consists of 20 images from on-site scenes where no welding operations were being performed. The experimental results are shown in Table 4, where T represents "True" and F represents "False." The first T or F indicates the ground truth, and the second represents the prediction made by the proposed method. Detailed error samples can be seen in Figure 8.

Table 4. Results of the control experiment

Group	Sample	Is welding	Safety check			
			With helmet	With safety vest	With weld mask	With glove
Group A	001.jpg	T/T	T/T	T/T	F/F	T/F
	002.jpg	T/T	T/T	T/T	F/F	T/T
	003.jpg	T/T	T/T	F/F	T/T	T/F
	004.jpg	T/F	F/F	F/F	T/-	T/-
	005.jpg	T/T	T/T	F/F	T/F	T/T
	006.jpg	T/T	T/T	T/F	T/T	T/T
	007.jpg	T/F	T/T	T/F	T/-	T/-
	008.jpg	T/T	T/T	T/T	F/F	T/T
	009.jpg	T/F	T/T	T/T	F/-	T/-
	0010.jpg	T/T	T/T	T/T	T/T	T/F
	0011.jpg	T/T	T/T	F/F	T/T	F/F
	0012.jpg	T/T	T/T	T/F	F/F	T/T
	0013.jpg	T/T	T/T	F/F	T/T	T/F
	0014.jpg	T/T	T/F	T/F	T/T	T/F
	0015.jpg	T/F	T/T	F/F	T/-	T/-
	0016.jpg	T/T	T/T	T/F	T/T	T/F
	0017.jpg	T/T	T/T	F/F	T/T	T/F

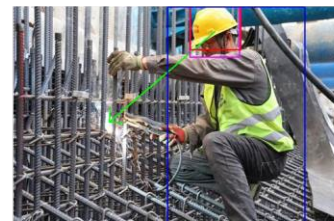
	0018.jpg	T/T	T/T	T/F	T/T	T/F
	0019.jpg	T/F	T/T	T/F	T/-	T/-
	0020.jpg	T/T	T/T	F/F	T/T	T/F
Precision		75%	95%	65%	93%	40%
Group B	0021.jpg	F/F	T/T	T/F	-/-	-/-
	0022.jpg	F/F	T/T	F/F	-/-	-/-
	0023.jpg	F/F	T/T	F/F	-/-	-/-
	0024.jpg	F/F	T/T	F/F	-/-	-/-
	0025.jpg	F/F	T/T	T/F	-/-	-/-
	0026.jpg	F/F	T/T	T/T	-/-	-/-
	0027.jpg	F/F	T/T	T/F	-/-	-/-
	0028.jpg	F/F	T/F	F/F	-/-	-/-
	0029.jpg	F/F	T/T	T/F	-/-	-/-
	0030.jpg	F/F	T/T	T/T	-/-	-/-
	0031.jpg	F/F	T/F	T/T	-/-	-/-
	0032.jpg	F/F	T/T	T/T	-/-	-/-
	0033.jpg	F/F	T/T	T/F	-/-	-/-
	0034.jpg	F/F	T/T	T/T	-/-	-/-
	0035.jpg	F/F	T/T	T/F	-/-	-/-
	0036.jpg	F/F	T/T	T/T	-/-	-/-
	0037.jpg	F/F	T/T	T/T	-/-	-/-
	0038.jpg	F/F	T/T	T/T	-/-	-/-
	0039.jpg	F/F	T/T	T/F	-/-	-/-
	0040.jpg	F/F	T/T	T/F	-/-	-/-
Precision		100%	90%	70%	-	-



(a) gaze estimate error



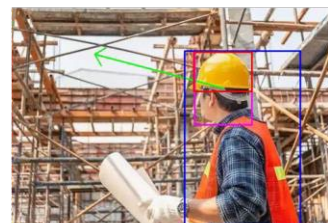
(b) welding area is obscured



(c) failed to identify the welding area



(d) failed to identify people with helmet



(e) failed to identify safety vest

Figure 8. Sample images of misclassified cases

5. DISCUSSION

Through the comparison experiment, it was found that the accuracy of determining welding activities was 75%. (1) The errors were mainly due to the following reasons: 1) As shown in Figure 8(a), the gaze estimation dataset lacked sufficient angles for upward and downward operations. 2) As shown in Figures 8(b) and 8(c), occlusion in the welding area made it difficult for object detection to effectively identify the target. Regarding object detection, the accuracy for helmet recognition was 95% in Group A and 90% in Group B. As shown in Figure 8(d), the primary cause of errors in some specific angles was that the Intersection over Union (IoU) between the head and helmet was less than 0.5. For the safety vest, the accuracy was 65% in Group A and 75% in Group B, significantly lower than the results from YOLOv8 model training. As shown in Figure 8(e), this was mainly because some workers in the experimental group wore orange reflective vests, and the training dataset lacked samples of this color. The accuracy for recognizing the weld mask was 93%, while the accuracy for detecting gloves was only 40%, which was due to poor model training, necessitating further expansion of the dataset.

In summary, the method proposed in this study shows certain potential for practical applications. However, limitations in the dataset and insufficient optimization of the mathematical model have led to some errors. Future improvements should focus on: 1) expanding and diversifying the dataset, particularly by adding samples for different working angles to enhance the precision of recognition of welding activities, and increasing types of PPE to address missed detection of PPE; and 2) further optimizing the mathematical relationship formulas to improve the identification of PPE wearing. These improvements will help enhance the method's effectiveness in real-world construction scenarios, thereby strengthening safety management on construction sites.

6. CONCLUSION

This study proposes a safety management method for welding operations based on gaze estimation and object detection. By combining these two technologies, it aims to address the limitations of traditional object detection methods in understanding scenes in high-risk construction tasks. Experimental validation shows that the method is feasible in identifying welding operations and checking whether workers are wearing personal protective equipment (PPE) as required. The existing errors are attributed to limitations in the dataset and the need for optimization of the mathematical model. Future improvements should focus on enriching the dataset by increasing the diversity of working angles and PPE samples, as well as further optimizing the mathematical relationships to enhance system accuracy and robustness. These improvements will help improve the method's effectiveness in real construction scenarios, thereby strengthening safety management on construction sites.

ACKNOWLEDGMENTS

This study was supported by the National Natural Science Foundation of China [Grant Nos. 52308319].

REFERENCES

- Bermejo, C., Chatzopoulos, D., & Hui, P. (2020). *EyeShopper: Estimating Shoppers' Gaze using CCTV Cameras*. Paper presented at the Proceedings of the 28th ACM International Conference on Multimedia, Seattle, WA, USA. <https://doi.org/10.1145/3394171.3413683>
- Chen, W., Li, C., & Guo, H. (2023). A lightweight face-assisted object detection model for welding helmet use, *Expert Systems with Applications*, 221, 119764
- Fan, L. F., Chen, Y. X., Wei, P., Wang, W. G., Zhu, S. C., & Ieee. (2018). *Inferring Shared Attention in Social Scene Videos*. Paper presented at the 2018 IEEE/CVF CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR).
- Fan, X., Wang, F., Song, D., Lu, Y., & Liu, J. (2021). GazMon: Eye Gazing Enabled Driving Behavior Monitoring and Prediction, *IEEE Transactions on Mobile Computing*, 204, 1420-1433
- Huang, L., Fu, Q. B., He, M. L., Jiang, D., & Hao, Z. Q. (2021). Detection algorithm of safety helmet wearing based on deep learning, *Concurrency and Computation-Practice & Experience*, 3313,
- Jiaqi, L., Xuefeng, Z., Guangyi, Z., & Mingyuan, Z. (2022). Standardized use inspection of workers' personal protective equipment based on deep learning, *Safety Science*, 150, 105689
- Katsini, C., Abdrabou, Y., Raptis, G. E., Khamis, M., Alt, F., & Acm. (2020). *The Role of Eye Gaze in Security and Privacy Applications: Survey and Future HCI Research Directions*. Paper presented at the PROCEEDINGS OF THE 2020 CHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS (CHI'20).
- Lieyun, D., Weili, F., Hanbin, L., Peter, E. D. L., Botao, Z., & Xi, O. (2018). A deep hybrid learning model to detect unsafe behavior: Integrating convolution neural networks and long short-term memory, *Automation in Construction*, 86, 118-124
- Numan, K., Muhammad Rakeh, S., Doyeop, L., Man-Woo, P., & Chansik, P. (2021). Utilizing safety rule

- correlation for mobile scaffolds monitoring leveraging deep convolution neural networks, *Computers in Industry*, 129, 103448
- Rahman, Y., Asish, S. M., Khokhar, A., Kulshreshth, A. K., & Borst, C. W. (2019). Gaze Data Visualizations for Educational VR Applications, *Sui '19*,
- Tafasca, S., Gupta, A., & Odobez, J.-M. (Year). ChildPlay: A New Benchmark for Understanding Children's Gaze Behaviour, *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 20935-20946
- Tom Ka Man, W., Siu Shing, M., & Alan Hoi Shou, C. (2020). Critical factors for the use or non-use of personal protective equipment amongst construction workers, *Safety Science*, 126, 104663
- Tonini, F., Dall'Asen, N., Beyan, C., Ricci, E., & Ieee. (2023). *Object-aware Gaze Target Detection*. Paper presented at the 2023 IEEE/CVF INTERNATIONAL CONFERENCE ON COMPUTER VISION (ICCV 2023).
- Weili, F., Botao, Z., Neng, Z., Peter, E. D. L., Hanbin, L., Jiayue, X., & Shuangjie, X. (2019). A deep learning-based approach for mitigating falls from height with computer vision: Convolutional neural network, *Advanced Engineering Informatics*, 39, 170-177
- Weili, F., Lieyun, D., Hanbin, L., & Peter, E. D. L. (2018). Falls from heights: A computer vision-based approach for safety harness detection, *Automation in Construction*, 91, 53-61
- Weili, F., Lieyun, D., Peter, E. D. L., Hanbin, L., Heng, L., Feniosky, P.-M., . . . Cheng, Z. (2020). Computer vision applications in construction safety assurance, *Automation in Construction*, 110, 103013

APPROCHING AN INTERACTIVE RESOLUTION ON NEW CITIZENS' JOB-HOUSING BALANCE AND VACANT HOUSING REDUCTION IN GUANGZHOU

Ziyao Dong¹, Yangbing Zhang², Yuejiang Su³, Yingqiao Fang⁴, Yilin Huang⁵ and Siyang Zheng⁶

1) Postgraduate student, School of Civil Engineering and Transportation, South China University of Technology, Guangzhou, China. Email: 202420107868@mail.scut.edu.cn

2) Corresponding Author, Associate Professor, School of Civil Engineering and Transportation, South China University of Technology, Guangzhou, China. Email: zhangybct@scut.edu.cn

3) Professor Level Senior Engineer, Ph.D. Candidate, School of Civil Engineering and Transportation, South China University of Technology, Guangzhou, China. Email: 250234329@qq.com

4) Research Student, School of Architecture, South China University of Technology, Guangzhou, China. Email: 2757184877@qq.com

5) Research Student, School of Civil Engineering and Transportation, South China University of Technology, Guangzhou, China. Email: 1020940917@qq.com

6) Research Student, School of Civil Engineering and Transportation, South China University of Technology, Guangzhou, China. Email: wayne_christian@163.com

Abstract: This study examines the job-housing balance in Guangzhou, focusing on its implications for new residents. Initially, it calculates the self-sufficiency of employment and residential populations across 11 administrative districts and 170 streets/towns in Guangzhou, along with peak commuting times for weekdays, to evaluate the job-housing relationship of new residents. Subsequently, the study assesses the housing vacancy rate in Guangzhou and analyzes the current state of vacant properties. Finally, by integrating the job-housing dynamics of new residents with the housing vacancy situation, the study proposes strategies aimed at addressing the challenges of vacant housing and job-housing imbalance through the rational allocation of unoccupied properties.

Keywords: New Citizens, Job-Housing Balance, Housing Supply, Vacant Housing.

1. INTRODUCTION

In recent years, the rapid urbanization and economic development of numerous major cities have led to a significant influx of new residents, resulting in a substantial increase in housing demand (Zeng et al., 2021). Although there are available housing options for most new residents, the pressures of daily life often compel them to select residential areas that are distant from their workplaces (Blumenberg & King, 2024). This situation typically results in longer commuting times, which not only diminishes the overall quality of life for these individuals but also places considerable strain on urban transportation systems (Huang et al., 2022). This issue is particularly pronounced in Guangzhou (Zheng et al., 2021). As a first-tier city with promising prospects for future development, Guangzhou has attracted many new residents from various provinces in search of employment (P. Wang et al., 2024). Confronted with high housing costs, these new residents frequently choose to reside in suburban areas far from their workplaces, thereby exacerbating the imbalance between job locations and housing availability (Song et al., 2019). This imbalance not only contributes to congestion during peak commuting hours but also poses significant challenges to the sustainable development of the city, creating substantial obstacles for urban planners and policymakers (Song et al., 2019).

The concept of job-housing balance is critical in urban planning, as it directly impacts the socio-economic dynamics of a city (Huang et al., 2021). This concept has its roots in Western urban design, dating back to the 19th century with Howard's "Garden City" model, which advocated that "residents' employment areas should be within walking distance" to mitigate issues such as urban overcrowding and traffic congestion (Howard, 1902). Subsequently, both domestic and international research has converged on a relatively unified definition: within a well-defined urban area, the number of housing units should be approximately aligned with the number of job opportunities (Cervero, 1989; Giuliano, 1992). This alignment allows residents to work nearby, minimizing travel distances and time costs (Zhao et al., 2011). A favorable job-housing ratio can effectively reduce traffic congestion, decrease environmental pollution, and enhance overall urban efficiency (Xiao et al., 2023). In summary, achieving a job-housing balance is not only a matter of urban planning but also a crucial factor in fostering sustainable urban development and improving the quality of life for residents in rapidly urbanizing cities like Guangzhou (Xiao et al., 2023).

The term "new citizens" originally referred to migrant workers who have resided in cities for an extended period and have relatively stable employment (Mu et al., 2024). Over time, the definition of this term has gradually expanded to include various groups who are not originally from the local area but have relocated to a city for diverse reasons, thereby overlapping with the concept of the floating population (Liao & Zhang, 2021). Currently, individuals who have transitioned to urban areas through employment, education, or other means and have established stable residency in these areas can collectively be referred to as new citizens (Chen et al., 2022).

For these new citizens, economic pressures and limited housing options are common challenges, making the achievement of a job-housing balance essential for their successful integration into urban life (Blumenberg & King, 2021).

Currently, the majority of scholarly research concentrates on proposing housing supply strategies, such as increasing the availability of affordable housing (P. H. Wang et al., 2024), while the issue of job-housing balance receives comparatively little attention. Furthermore, most new residents are employed in the city center, where land resources are limited and there is a shortage of space for new housing construction. This situation, combined with high rental prices in the area, exacerbates the imbalance between job locations and housing availability. Nevertheless, there remain some vacant housing units in the city center. Effectively utilizing these vacant properties could not only address the issue of housing vacancy but also enhance the job-housing relationship and mitigate the imbalance between jobs and housing.

Therefore, this study focuses on Guangzhou as the research area. Initially, it assesses the self-sufficiency of employment and residential populations across 11 administrative districts and 170 streets/towns in Guangzhou, along with the commuting times during peak hours on weekdays, to evaluate the job-housing relationship of new citizens in the city. Subsequently, the study measures the housing vacancy rate in Guangzhou and analyzes the current status of vacant properties. Finally, by integrating the job-housing relationship of new citizens with the housing vacancy situation in Guangzhou, this research proposes targeted strategies aimed at addressing both the challenges of vacant housing and the imbalance between employment and housing through the rational allocation of these vacant properties.

2. METHOD

2.1 Data Description

This study focuses on Guangzhou, which comprises 11 administrative districts and 170 streets and towns. According to statistics, by the end of 2019, Guangzhou's administrative area spanned 7,434 square kilometers, with a permanent resident population of 15.3 million (data sourced from the 2020 Guangzhou Statistical Yearbook). As the capital of Guangdong Province, Guangzhou has consistently demonstrated high-quality economic development, positioning itself as a crucial driver of growth for the province. Furthermore, its advantageous development conditions have generated a substantial number of employment opportunities.

Mobile signaling data, recognized as a significant form of big data, has attracted increasing attention from researchers due to rapid advancements in mobile communication technology and the enhanced availability of data from telecommunications operators in recent years. Additionally, mobile signaling data can be utilized to reflect users' travel trajectories, which can subsequently be employed to identify users' residential and workplace locations. This will aid in measuring the job-housing balance indicators. The mobile signaling data employed in this study was provided by China Telecom and encompasses the period from 00:00 on April 16, 2019, to 24:00 on May 1, 2019, specifically focusing on the non-Guangzhou resident population within the city.

Remote sensing data refers to information about the Earth's surface and its features that is collected from a distance, typically using sensors mounted on satellites or aircraft. This method enables the acquisition of data without direct contact with the ground and is widely applied across various fields, including environmental monitoring, land use planning, climate research, disaster management, and urban development. Furthermore, remote sensing data can be effectively utilized to measure housing vacancy rates in different regions. Consequently, this study will employ cloud-free composite data from April 2019, provided by the National Oceanic and Atmospheric Administration (NOAA), to assess housing vacancy rates across various administrative districts and streets/towns in Guangzhou. As a next-generation nighttime light dataset, NPP-VIIRS provides vcm data that effectively mitigates the influence of stray light, ensuring excellent spatial continuity, high resolution (500m×500m), and strong timeliness, with updates occurring on a monthly basis.

2.2 Data Processing

(1) Mobile Signal Data

The study commences with an analysis of mobile signaling data, which encompasses user codes, the city and district of number attribution, the longitude and latitude of recorded base stations, as well as the start and end times of signaling records and their respective durations. A total of 172,699,047 signaling records were analyzed, involving 7,253,819 users and 27,241 base stations. The original dataset was filtered to exclude signaling records with durations of less than 30 minutes, resulting in a refined dataset characterized by longer signaling durations and a reduced number of records.

The distribution of base stations in the central urban area generally adheres to a regular grid pattern of 500m×500m, while in suburban areas, the spacing between base stations ranges from 1 to 2 km. By examining the travel trajectories reflected in the mobile signaling data and integrating the cumulative duration of user stops at various locations with established travel patterns, we successfully identified the residential and workplace locations of 883,995 users.

After acquiring the mobile signaling records and base station information of users, criteria for

identifying workplaces and residences were established. This study considers users' travel patterns, the extended duration of signaling records, and the limited number of records available. The methods based on the shortest distance and access frequency exhibited relatively low success rates in accurately identifying workplaces and residences. Consequently, the method utilizing cumulative stay time at visited base stations was selected for identification, implemented through programming in Python.

To enhance the efficiency of the identification process, the workplace was identified first. If the actual workplace could not be determined, the identification of the residence was not pursued further. Working hours were defined as 08:00 to 18:00, and the base station where the user accumulated the longest stay time during this period on a given workday was considered the potential workplace for that day. Following the identification of potential workplaces for each day, the actual workplace was defined as the location that appeared most frequently, with a minimum occurrence of three days.

Similarly, the residence identification period was established from 20:00 in the evening to 06:00 the following day. After identifying potential residences based on cumulative stay duration for each day, the actual residence was defined as the location that appeared most frequently, with a minimum occurrence of five days. Additionally, this study does not account for scenarios in which individuals work from home. Therefore, cases where the workplace and residence are the same location were excluded from the analysis.

Further statistical analysis was performed to evaluate various indicators of job-housing balance, including the number of residents, the number of workers, and the self-sufficiency ratio (local employment/resident population) for each street or town. Using the Baidu Maps route planning API, we calculated the commuting times for the aforementioned users during peak hours on weekdays, considering both private car and public transportation modes.

(2) Remote Sensing Data

The monthly DNB data provided by the platform categorizes global regions into six segments and offers two products: vcm and vcmssl, with the former effectively eliminating the influence of stray light. In this study, we utilize the vcm product for Tile 3, which is subsequently imported into Arc Map and clipped to the geographical boundaries of Guangzhou.

The primary steps for calculating the housing vacancy rate in this research are outlined as follows: First, outliers and negative values are systematically excluded from the remote sensing data. Next, the urban area is delineated based on the relationship between brightness thresholds and the corresponding changes in perimeter. Following this, the top 20% of brightness values from the statistical curve of grid brightness values are designated as the housing occupancy threshold. The housing vacancy rates for the remaining grids are then calculated using linear interpolation. Finally, the average housing vacancy rate for the grids within each street, town, and administrative district is computed.

3. RESULTS

Table 1 presents the employment figures, resident counts, local employment numbers, and average self-sufficiency ratios for each district. Table 2 identifies the two streets/towns within each district that exhibit the highest and lowest self-sufficiency ratios, along with their corresponding employment numbers, resident counts, and local employment figures. Table 3 displays the average commuting times for users in each district, including average driving times, average public transportation times, and overall average commuting times. Additionally, Table 4 highlights the two streets/towns within each district that have the longest and shortest average commuting times during peak hours, along with their respective driving and public transportation durations. Finally, Table 5 presents the average housing vacancy rates for each district, as well as the two streets/towns with the highest and lowest vacancy rates within each district. This comprehensive analysis provides valuable insights into the dynamics of job-housing balance and commuting patterns, contributing to a better understanding of urban mobility and residential stability.

Table 1. Employment and Resident Population, and Self-Sufficiency by District

Administrative District	Employment (Persons)	Resident Population (Persons)	Local Employment (Persons)	Average Self-Sufficiency (%)
Tianhe	162,598	118,908	5,771	4.85
Baiyun	126,120	159,554	29,326	18.38
Liwan	38,256	51,323	1,807	3.52
Haizhu	89,333	99,096	5,028	5.07
Yuexiu	87,661	46,617	700	1.50
Panyu	123,657	146,590	33,534	22.88
Huangpu	72,456	66,922	13,486	20.15
Huadu	61,848	65,083	24,223	37.22
Zengcheng	55,222	64,808	27,430	42.33

Nansha	46,048	44,536	22,388	50.27
Conghua	20,796	20,558	8,866	43.13

Table 2. Employment and Resident Population, and Self-Sufficiency by Selected Streets/Towns

Administrative District	Street/Town	Employment (Persons)	Resident Population (Persons)	Local Employment (Persons)	Self-Sufficiency (%)
Tianhe	Xintang Street	6,075	4,427	726	0.16
	Shadong Street	759	1,258	15	1.19
Baiyun	Zhongluotan Town	10,120	10,186	5,486	53.86
	Tangjing Street	3,626	5,923	137	2.31
Liwan	Hailong Street	2,119	3,520	495	14.06
	Shamian Street	411	247	0	0.00
Haizhu	Pazhou Street	16,066	4,994	1,127	22.57
	Sushesha Street	1,807	2,666	9	0.34
Yuexiu	Kuanquan Street	2,258	2,572	74	2.88
	Dadong Street	3,086	1,711	2	0.12
Panyu	Hualong Town	5,348	3,915	1,955	49.94
	Shiqiao Street	9,426	12,417	1,084	8.73
Huangpu	Jiulong Town	5,536	5,460	2,727	49.95
	Huangpu Street	3,859	5,352	179	3.35
Huadu	Shiling Town	8,932	8,567	5,086	59.37
	Timian Town	410	397	39	9.82
Zengcheng	Shitan Town	4,865	5,376	3,090	57.48
	Yongning Street	10,362	14,714	5,037	34.23
Nansha t	Longxue Street	3,764	750	593	79.07
	Zhujiang Street	1,849	1,696	504	29.72
Conghua	Lvtian Town	284	247	232	93.93
	Jiekou Street	1,880	2,976	394	13.24

Table 3. Average Commute Time (in minutes) during Peak Hours by District

Administrative District	Average Driving Time (min)	Average Public Transport Time (min)	Average Commute Time (min)
Tianhe	22.85	59.46	41.15
Baiyun	23.00	67.36	45.18
Liwan	23.00	58.25	40.63
Haizhu	22.09	61.28	41.69
Yuexiu	24.64	51.58	38.11
Panyu	23.14	67.73	45.43
Huangpu	23.15	69.79	46.47
Huadu	20.98	74.32	47.65
Zengcheng	22.99	84.21	53.60
Nansha	21.14	77.28	49.21
Conghua	23.93	89.98	56.96

Table 4. Commute Time (in minutes) during Peak Hours by Selected Streets/Towns

Administrative District	Street/Town	Driving Time (min)	Public Transport Time (min)	Average Commute Time (min)
Tianhe	Qianjin Street	23.8	113.4	68.6
	Xinghua Street	21.3	51.6	36.45
Baiyun	Taihe Town	25.9	81.8	53.85
	Tongde Street	20.1	57.3	38.7
Liwan	Duobao Street	25.2	70.5	47.85
	Zhanqian Street	22.7	47.8	35.25
Haizhu	Shayuan Street	21.8	109.2	65.5
	Haizhuang Street	20.6	50.2	35.4
Yuexiu	Liurong Street	38.6	72.5	55.55
	Renmin Street	22.4	48.1	35.25
Panyu	Shilou Town	26.5	74.9	50.7
	Hualong Town	16	63	39.5

Huangpu	Huangpu Street	25.5	114.7	70.1
	Dasha Street	19.7	56.2	37.95
Huadu	Timian Town	28.9	94.3	61.6
	Huacheng Street	18.1	61.5	39.8
Zengcheng	Zhengguo Town	34.7	148.7	91.7
	Licheng Street	18.6	62.7	40.65
Nansha	Longxue Street	18.6	98.3	58.45
	Lanhua Town	19.7	64.5	42.1
Conghua	Liangkou Town	47.6	156.9	102.25
	Taiping Town	18.4	65.4	41.9

Table 5. Housing Vacancy Rates (%) by District and Selected Streets/Towns

Administrative District and Average Vacancy Rate (%)	Street	Vacancy Rate (%)
Tianhe (15.46)	Shahe Street	33.90
	Xian Village Street	1.65
Baiyun (16.26)	Jingxi Street	30.54
	Tangjing Street	1.54
Liwan (20.08)	Qiaozhong Street	35.10
	Zhanqian Street	1.97
Haizhu (17.45)	Nanzhou Street	31.50
	Shayuan Street	7.24
Yuexiu (16.70)	Guangta Street	32.55
	Liuhua Street	3.10
Panyu (23.42)	Xinzhaoh Town	32.80
	Shibi Street	14.22
Huangpu (17.78)	Changzhou Street	30.44
	Huangpu Street	3.40
Huadu (12.64)	Xiuquan Street	29.88
	Huadong Town	5.34
Zengchengt (22.04)	Zengjiang Street	33.22
	Yongning Street	18.03
Nansha (15.16)	Wanquingsha Town	33.04
	Longxue Street	8.70
Conghua (18.88)	Taiping Town	34.07
	Jiekou Street	13.53

4. DISCUSSION

According to Table 1, Nansha District has the highest average self-sufficiency in employment and resident population among the administrative districts in Guangzhou, while Yuexiu District has the lowest. This disparity is primarily influenced by a combination of factors, including economic development levels, industrial structure, and living costs. Nansha District, designated as a new economic zone in Guangzhou, has undergone rapid development in recent years, attracting a substantial influx of enterprises and investments. The flourishing manufacturing and high-tech industries in this district have generated numerous job opportunities. Furthermore, the relatively low living costs in Nansha have drawn a significant number of migrant workers, contributing to a robust local employment population. In contrast, Yuexiu District, as the central urban area of Guangzhou, is economically advanced but faces constraints due to limited land resources. Its industrial structure is predominantly oriented towards the service and commercial sectors, which results in relatively fewer job opportunities. Consequently, residents in Yuexiu are more likely to seek employment in other districts, leading to a lower local employment population in Yuexiu District. This analysis highlights the complex interplay of economic and structural factors that influence local employment dynamics in Guangzhou's administrative districts.

Table 2 indicates that Lyutian Town in Conghua District has the highest self-sufficiency in employment and resident population among the towns in Guangzhou, while Shamian Street in Liwan District has the lowest. Lütian Town's economy is primarily supported by agriculture and rural development. In 2021, it was recognized as a provincial-level demonstration town for leisure agriculture and rural tourism for the year 2020. The rural cultural and tourism sectors have generated numerous employment opportunities for local residents. Additionally, the low cost of living and favorable living conditions in rural areas attract individuals to seek both employment and residence in Lütian Town, resulting in an impressive self-sufficiency rate of 93.93%. Conversely, Shamian Street is well-known for its rich historical and cultural heritage, which draws a significant number of visitors.

The local economy is largely centered on the service sector, including restaurants, hotels, and tourism, which primarily provides employment opportunities for migrant workers rather than local residents. Moreover, the relatively limited job opportunities available in Shamian Street lead many individuals to pursue employment elsewhere. Consequently, these factors have resulted in a local employment population of zero in Shamian Street.

According to Tables 3 and 4, Conghua District exhibits the longest average commuting time during peak hours on weekdays in Guangzhou, while Yuexiu District records the shortest. Among the towns, Liangkou Town in Conghua District has the highest average commuting time during these peak hours, in contrast to Zhanqian Street in Liwan District, which has the lowest. Conghua District is situated on the outskirts of Guangzhou, far from the city center, and is characterized by relatively underdeveloped transportation infrastructure and limited public transport options. Consequently, residents in this district tend to spend more time commuting. In contrast, both Yuexiu and Liwan Districts are located in the central area of Guangzhou, benefiting from favorable geographical positions and well-established transportation infrastructure. The public transport systems in these districts, which include multiple subway and bus lines, enable quick access to various destinations, thereby effectively reducing commuting times. This analysis highlights the significant impact of geographical location and transportation infrastructure on commuting patterns within Guangzhou, underscoring the need for targeted improvements in areas with longer commuting times.

Table 5 indicates that Panyu District has the highest average housing vacancy rate in Guangzhou, while Huadu District has the lowest. Among the streets, Qiaozhong Street in Liwan District exhibits the highest vacancy rate, whereas Tangjing Street in Baiyun District has the lowest. Panyu District, located in the southern part of Guangzhou, is relatively distant from the city center and displays uneven internal development. Certain areas have experienced rapid growth, attracting a significant number of residents, while others have lagged behind, resulting in an overall increase in the vacancy rate. Furthermore, in recent years, Panyu District has witnessed the construction of numerous residential projects, particularly large-scale developments, which have substantially increased the supply of new properties and, consequently, the number of vacant homes in the market. In contrast, Huadu District, situated in the northern part of Guangzhou, benefits from convenient transportation links, including its proximity to Guangzhou Baiyun International Airport. The district has experienced rapid economic growth in recent years, attracting a considerable influx of businesses and investments, which has enhanced employment opportunities. Additionally, the ongoing improvement of supporting facilities, such as education, healthcare, and commercial services, has made Huadu District an appealing location for residents, thereby contributing to a reduction in the housing vacancy rate. Qiaozhong Street, located in the central area of Liwan District, features a property market that attracts numerous investors, with some properties being perceived as investment assets rather than personal residences. This trend has led to an increase in the vacancy rate for this street. Conversely, Tangjing Street benefits from significant commercial and industrial development in the surrounding area, resulting in high housing demand, particularly among young families and migrant workers. The well-developed supporting facilities, good transportation accessibility, and an active rental market with moderate rental prices contribute to a notably low housing vacancy rate of just 1.54% in this area.

Based on the preceding analysis, it is clear that a significant imbalance exists between employment and housing across various administrative districts in Guangzhou, with certain districts displaying notably high housing vacancy rates. Effectively leveraging these vacant properties to address the employment-housing imbalance experienced by new residents in Guangzhou could not only mitigate the issue of housing vacancies and reduce resource waste but also enhance the employment-housing relationship for these residents, thereby alleviating the overall imbalance. In this context, the present study integrates employment-housing balance indicators for new residents in Guangzhou with housing vacancy rates to offer targeted recommendations for employment-housing location choices in specific streets and towns within selected administrative districts (as illustrated in Table 6). It is anticipated that these recommendations will play a significant role in addressing the challenges of housing vacancy and the employment-housing imbalance.

Table 6. Recommendations for Employment-Housing Location Matching

District	Employment Location	Residence Location
Haizhu District	Sha Garden Street	Liwan District Hua Di Street, Lingnan Street
	Ruibao Street, Nanshitou Street	Haizhu District Nanzhou Street, Fengyang Street
Tianhe District	Qianjin Street	Tianhe District Huangcun Street
	Longdong Street	Tianhe District Fenghuang Street
Huangpu District	Huangpu Street	Huangpu District Changzhou Street, Dasha Street
	Yuzhu Street	Huangpu District Changzhou Street, Dasha

		Street
	Xiagang Street	Huangpu District Sui Dong Street, Panyu District Hua Long Town
Baiyun District	Shi Jing Street	Baiyun District Shi Men Street, Baiyun Lake Street, Songzhou Street, Tongde Street
Liwan District	Shiweitang Street, Duo Bao Street	Liwan District Qiao Zhong Street, Lingnan Street, Changhua Street, Hua Di Street
Panyu District	Xin Zao Town	Panyu District Xiao Gu Wei Street, Nan Cun Town, Hua Long Town
Nansha District	Longxue Street, Heng Li Town, Wan Qing Sha Town, Huang Ge Town, Nansha Street, Da Gang Town, Zhu Jiang Street, Dong Yong Town	Panyu District Xiao Gu Wei Street, Nan Cun Town, Hua Long Town
Zengcheng District	Pai Tan Town, Xiao Lou Town, Xian Cun Town, Zhu Cun Street, Zhong Xin Town, Shi Tan Town	Panyu District Xiao Gu Wei Street, Nan Cun Town, Hua Long Town
Conghua District	Liangkou Town, Liu Xi He Forest Farm, Wen Quan Town, Jiangpu Street, Ao Tou Town, Lv Tian Town, Cheng Jiao Street	
Huadu District	Chi Ni Town, Ti Mian Town, Hua Dong Town, Shi Ling Town, Tan Bu Town	Choose nearby streets or towns based on local residence or in conjunction with the actual work location and the availability of housing supply in the market.
Baiyun District	Tai He Town, Jiang Gao Town, Baiyun Lake Street, Ren He Town, Shi Men Street, Tong He Street	
Huangpu District	Jiu Long Town, Yong He Street, Sui Dong Street, East District Street	
Panyu District	Shi Lou Town, Hua Long Town	

5. CONCLUSIONS

The sizes of various streets and towns in Guangzhou exhibit significant variation. The central urban districts are characterized by smaller, densely populated areas, while the suburban regions cover larger expanses. Therefore, it is crucial to develop housing supply policies that are specifically tailored to the unique conditions of the central urban area, thereby guiding local workers to reside either within the area or in nearby regions that have higher housing vacancy rates. In contrast, the non-central urban districts—including Conghua District, Huadu District, Zengcheng District, Nansha District, as well as certain streets and towns in Baiyun District, Huangpu District, and Panyu District—are marked by their vast areas and comparatively underdeveloped public transportation systems. This results in longer average commuting times for residents. Consequently, it is advisable to encourage workers to either live locally or choose nearby streets or towns that align with their actual work locations. Based on the preceding analysis of employment and residential self-sufficiency, commuting times, and housing vacancy rates across the streets and towns in Guangzhou's administrative districts, the following three recommendations are proposed.

First, it is crucial to ensure a dedicated supply of affordable rental housing for newly introduced talents and recent graduates. This provision will facilitate their access to housing, encourage them to remain in the area, enable comfortable living conditions, and support their professional development. Furthermore, it is important to develop cost-effective housing solutions specifically designed for frontline builders who play a vital role in urban operations. This strategy not only addresses housing challenges but also aligns with and supports various urban development initiatives.

Second, building on the policies of the national affordable housing program, it is essential to establish a shared ownership housing system to address the housing challenges faced by the "sandwich" group, which exists between low-income and market-rate populations. This initiative aims to support struggling families in achieving their aspiration of homeownership. Under this system, beneficiaries who purchase shared ownership housing will be permitted to list their properties for sale after a five-year period. They will have the option to either buy out the government's share or sell their own share, with the government retaining the right of first refusal for repurchase. During the homeowner's occupancy, the housing rights associated with the government's share will be transferred to the homeowner at no cost. However, the homeowner will be responsible for all maintenance and property management expenses related to the entire property. This approach not only facilitates access to homeownership for disadvantaged groups but also promotes a sustainable housing market by ensuring that properties remain available for future buyers.

Third, a public rental housing system should be established based on a "government-supported,

enterprise-operated" management model, with a strong emphasis on promoting stable employment in Guangzhou. This system will eliminate household registration and income restrictions, adopting a leasing strategy characterized by "limited leasing, rental only without sale, and slightly below market rates." Furthermore, a standard will be set to ensure that the total lease term does not exceed six years. However, beneficiaries will have the option to exit the program "flexibly" after this six-year period, allowing them to continue renting at market rates within public rental housing for an additional one to two years as a transitional measure. This approach upholds a "people-centered" philosophy while maintaining the essential characteristics and functions of the public rental housing system in addressing temporary housing challenges.

ACKNOWLEDGMENTS

This work was funded by the Fundamental Research Funds for the Central Universities, China under Grant [2023ZYGXZR075] and the Guangdong Basic and Applied Basic Research Foundation, China under Grant [2023A1515011006].

REFERENCES

- Blumenberg, E.;King, H. (2021). Jobs-Housing Balance Re-Re-Visited. *Journal of the american planning association*, 87(4), 484-496.
- Blumenberg, E.;King, H. (2024). Young workers, jobs-housing balance, and commute distance: Findings from two high-housing-cost US regions. *Cities*, 147.
- Cervero, R. (1989). Jobs-housing balancing and regional mobility. *Journal of the american planning association*, 55(2), 136-150.
- Chen, L.;Du, H.;Hui, E. C.-m.;Tan, J.;Zhou, Y. (2022). Why do skilled migrants' housing tenure outcomes and tenure aspirations vary among different family lifecycle stages? *Habitat International*, 123, 102553.
- Giuliano, G. (1992). Is Jobs-Housing Balance a Transportation Issue? *Transportation Research Record Journal of the Transportation Research Board*, 1305(1305), 305-312.
- Howard, E. (1902). Garden cities of tomorrow. *Swan Sonnenschein & Co., Ltd., London*.
- Huang, J.;Hu, Y. J.;Wang, J. E.;Li, X. (2021). A tale of two cities: Jobs-housing balance and urban spatial structures from the perspective of transit commuters. *Environment and Planning B-Urban Analytics and City Science*, 48(6), 1543-1557.
- Huang, Y. D.;Du, Q.;Li, Y.;Li, J. T.;Huang, N. (2022). Effects of Metro Transit on the Job-Housing Balance in Xi'an, China. *Journal of Urban Planning and Development*, 148(1).
- Liao, Y.;Zhang, J. (2021). Hukou status, housing tenure choice and wealth accumulation in urban China. *China Economic Review*, 68, 101638.
- Mu, X. Y.;Li, H. Q.;Cui, C.;Song, W. X. (2024). Housing career disparities in a transitional society: Comparing local residents, new citizens, and the floating population in Nanjing. *Cities*, 155.
- Song, C.;Jin, A.;Ma, X. Y.;Jee. (2019, Jul 14-17). Job-Housing Spatial Relationship Evolution in Guangzhou Analysis, Reflection and Coping Strategies. [2019 5th international conference on transportation information and safety (ictis 2019)]. 5th International Conference on Transportation Information and Safety (ICTIS), Liverpool, ENGLAND.
- Wang, P.;Li, K.;Zhang, W. (2024). China's New Housing Security Model: Evaluation of the Job-Housing Balance in Affordable Rental Housing, Shanghai. *Land*, 13(7).
- Wang, P. H.;Li, K. R.;Zhang, W. B. (2024). China's New Housing Security Model: Evaluation of the Job-Housing Balance in Affordable Rental Housing, Shanghai. *Land*, 13(7).
- Xiao, W. Y.;Wei, Y. D.;Chen, W. (2023). Skills mismatch, jobs-housing relationship and urban commuting. *Travel Behaviour and Society*, 33.
- Zeng, H.;Ke, Q.;Yu, X. (2021). Investigating the factors affecting the residential satisfaction of new-generation migrants: a case study of Hangzhou in China. *International Journal of Urban Sciences*, 25(1), 16-30.
- Zhao, P.;Lü, B.;De Roo, G. (2011). Impact of the jobs-housing balance on urban commuting in Beijing in the transformation era. *Journal of Transport Geography*, 19(1), 59-69.
- Zheng, Z.;Zhou, S. H.;Deng, X. D. (2021). Exploring both home-based and work-based jobs-housing balance by distance decay effect. *Journal of Transport Geography*, 93.



Part 3

**Innovations for Operation
and Maintenance**

LIGHTWEIGHT TRANSFER LEARNING FOR WATER BODY SEGMENTATION USING ADAPTOR-BASED FINE-TUNING

Jiapei Zhao¹, Nobuyoshi Yabuki², Tomohiro Fukuda³

1) Ph.D. Student, Division of Sustainable Energy and Environmental Engineering, Osaka University, Osaka, Japan. Email: u812893b@ecs.osaka-u.ac.jp

2) Ph.D., Prof., Division of Sustainable Energy and Environmental Engineering, Osaka University, Osaka, Japan. Email: yabuki@see.eng.osaka-u.ac.jp

3) Ph.D., Assoc. Prof., Division of Sustainable Energy and Environmental Engineering, Osaka University, Osaka, Japan. Email: fukuda.tomohiro.see.eng@osaka-u.ac.jp

Abstract: This paper presents a lightweight transfer learning approach for water body segmentation by applying Adaptor-based fine-tuning on general image datasets. Traditional deep learning models often require full-scale retraining for each new task, which is computationally expensive and time-consuming. In contrast, Adaptor networks—lightweight modules that selectively fine-tune task-specific layers while retaining most pre-trained model parameters—offer an efficient alternative. Water bodies present unique challenges for segmentation, such as varying lighting, reflections, and seasonal fluctuations. These factors can confuse distinguishing water from land, particularly in cases where reflections resemble adjacent features. Adaptor-based fine-tuning helps to reduce computational costs while ensuring the model captures the fine distinctions between similar regions like shallow water and land. This paper evaluated the method on the ATLANTIS dataset, which includes diverse categories of water bodies such as lakes, rivers, and wetlands. This dataset is recognized as a comprehensive collection for evaluating semantic segmentation performance in varied environmental conditions. The results indicate that Adaptor-based fine-tuning achieves comparable performance to fully fine-tuned models, with a significant reduction in computational costs and training time. The method also demonstrated high precision in segmenting water bodies under challenging conditions, such as occlusions and reflections. This study highlights the potential of lightweight transfer learning in resource-constrained environments, with applications in environmental monitoring, hydrological modeling, and geographic information systems (GIS). By demonstrating the effectiveness of Adaptor networks, this work contributes to the broader field of efficient transfer learning, showcasing how minimal adjustments to pre-trained models can yield accurate task-specific performance.

Keywords: Adaptor Networks, Water Body Segmentation, Lightweight Transfer Learning, Fine-Tuning.

1. INTRODUCTION

Water body segmentation is a fundamental task in environmental monitoring, particularly for applications in geographic information systems (GIS), hydrological modeling, and urban planning (Kadhim & Premaratne, 2023). The increasing frequency of extreme weather events due to climate change makes accurate segmentation of water bodies crucial for disaster management, urban flood prediction, and resource allocation (Zaffaroni & Rossi, 2020). Satellite imagery provides an efficient means for this purpose, but the complexity of natural environments, including reflections, occlusions, and seasonal variations, complicated the accurate delineation of water boundaries (Saleh et al., 2018).

Traditional deep learning approaches, such as convolutional neural networks (CNNs), have been widely employed for semantic segmentation (Pinaya et al., 2020; Yuan et al., 2021). However, these methods often require a complete retraining process for each new segmentation task, making them computationally expensive and unsuitable for scenarios with limited computational resources (Chen et al., 2017; Zhao et al., 2017). Moreover, the need for vast labeled datasets and the high computational cost associated with full-scale retraining limit the scalability of these models to new environments and tasks. In contrast, recent advancements leverage Vision Transformers (ViTs), which have shown great promise for segmentation tasks due to their ability to model long-range dependencies efficiently (Dosovitskiy, 2020).

To address these challenges, we were the first to apply Adaptor-based fine-tuning with Vision Transformer (ViT) architecture to the task of waterbody segmentation, proposing a lightweight transfer learning approach. This approach leverages Vision Transformers (ViTs), specifically the SegFormer architecture (Xie et al., 2021), which has shown promise in capturing multi-scale features suitable for segmentation (Xie et al., 2021). By selectively fine-tuning task-specific layers using lightweight Adaptor modules (Liu et al., 2023), this method retains most pre-trained model parameters while adapting efficiently to the unique characteristics of water bodies, such as their dynamic boundaries and varying appearances. This strategy effectively reduces the computational load, facilitating the deployment of the model in resource-constrained environments (Dong et al., 2023). The experimental evaluation presented in this paper highlights the efficacy of our method when applied to 15 different water body categories, demonstrating its robustness and efficiency in real-world scenarios.

2. RELATED WORK

Recent advancements in semantic segmentation have seen a shift from conventional convolutional neural networks (CNNs) to transformer-based architectures, such as Vision Transformers (ViTs) (Dosovitskiy, 2020). These models have shown notable improvements in capturing global context and multi-scale information, which are essential for accurately delineating objects. The SegFormer model, in particular, has gained prominence for its hierarchical transformer architecture, which is effective for extracting multi-scale features suitable for segmentation tasks (Xie et al., 2021).

Adaptor-based fine-tuning methods have also been explored in natural language processing (NLP) applications, such as BERT (Kenton & Toutanova, 2019) and T5 (Raffel et al., 2020), to reduce computational costs while retaining model performance. This technique has gradually been introduced to the computer vision domain to address similar challenges (Dong et al., 2023). This approach draws upon these advancements, extending the idea of adaptor modules to the segmentation of water bodies—a task characterized by diverse and challenging environmental conditions, including fluctuating water levels, occlusions, and reflections.

Prior studies have utilized datasets such as COCO for semantic segmentation of multiple categories. For instance, the ATLANTIS dataset, proposed by Erfani et al., (2022), contains 5,195 training images and 1,296 testing images, with a wide variety of water body types. This study focuses on 15 specific categories from ATLANTIS to validate the method. Compared to state-of-the-art models like DNLNet, GCNet, and AQUANet (Cao et al., 2019; Erfani et al., 2022; Ni et al., 2022), this adaptor-based method offers improved performance by selectively fine-tuning high-frequency components and embedding features critical to water body segmentation.

3. METHOD

In this section, a lightweight transfer learning framework for water body segmentation by applying Adaptor-based fine-tuning is introduced. The framework utilizes Vision Transformers pre-trained on large-scale image datasets such as ImageNet (Deng et al., 2009). This method focuses on selectively fine-tuning layers critical to task-specific segmentation, which allows to avoid the computational cost of full-scale retraining. By targeting high-frequency components and image embeddings, the model effectively adapts to challenges such as varying lighting conditions, occlusions, and reflections in water body segmentation.

Figure 1 presents an overview of the Adaptor-based fine-tuning approach. The frozen pretrained Transformer model, represented with an ice symbol, indicates the components that remain unchanged during the training process. The Adaptor module, marked with a flame symbol, highlights the elements that are specifically fine-tuned to adapt to the target -water body segmentation task. This Adaptor module leverages the generalization capabilities of the pretrained model while enabling efficient adaptation to the specific requirements of water body segmentation. The input image undergoes processing through the model, producing a segmented output that accurately delineates the water body.

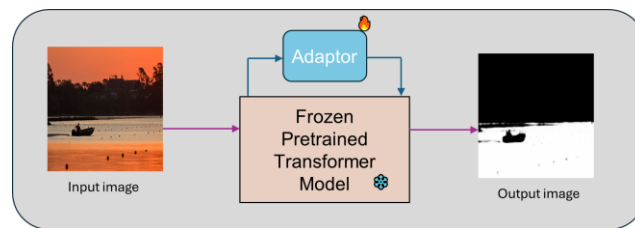


Figure 1. Overview of Adaptor-Based Fine-Tuning for Water Body Segmentation

3.1 Pre-trained Model Architecture

This approach builds upon the hierarchical structure of SegFormer (Xie et al., 2021), a Vision Transformer (ViT) model renowned for its multi-scale feature extraction. The encoder of SegFormer is composed of a series of transformer blocks that progressively capture finer image details. Each block outputs features at different resolutions, providing multi-level representations.

As Figure 1 shows, given that the pre-trained model already captures essential structural features, its backbone layers are frozen to retain general visual knowledge learned from ImageNet. Only a small number of task-specific layers are leveraged for fine-tuning. These layers learn to specialize in water body segmentation without the need to modify the majority of the pre-trained parameters, drastically reducing computational overhead. The frozen layers are shown in Equation (1):

$$\theta_f \in \mathbb{R}^d \quad (1)$$

where θ_f represents the frozen parameter set, the symbol \mathbb{R} represents the set of real numbers, and d is the number of parameters in the backbone.

3.2 High-frequency Component Extraction

One of the core aspects of the Adaptor model is the extraction and fine-tuning of high-frequency components (HFC) (Wang et al., 2020), which capture fine structural details crucial for water body segmentation. These high-frequency components are extracted using the Fourier Transform. The process decomposes the input image into low- and high-frequency components, where the high-frequency components focus on sharp boundaries and textures, essential for delineating water from land.

$$f(\tau) = \begin{cases} 1, & \text{if } \frac{4(\frac{H}{2}-x)(\frac{W}{2}-y)}{HW} \leq \tau \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

As Equation (2) shows, H and W are the height and width of the image, x and y are the coordinates of a given pixel, and τ controls the frequency threshold. After applying this mask, the high-frequency component is obtained using the inverse Fourier transform using Equation (3). The high-frequency components I_{hf} are computed by applying a binary mask M_{hf} , which retains frequencies above a threshold τ . fft is the Fourier transform and $ifft$ is its inverse, and I represents the input image.

$$I_{hf} = ifft(M_{hf}) \cdot fft(I) \quad (3)$$

3.3 Adaptor-based Fine-tuning

Figure 2 shows the structure of an Adaptor module used for fine-tuning. The tunable components (marked with flames) are updated. The Embedding Tune and HFC Tune modules adapt the embeddings and high-frequency features, respectively, and their outputs are combined. The GELU activation introduces non-linearity. The MLP_{tune}^i is specifically fine-tuned, while MLP_{up} layer, both shared and unshared, manage up-projection, allowing for a balance between shared learning and task-specific adaptation.

Figure 3 illustrates the integration of Adaptor modules within a Vision Transformer (ViT) for water body segmentation. It includes both frozen and tunable components: the frozen parts (marked with an ice crystal) remain unchanged during training, while the tunable components (marked with flames) are updated. The model starts with patch embedding of the input dataset, followed by a series of transformer layers that leverage frozen pre-trained features. The key innovation lies in integrating Adaptors between these transformer layers. Each Adaptor is specifically fine-tuned to handle unique aspects of water body segmentation, such as distinguishing between water and non-water regions under varying environmental conditions. This selective fine-tuning mechanism effectively balances computational efficiency with task-specific adaptability, resulting in robust segmentation outcomes suitable for diverse real-world scenarios.

The core contribution of this work is the Adaptor-based fine-tuning mechanism, which allows selective tuning of the frozen model layers. The Adaptors are lightweight modules designed to modify both the image embeddings and the high-frequency components in a task-specific manner.

Two types of tunable components are defined. One is Embedding Tune: This module fine-tunes the image embeddings E by learning a projection to a lower-dimensional space using the Equation (4):

$$F_{pe} = L_{pe}(E) \quad (4)$$

where L_{pe} is a linear layer that maps the embeddings to a task-specific feature space, and F_{pe} is the fine-tuned embedding output. Another one is HFC Tune: For high-frequency components I_{hf} , an additional layer of fine-tuning is applied. The high-frequency patches are mapped into a low-dimensional representation through a linear layer using the Equation (5):

$$F_{hfc} = L_{hfc}(I_{hf}) \quad (5)$$

where, L_{hfc} is a linear transformation that compresses the high-frequency information into a task-specific form, F_{hfc} , which is then combined with the embeddings.

The final feature used for segmentation is the combination of both embedding and high-frequency tuned components using the Equation (6):

$$p_i = MLP_{adaptor}(GELU(F_{pe} + F_{hfc})) \quad (6)$$

where p_i is the prompt for the i -th transformer layer and $MLP_{adaptor}$ is a multi-layer perceptron tasked with merging these features. The activation function $GELU$ introduces non-linearity, and the summed features are passed to subsequent layers in the transformer.

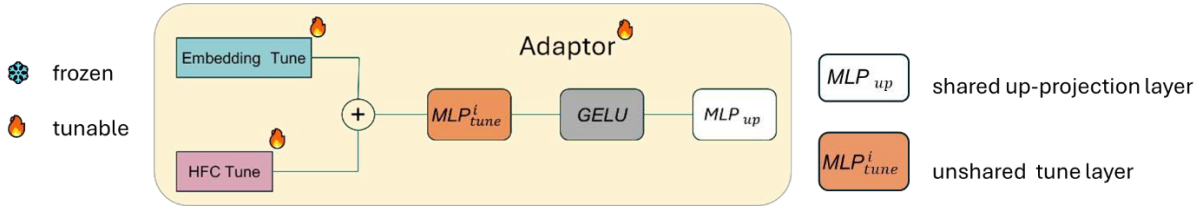


Figure 2. The Architecture of Adaptor-Based Fine-Tuning Modules

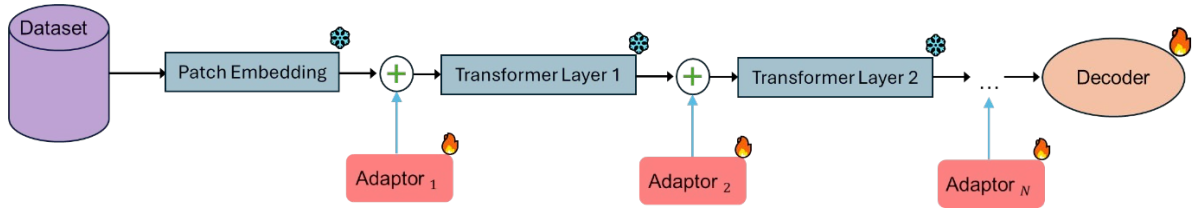


Figure 3. Adaptor Integration in Vision Transformer for Water Body Segmentation

3.4 Efficiency of the Adaptor Network

The Adaptor network works by adjusting only the critical layers responsible for handling task-specific challenges in segmentation, such as reflections or occlusions that are common in water body images. This selective fine-tuning drastically reduces the computational cost compared to traditional methods that retrain entire models.

Given the task of water body segmentation, which often involves challenging scenarios such as subtle distinctions between water and land under different environmental conditions, the fine-tuned high-frequency and embedding components enhance the model's robustness. The compact nature of the Adaptor enables efficient transfer learning without sacrificing performance.

The efficiency of the model can be quantified by comparing the total number of parameters trained in traditional full-scale fine-tuning versus the parameters trained in the Adaptor-based method. θ_t denotes the parameters trained in a traditional method and θ_a denotes the parameters trained in the Adaptor-based method, achieving a significant reduction in trainable parameters, as Equation (7) shows:

$$\frac{\theta_a}{\theta_t} \ll 1 \quad (7)$$

where θ_a is typically an order of magnitude smaller than θ_t . The computational complexity is further reduced due to the focused fine-tuning of high-frequency components. Instead of performing a full forward and backward pass for all model layers, the complexity is concentrated in the computation of the high-frequency components and their subsequent fine-tuning, as Equation (8) shows.

$$O(Adaptor) = O(L_{pe}) + O(L_{hfc}) \quad (8)$$

$O(Adaptor)$ represents the overall computational complexity of the Adaptor-based method. $O(L_{pe})$ represents the computational complexity of the linear layer L_{pe} used for fine-tuning the embeddings. $O(L_{hfc})$ represents the computational complexity of the linear layer L_{hfc} used for fine-tuning the high-frequency components. This complexity remains manageable, even for large images, as only the most important task-specific features are adjusted, making the method suitable for real-world applications in resource-constrained environments.

4. EXPERIMENTS

The experiments were conducted using Adaptor-based fine-tuning on the SegFormer model as backbone. The training was performed on a single NVIDIA GeForce RTX 3080 GPU for 50 epochs, using a batch size of 4 and an initial learning rate of 0.0002, with the Adam optimizer.

4.1 Dataset

The ATLANTIS dataset (Erfani et al., 2022) is recognized as the largest annotated collection for semantic segmentation of water bodies and their related facilities, consisting of 56 categories of 5,195 training images and

1,296 testing images as Table 1 shows. This study specifically focused on 15 distinct water body categories without their related facilities from the ATLANTIS dataset, as Table 2 shows. This subset includes 1,609 training images, 260 validation images, and 662 testing images. The images are resized to 352×352 for consistent evaluation. For performance assessment, the mean Intersection over Union (mIoU) is utilized as the evaluation metric.

Table 1. Complete List of Categories from the ATLANTIS Dataset

Number	Label	Number	Label	Number	Label	Number	Label
0	background	1	bicycle	2	boat	3	Breakwater
4	bridge	5	building	6	bus	7	canal
8	car	9	cliff	10	culvert	11	cypress tree
12	dam	13	ditch	14	fence	15	hydrant
16	fjord	17	flood	18	glaciers	19	hot spring
20	lake	21	levee	22	lighthouse	23	mangrove
24	marsh	25	motorcycle	26	offshore	27	parking
28	person	29	pier	30	pipeline	31	pole
32	puddle	33	rapids	34	reservoir	35	river
36	river delta	37	road	38	sea	39	ship
40	shoreline	41	sidewalk	42	sky	43	snow
44	spillway	45	swimming pool	46	terrain	47	traffic sign
48	train	49	truck	50	umbrella	51	vegetation
52	wall	53	water tower	54	water well	55	waterfall
56	wetland						

Table 2. Selected Water Body Categories Used for Segmentation in This Study

Number	Label	Number	Label	Number	Label	Number	Label
0	background	7	canal	13	ditch	16	fjord
17	flood	19	hot spring	20	lake	32	puddle
33	rapids	34	reservoir	35	river	36	River delta
38	sea	45	swimming pool	55	waterfall	56	wetland

4.2 Data Preprocessing

In the data preprocessing stage, the raw images were transformed to facilitate effective segmentation. As illustrated in Figure 4, the original images were annotated with multi-class masks, distinguishing between various elements within the scene, including different types of water bodies and non-water features. To enhance the focus of water body segmentation and simplify the model's learning process, these original multi-class masks were converted into binary masks. This conversion produced processed masks, where the water bodies were distinguished from the background, thereby framing the segmentation task as a foreground-background classification problem.

This approach reduced the complexity of the segmentation model by focusing solely on the target water bodies, thus allowing the model to concentrate on differentiating water from non-water areas without the need to classify multiple object categories. The preprocessing pipeline ensured that each image was represented consistently, facilitating robust training of the segmentation network across diverse types of water body environments.

4.3 Comparison with Existing Models

The results of the introduced Adaptor-based fine-tuning method for water body segmentation are presented in Table 3 and Table 4. The evaluation metric used is Intersection over Union (IoU), which measures the accuracy of segmentation by comparing the predicted output with the ground truth data.

Table 3 presents the performance comparison of this method, Adaptor, against various state-of-the-art models such as DNLNet, GCNet, DeepLabv3, and AQUANet across 15 water body categories. The Adaptor method demonstrates a clear improvement in the mean IoU (mIoU) with 79.38%, outperforming all the baseline models significantly. For example, fjord: The Adaptor method achieves an IoU of 91.7%, which is much higher compared to DNLNet (48.8%) and GCNet (44.7%). River delta: Adaptor achieves 88.8% IoU, surpassing all other

methods, including PSPNet (65.5%) and OCNet (65.9%). Flood: The performance of the Adaptor method is 84.6%, whereas models like CCNet and EMANet achieve only 26.9% and 23.0%, respectively.

These results indicate that the selective fine-tuning of task-specific layers using Adaptor-based networks effectively captures the unique characteristics of water bodies under varying conditions, such as occlusions and reflections, leading to superior segmentation accuracy.

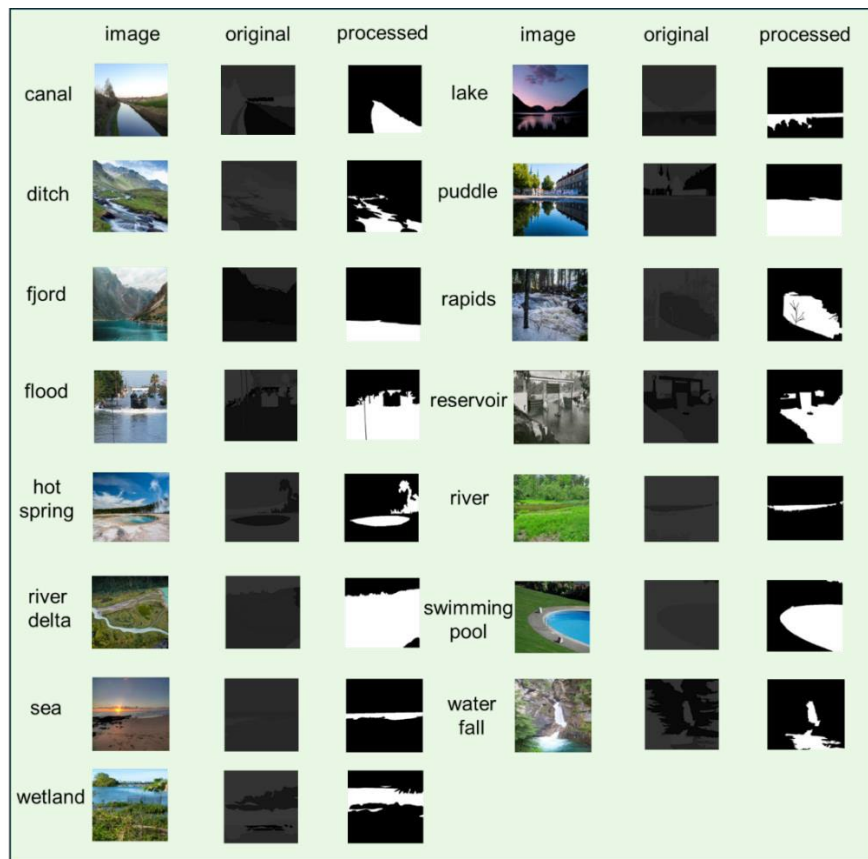


Figure 4. Transformation of Original Dataset Labels into Processed Masks for Model Training and Testing

4.4 Comparison with different fine-tuning methods

Figures 5, 6, and 7 demonstrate the performance of different fine-tuning methods, namely AdaptFormer, Linear, and Adaptor, on water body segmentation across various types of water bodies. Each figure presents a comparison of segmentation results for distinct water body categories, highlighting the Intersection over Union (IoU) values achieved by each method.

Table 4 illustrates the performance comparison between the Adaptor method and other related approaches, Adaptformer, and the Linear fine-tuning method, all based on the same backbone model, Segformer. The experimental results show that the Adaptor method consistently outperforms both Adaptformer and the Linear approach across various water body categories. Specifically, in the canal category, the Adaptor method achieved an Intersection over Union (IoU) score of 83.6%, significantly surpassing Adaptformer (26.0%). In the lake category, the Adaptor method reached an IoU of 90.9%, compared to Adaptformer at 37.1% and Linear at 85.1%. For the reservoir category, the Adaptor method achieved an IoU of 82.7%, while the Linear method performed slightly lower, with an IoU of 75.2%. The results demonstrate that the Adaptor-based fine-tuning strategy is more effective than both Adaptformer and Linear, in challenging water body categories. The method's ability to tune high-frequency components and embedding features enables it to handle variations in water appearance, such as seasonal changes and reflective surfaces.

The experimental results clearly highlight the strength of the Adaptor-based fine-tuning method in achieving high segmentation accuracy across diverse water body types. By focusing on selective tuning, the approach significantly reduces computational costs compared to full-scale retraining while maintaining a high level of accuracy. These results demonstrate the feasibility of using Adaptor-based models for real-world water body segmentation tasks, offering both efficiency and robustness.

Table 3. Performance Comparison of State-of-the-Art Models for Water Body Segmentation, the best result is bold

Method	IoU (%)															
	canal	ditch	fjord	flood	hot spring	lake	puddle	rapids	reservoir	river	river delta	sea	swimming pool	waterfall	wetland	mIoU (%)
DNLNet	54.4	26.3	48.8	36.3	55.3	35.5	52.3	40.4	32.1	31.3	37.1	61.7	52.4	48.7	54.6	44.48
GCNet	56.6	19.0	44.7	34.8	36.1	35.8	39.4	39.9	41.6	32.4	67.0	62.2	42.9	50.7	59.7	44.19
OCRNet	52.4	19.4	46.9	34.9	58.8	30.4	39.7	42.5	29.8	31.9	55.5	55.4	43.6	56.8	51.5	43.30
CCNet	41.1	17.4	35.2	26.9	47.9	18.6	43.8	29.9	16.6	23.7	48.3	53.3	38.4	51.1	34.1	35.09
EMANet	46.1	16.6	27.1	23.0	63.7	17.2	43.6	42.2	17.2	21.0	68.6	53.5	36.1	52.1	36.2	37.61
ANNet	50.9	22.8	31.6	32.0	58.1	25.6	52.9	48.4	20.8	28.6	56.8	60.4	43.9	57.9	51.4	42.81
DANet	50.5	34.1	37.1	37.0	61.6	23.8	51.5	42.8	30.2	31.5	63.5	60.4	43.1	55.2	54.6	45.13
DeepLabv3	52.5	27.2	52.3	43.8	42.5	31.1	54.2	46.0	34.4	27.1	51.1	61.5	53.6	52.8	52.9	45.40
PSPNet	53.8	29.0	42.9	46.5	53.9	29.7	54.7	38.2	29.8	28.8	65.5	63.5	47.7	48.4	47.5	45.33
OCNet	56.4	33.6	48.0	37.3	55.2	29.2	50.6	43.8	35.1	35.6	65.9	62.7	47.9	53.1	54.9	47.29
AQUANet	55.0	27.7	53.4	47.0	60.5	33.2	54.4	46.3	39.0	34.7	63.2	64.2	44.9	53.0	66.1	49.51
Adaptor	83.6	58.1	91.7	84.6	74.1	90.9	74.9	86.1	82.7	84.2	88.8	86.7	61.5	71.9	72.4	79.38

Table 4. Performance Comparison of Adaptor-Based Fine-Tuning with Other Methods, the best result is bold

Method	IoU (%)															
	canal	ditch	fjord	flood	hot spring	lake	puddle	rapids	reservoir	river	river delta	sea	swimming pool	waterfall	wetland	mIoU (%)
adaptformer	26.0	12.6	30.4	19.1	25.4	37.1	13.3	33.0	32.9	31.5	24.8	35.1	26.6	31.0	24.6	26.68
linear	79.3	50.2	83.2	76.9	68.7	85.1	59.2	79.0	75.2	78.2	78.4	82.1	59.1	68.7	65.0	72.65
Adaptor	83.6	58.1	91.7	84.6	74.1	90.9	74.9	86.1	82.7	84.2	88.8	86.7	61.5	71.9	72.4	79.38

5. DISCUSSION

The experimental results demonstrate that the Adaptor-based fine-tuning approach provides superior segmentation performance across diverse water body categories compared to conventional full-scale retraining methods. This approach, which utilizes Adaptors, significantly reduces computational costs without sacrificing accuracy. The improvement in mean Intersection over Union (mIoU), highlights the efficacy of leveraging high-frequency component tuning alongside embedding-based adjustments. The significant gains observed for categories such as fjord, river delta, and flood indicate that the Adaptor-based method effectively handles challenges like occlusions, reflections, and seasonal variations. Unlike traditional models, which often struggle to distinguish between visually similar regions, this approach's ability to retain pre-trained knowledge while adapting to specific features of water bodies leads to a more robust segmentation output.

Another key observation is the effectiveness of Adaptor networks when compared to recent related approaches such as adaptformer and linear fine-tuning. The Adaptor method consistently outperforms these methods, particularly in challenging categories such as "canal" and "lake," which require precise boundary detection. The combination of embedding and high-frequency component tuning allows the ViT model to handle complex water body characteristics, ensuring that subtle differences between water and non-water regions are accurately captured. The reduction in trainable parameters, quantified in equation (7), further validates the efficiency of the method, making it suitable for deployment in real-world, resource-constrained environments. This efficiency is critical for applications in environmental monitoring, where computational resources are often limited, and rapid processing is required for decision-making.

Compared to existing segmentation models, the adaptor-based method significantly reduces the number of parameters. Models such as ANNet (63.1M), GCNet (28.1M), and DeepLabv3 (15.4M) exhibit substantially

higher parameter counts, whereas the adaptor-based method achieves remarkable efficiency with only 0.55M parameters. This parameter efficiency makes it highly suitable for resource-constrained environments, enabling a substantial reduction in computational demands while maintaining competitive performance. Notably, training this model required only 3 hours. Furthermore, the existing segmentation models used for comparison were trained from scratch on the dataset in this study, rather than leveraging pre-trained models, with training times ranging from 7 hours to 2 days. The adaptor-based method demonstrates a clear advantage in terms of speed and efficiency.


























	Input	GT	adaptformer	linear	adaptor
canal			 53.1%	 90.9%	 91.0%
ditch			 54.8%	 70.1%	 61.1%
fjord			 37.5%	 96.7%	 98.2%
flood			 33.6%	 90.6%	 96.5%
hot spring			 45.1%	 65.5%	 74.7%

Figure 5. IoU for Various Water Body Segmentation using Different Light-Weight Fine-Tuning Methods


























	Input	GT	adaptformer	linear	adaptor
lake			 56.0%	 62.3%	 67.6%
puddle			 27.5%	 84.0%	 96.6%
rapids			 55.8%	 61.5%	 79.3%
reservoir			 17.2%	 81.5%	 87.1%
river			 74.3%	 69.1%	 69.1%

Figure 6. IoU for Various Water Body Segmentation using Different Light-Weight Fine-Tuning Methods


























	Input	GT	adaptformer	linear	adaptor
river delta			 5.4%	 74.8%	 89.5%
sea			 82.6%	 61.5%	 75.0%
swimming pool			 36.2%	 60.1%	 52.9%
water fall			 17.0%	 81.1%	 83.1%
wetland			 64.5%	 80.8%	 82.7%

Figure 7. IoU for Various Water Body Segmentation using Different Light-Weight Fine-Tuning Methods

6. CONCLUSIONS

This paper presents a lightweight, Adaptor-based fine-tuning framework for water body segmentation that significantly improves both computational efficiency and segmentation accuracy. By adding Adaptors in Vision Transformers, the approach avoids the need for full-scale retraining, thereby reducing the computational load and training time. The experimental results indicate that this Adaptor-based method outperforms existing state-of-the-art models and related fine-tuning approaches, achieving a mean IoU of 79.38% across 15 water body categories. The ability to effectively capture high-frequency details and adapt embeddings to specific segmentation tasks allows for robust performance, even under challenging conditions such as occlusions, reflections, and seasonal variations. This research's contributions include demonstrating the feasibility of using lightweight Adaptor networks for effective transfer learning, particularly in the context of water body segmentation. Future work will explore extending this methodology to other environmental monitoring tasks, leveraging the flexibility and efficiency of Adaptor-based models to address various segmentation challenges in remote sensing and geographic information systems (GIS).

ACKNOWLEDGMENTS

This work was supported by JST SPRING, Grant Number JPMJSP2138.

REFERENCES

- Cao, Y., Xu, J., Lin, S., Wei, F., & Hu, H. (2019). GCNet: Non-Local Networks Meet Squeeze-Excitation Networks and Beyond. 0–0. https://openaccess.thecvf.com/content_ICCVW_2019/html/NeurArch/Cao_GCNet_Non-Local_Networks_Meet_Squeeze-Excitation_Networks_and_Beyond_ICCVW_2019_paper.html.
- Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4), 834–848.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255. <https://ieeexplore.ieee.org/abstract/document/5206848/>
- Dong, W., Yan, D., Lin, Z., & Wang, P. (2023). Efficient Adaptation of Large Vision Transformer via Adapter Re-Composing. *Advances in Neural Information Processing Systems*, 36, 52548–52567.
- Dosovitskiy, A. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv Preprint arXiv:2010.11929*.

- Erfani, S. M. H., Wu, Z., Wu, X., Wang, S., & Goharian, E. (2022). ATLANTIS: A benchmark for semantic segmentation of waterbody images. *Environmental Modelling & Software*, 149, 105333.
- Kadhim, I. J., & Premaratne, P. (2023). A Novel Deep Learning Framework for Water Body Segmentation from Satellite Images. *Arabian Journal for Science and Engineering*, 48(8), 10429–10440. <https://doi.org/10.1007/s13369-023-07680-5>
- Kenton, J. D. M.-W. C., & Toutanova, L. K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of naacL-HLT*, 1, 2. <https://www.waqaarana.me/assets/papers/N19-1423.pdf>
- Liu, W., Shen, X., Pun, C.-M., & Cun, X. (2023). Explicit visual prompting for low-level structure segmentations. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 19434–19445. http://openaccess.thecvf.com/content/CVPR2023/html/Liu_Explicit_Visual_Prompting_for_Low-Level_Structure_Segmentations_CVPR_2023_paper.html
- Ni, J., Wu, J., Elazab, A., Tong, J., & Chen, Z. (2022). DNL-Net: Deformed non-local neural network for blood vessel segmentation. *BMC Medical Imaging*, 22(1), 109. <https://doi.org/10.1186/s12880-022-00836-z>
- Pinaya, W. H. L., Vieira, S., Garcia-Dias, R., & Mechelli, A. (2020). Convolutional neural networks. In *Machine learning* (pp. 173–191). Elsevier. <https://www.sciencedirect.com/science/article/pii/B9780128157398000109>
- Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., & Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140), 1–67.
- Saleh, F. S., Aliakbarian, M. S., Salzmann, M., Petersson, L., & Alvarez, J. M. (2018). Effective use of synthetic data for urban scene semantic segmentation. *Proceedings of the European Conference on Computer Vision (ECCV)*, 84–100. http://openaccess.thecvf.com/content_ECCV_2018/html/Fatemeh_Sadat_Saleh_Effective_Use_of_ECCV_2018_paper.html
- Wang, H., Wu, X., Huang, Z., & Xing, E. P. (2020). High-frequency component helps explain the generalization of convolutional neural networks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8684–8694. http://openaccess.thecvf.com/content_CVPR_2020/html/Wang_High-Frequency_Component_Helps_Explains_the_Generalization_of_Convolutional_Neural_Networks_CVPR_2020_paper.html
- Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., & Luo, P. (2021). SegFormer: Simple and efficient design for semantic segmentation with transformers. *Advances in Neural Information Processing Systems*, 34, 12077–12090.
- Yuan, K., Zhuang, X., Schaefer, G., Feng, J., Guan, L., & Fang, H. (2021). Deep-learning-based multispectral satellite image segmentation for water body detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 7422–7434.
- Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid scene parsing network. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2881–2890. http://openaccess.thecvf.com/content_cvpr_2017/html/Zhao_Pyramid_Scene_Parsing_CVPR_2017_paper.html

SEMANTIC-ENRICHED IMAGE RETRIEVAL FOR BRIDGE DAMAGE ASSESSMENT

Chengzhang Chai¹, Jiucui Liu², Yan Gao³, Guanyu Xiong⁴ and Haijiang Li^{5*}

1) Ph.D. Candidate, School of Engineering, Cardiff University, Cardiff, UK. Email: chaic1@cardiff.ac.uk

2) Ph.D. Candidate, School of Engineering, Cardiff University, Cardiff, UK. Email: liuj151@cardiff.ac.uk

3) Ph.D., School of Engineering, Cardiff University, Cardiff, UK. Email: gaoy74@cardiff.ac.uk

4) Ph.D. Candidate, School of Engineering, Cardiff University, Cardiff, UK. Email: xiongg@cardiff.ac.uk

5) Prof., School of Engineering, Cardiff University, Cardiff, UK. Email: lih@cardiff.ac.uk

* Corresponding author

Abstract: Periodic bridge damage inspections result in a vast number of image records stored in a database, which can be used as a reference for the subsequent damage assessment. However, the currently used content-based image retrieval (CBIR) techniques are limited by the 'semantic gap'. They tend to consider only the low-level visual features in an image and ignore the high-level semantic information. This study proposes a semantic-enriched image retrieval framework (SEIR-Net) for bridge damage assessment. The framework enables the image encoder to extract low-level visual features and high-level semantic information by fine-tuning the multi-modal image captioning model (CNN-LSTM). The high-dimensional vectors extracted using the fine-tuned encoder are stored in the FAISS vector database, and efficient retrieval is achieved based on L2 Euclidean distance. Retrieval evaluation was performed on a damage dataset constructed on the real-world bridge inspection report, and our proposed method outperforms the commonly used VGG-16 and ResNet-50 models on the mAP and Recall@K (K=1, 2, 4) metrics. These results suggest that incorporating the semantic content of damage in image retrieval would be more beneficial for assessment references. In summary, this study effectively enhances the utility of historical image records in bridge damage assessment through semantic-enriched image retrieval techniques.

Keywords: Deep learning, Semantic enriched, Image retrieval, Bridge damage

1. INTRODUCTION

Bridges are one of the key components in the global transport network. The aging of bridges is now a growing problem. The American Society of Civil Engineers (ASCE) released the America's Infrastructure 2021 Report Card, which states that there are more than 617,000 bridges across the United States (ASCE, 2021). 42% of these bridges are at least 50 years old, and as many as 46,154 (7.5% of the nation's bridges) are considered to be at risk for structural deficiencies (Ali et al., 2022; Cha et al., 2024; Spencer et al., 2019). These figures highlight the urgency of carrying out inspections and raise a widespread concern.

Regular damage inspections generate many image records stored in a database, which can be a valuable reference. As the saying goes, 'A picture is worth a thousand words.' Engineers can retrieve images of the same type or similar scenarios from the historical records as a reference to assist in on-site diagnosis during subsequent damage assessment. Therefore, the effective use of such image information has become a vital issue in bridge maintenance assessment.

Image retrieval techniques are generally classified into two categories: tag-based methods and content-based methods. Tag-based image retrieval (TBIR) relies on information provided by metadata or text for retrieval (Lee et al., 2017). It usually includes pre-defined template information such as name, time of capture, geographic location, device information, and other relevant details. For example, Sun et al. proposed a generic framework for tag-driven social image retrieval, starting from five orthogonal dimensions (tag relatedness, tag discrimination, tag length normalization, tag query matching model, and query model) to improve the accuracy of image retrieval (Sun et al., 2011). Li et al. developed a retrieval system called 'BIMSeek', which calculates the similarity between components based on the attribute information (including geometric features and semantic attributes) of BIM components, thus enabling efficient retrieval of components (Li et al., 2020). Ma et al. used photos' metadata (including information such as photographed location, camera perspective, and image semantic content) to build an image retrieval system that efficiently retrieves large amounts of facility management data (Ma et al., 2021).

In comparison, content-based image retrieval (CBIR), which typically relies on visual features of images such as texture, shape, colour, and spatial information, has made more significant progress in recent years (Dubey, 2021; Li et al., 2021). In an early exploratory phase, Brilakis and Soibelman completed an attempt at content-based construction image retrieval using the technique of blind relevance feedback (Brilakis & Soibelman, 2005). The developed system makes better use of the image's texture features and shape features to quickly find relevant images compared to traditional tag-based retrieval methods. In addition, Brilakis et al. used a clustering approach to assign the material feature information extracted by Fourier analysis and edge detection to the nearest centroid clusters (Brilakis et al., 2006). The presence of construction materials in the image is detected by matching with pre-defined material samples, which in turn completes the image retrieval. The above explorations mainly started from a single feature, i.e., only the visual features of the image were considered for retrieval. Based on this, Brilakis

and Soibelman made further improvements by introducing meta-information, including date and location, to achieve more accurate multi-feature construction material image retrieval (Brilakis & Soibelman, 2008). The win of the AlexNet model in the 2012 ImageNet image classification competition marked an essential breakthrough in deep learning and convolutional neural networks (CNN). Feature extraction methods in image retrieval tasks have thus gradually shifted from traditional manual feature-based to deep learning-based approaches. Ha et al. used a rendered BIM model to construct the dataset and a pre-trained VGG network for image feature extraction (Ha et al., 2018). Firstly, the similarity between the rendered image and the real captured indoor image is evaluated, after which the positional and directional information contained in the rendered image is combined to jointly complete the estimation of indoor positioning. Wang et al. argued that using existing content-based methods to retrieve construction site scenes is limited to utilising simple visual features of the entire image, which makes it difficult to distinguish similar details in related scenes (Wang et al., 2023). Therefore, they used the object detection model to acquire the critical sub-regions in the construction image and employed CNN to extract these fine features, thus effectively improving the performance of the retrieval system.

It is worth noting that existing applications of image retrieval tasks in civil engineering fall into two main categories. One category is the safety management and identification of workers, materials and machinery in construction site scenarios, and the other category is the identification of components in BIM models, thus enabling the alignment of images with BIM data. However, there are few applications in the field of bridge damage detection. The closest to our study is the retrieval of bridge deck images by Wogen et al. (Wogen et al., 2024). This study used a Siamese convolutional neural network to extract image features. Based on a composite similarity measure that combines image features with deck geolocation information, effective retrieval of bridge inspection images was achieved. However, this study's shortcomings are firstly that fine-tuning the Siamese convolutional neural network relies on manually labelled 14,959 positive and negative samples. Such labelling is time-consuming and challenging. Secondly, this study focuses on single-deck component retrieval and does not explore different damage types in real-world bridges. Finally, GPS data may be incomplete or inaccurate in real-world applications, especially in the defective areas underneath the bridge structure. This situation can directly affect the retrieval validity of the model. In the scenario of bridge damage assessment, relying solely on the simple visual features of the images extracted by CNN is often insufficient to retrieve the corresponding damaged images accurately and comprehensively. Therefore, additional information should be provided to assist the retrieval as much as possible.

In this study, a novel semantic-enriched image retrieval framework (SEIR-Net) is developed to address these challenges. This framework is different from existing methods that rely solely on CNN encoders to extract image low-level visual features. Our core idea is to make this encoder capable of spanning from low-level visual features, such as edges, textures, etc. to high-level semantic features by fine-tuning the multi-modal image captioning model (CNN-LSTM). This feature extraction method can capture rich semantic information in the image, such as the type of bridge damage, the location of occurrence, etc., which enables the retrieval process to match the image accurately based on visual and semantic cues. The main contributions of this research are as follows:

1. A semantic-enriched image retrieval framework is proposed. By fine-tuning the CNN-LSTM multi-modal model, the cross-entropy loss function is used to adjust the positional distribution of the encoder vectors in the feature space to provide it with the ability to capture the high-level semantic information implicit in the images.
2. A Faiss-based vector database is adopted for efficient storage and fast retrieval. By storing the extracted high-dimensional image features in the Faiss vector database, the L2 Euclidean distance is used as the retrieval algorithm to calculate the similarity of the feature vectors in the space and ultimately achieve accurate retrieval.
3. Experiments on a real-world constructed bridge damage dataset verify its effectiveness as an assessment tool. This method can assist engineers in efficiently retrieving historical damage images, which provides practical support for bridge damage assessment.

2. METHODOLOGY

2.1 Problem Statement

As can be seen from the research related to content-based image retrieval, using convolutional neural network to extract visual features from image data for similarity matching has become one of the most effective methods for image retrieval. This success can be attributed to CNN's automatic feature learning capability and the fact that the extracted high-dimensional feature vectors can be efficiently metricised by the distance function for similar image matching.

However, the core challenge in image retrieval is the semantic gap between its high-level semantic information and low-level visual features (Wan et al., 2014). The CNN network used in existing retrieval methods is mainly responsible for extracting low-level visual features, but has limited ability to capture high-level semantic information contained in the image, thus affecting the accuracy and relevance of the retrieval effect. This problem is even more critical in the scenario of bridge damage assessment. When engineers conduct the bridge damage assessment, they need to retrieve historical images with similar damage information to assist in diagnosis. Therefore, high-level semantic information in images (such as damage type and location) plays a vital role in

retrieval. Effectively bridging the semantic gap between low-level visual features and high-level semantic information has become a critical challenge in applying the image retrieval technique for bridge damage assessment.

2.2 Overall Framework

This study proposes a semantic-enriched image retrieval framework (shown in Figure. 1) to solve the above conflict, which consists of four parts: dataset construction, semantic enrichment, feature indexing, and image retrieval. Firstly, the dataset is constructed by preparing the damaged images with corresponding descriptions based on the bridge inspection report, which will be described in detail in the Section 3.1. Following that, in the semantic enrichment stage, the encoder is trained through supervised learning to capture high-level semantic information and low-level visual features from the images. Next, in the feature indexing stage, the feature vectors of the test image set are extracted using the fine-tuned encoder and saved to the FAISS vector database for efficient storage. Finally, the query image is input in the image retrieval stage, and the fine-tuned encoder extracts the query features. Similarity matching is then performed with the stored feature vectors in the FAISS vector database to retrieve the most similar historical images.

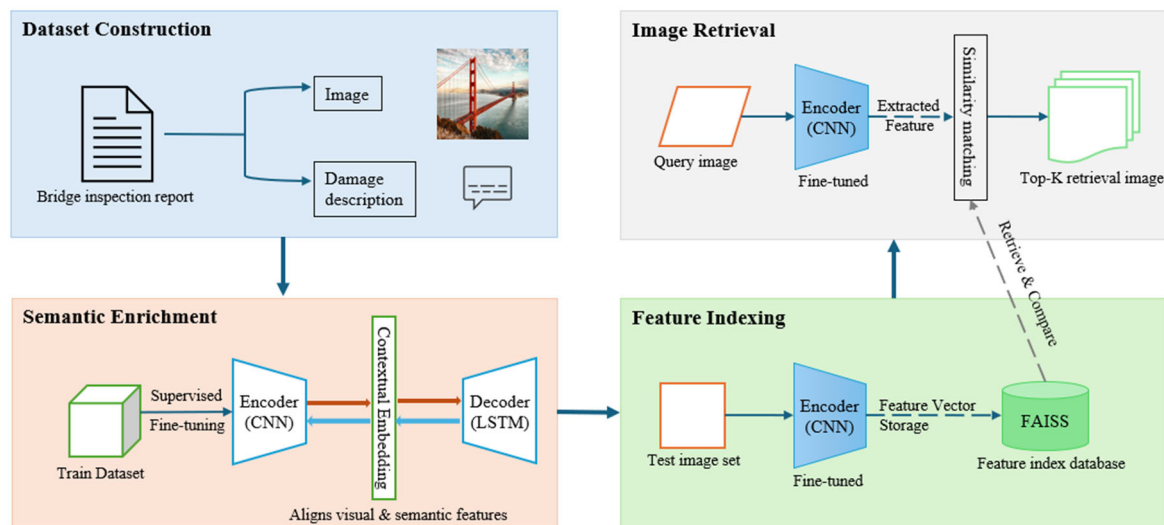


Figure 1. Semantic-enriched image retrieval framework

2.3 Semantic Enrichment

The semantic enrichment component is the core part of this SEIR-Net framework, as shown in Figure 2. Its purpose is to fine-tune the multi-modal image captioning model (CNN-LSTM) so that the fine-tuned encoder not only extracts the image's low-level visual features but also considers the high-level semantic information simultaneously. In this framework, the CNN acts as an encoder, which is mainly responsible for extracting low-level visual features (edges, textures, and colours) within an image, and the LSTM acts as a decoder, which can convert the visual features extracted by the CNN into semantic text sequences through sequence modelling. The model performs supervised learning on image-text pairs through a cross-entropy loss function. The position of the encoder in the feature space is continuously optimised through the backpropagation of parameters. As training proceeds, the CNN encoder gradually learns to extract visual features that are highly correlated with the semantic sequence of text. This process allows the encoder to capture both low-level visual and high-level semantic information from images in the end.

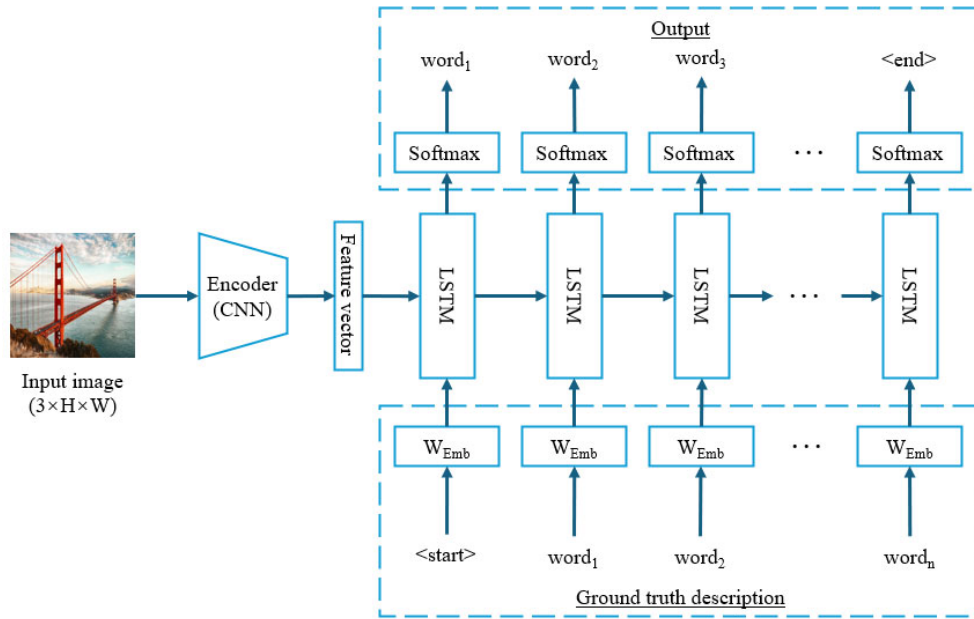


Figure 2. Semantic enrichment using CNN-LSTM

2.4 Feature Indexing and Image Retrieval

In the feature indexing stage, the features of the image set are first extracted using the encoder part that has been fine-tuned by supervised learning in the semantic enrichment stage. After high-dimensional vectorisation, these extracted features are stored in the Facebook AI Similarity Search (FAISS) vector database. FAISS is an open-source library for efficient similarity search and clustering of dense vectors (Douze et al., 2024). In the framework of SEIR-Net, the vector storage uses the IndexFlatL2 index structure. This index structure implements vector retrieval based on L2 Euclidean distance, which can process large-scale data quickly and supports similarity search of vectors in high-dimensional spaces. At the same time, the index will be preliminarily optimised to obtain the distribution characteristics of the vectors before storage to better accelerate the subsequent similarity matching.

In the retrieval stage, the engineer will provide an image of the bridge damage to be queried, and the system will perform feature extraction using a fine-tuned encoder (CNN part) to obtain high-dimensional vectors. Then, the distance metric function (L2 Euclidean distance) is used to match the similarity with the feature vectors in the FAISS database, as shown in Equation 1. The Euclidean distance (L2 paradigm) is a metric formula between two vectors in Euclidean space. It is calculated by summing the squares of the differences between two vectors in each dimension and then taking the square root. The smaller the distance value, the more similar the two vectors are in the feature space. The system will sort by similarity distance and retrieve the K historical images from the FAISS database that are most similar to the query image, i.e., Top-K retrieval results. These retrieval results will be returned to the engineer as a matched result set of the query image to provide reference support for bridge damage assessment.

$$d(I_1, I_2) = \sqrt{\sum_{i=1}^k (I_{1i} - I_{2i})^2} \quad (1)$$

2.5 Loss Function

The loss function is mainly used for supervised training of multi-modal image captioning models in the semantic enrichment stage of the SEIR-Net framework. We use the cross-entropy loss function shown in Equation 2, and the parameters of the encoder and decoder will be continuously updated by backpropagation during the training process. This update allows the encoder to gradually acquire high-level semantic information based on its ability to extract low-level visual features. The function continuously compares the difference between the text sequence output by the LSTM decoder and the reference description to measure the difference between the predicted probability distribution and the actual target distribution. The minimisation of this loss can contribute to a better match between the image features extracted by the model and the semantic description of the text sequence. Eventually, by alignment, the model encoder can represent both low-level visual features and high-level semantic information within the feature space.

$$L_{ce}(\theta) = - \sum_{t=1}^T \log(p_{\theta}(w_t^* | w_{1:t-1}^*)) \quad (2)$$

In the encoder-decoder framework of the image captioning model, the attention mechanism is sometimes prone to focus on local regions and ignore the global context (Zohourianshahzadi and Kalita, 2022). This phenomenon can lead to insufficiently accurately generated descriptions, and the encoder cannot extract enough semantic information in the feature space. Therefore, this study additionally uses the bi-stochastic attention regularisation to address this potential problem. This regularisation term enables the model to generate more comprehensive descriptions using penalty weights to ensure the attention is more evenly distributed.

3. EXPERIMENTS AND RESULTS

3.1 Experimental Design

Dataset: The core of the SEIR-Net framework is fine-tuning the CNN-LSTM multi-modal model in the semantic enrichment stage. This training process requires the preparation of bridge damage image-description datasets for supervised learning. Considering that no damage image-description dataset is publicly available, we chose the bridge inspection report provided by Centregreat Rail (CGR) to construct the dataset. This bridge inspection report details the investigation of bridge damage by engineers on the site. From it, we extracted images related to various damage types, each containing a textual description of the damage type and the location of the damage. In addition, we cleaned the data on the above images with corresponding descriptions to ensure that this constructed dataset was of high quality. The cleaned dataset has 429 sets of samples, divided into training set, validation set, and testing set in the ratio of 70%, 10%, and 20%.

Evaluation metrics: To comprehensively evaluate the performance of the SEIR-Net framework in the semantic enrichment and image retrieval stages, we used different evaluation metrics for measurement. To evaluate the effectiveness of the multi-modal model in generating text, we used the Bilingual Evaluation Understudy (BLEU) series of metrics, as shown in Equation 3, which has a score between 0 and 1. BLEU is a widely used evaluation criterion in natural language processing (Papineni et al., 2002). It is classified into BLEU-1 to BLEU-4 depending on the n-gram, where n-gram refers to the number of consecutive words. Thus, BLEU measures the model's ability to generate text at different levels, and a higher BLEU score also ensures that the encoder has captured relevant high-level semantic information in the feature space. For the evaluation of the effectiveness of the retrieval stage, we use the mean average precision (mAP) and recall, which are commonly used in image retrieval tasks, as evaluation metrics. The precision (P) is the proportion of relevant images in the returned retrieved images during a single retrieval, as shown in Equation 4. The recall (Q) is the ratio of the number of relevant images in the returned retrieved images to the number of actual relevant images in a single retrieval process, as shown in Equation 5. mAP evaluates the overall performance of the retrieval system by calculating the precision rate and averaging it using all the data with different recall. A higher mAP indicates that the model can retrieve more accurate images across various queries. recall@K, on the other hand, measures the model's ability to find relevant images in the first K retrieval results.

$$\text{BLEU} = \text{BP} \cdot \exp \left(\sum_{n=1}^N w_n \log p_n \right) \quad (3)$$

$$P = \frac{\text{Number of retrieved images related to the query image}}{\text{Number of retrieval images}} \quad (4)$$

$$R = \frac{\text{Number of retrieved images related to the query image}}{\text{Number of relevant images}} \quad (5)$$

Implementation detail: The implementation of the SEIR-Net framework is based on the environment of Python 3.8 and Pytorch 2.0.1. For hardware, the CPU processor is Intel(R) Core (TM) i7-10700KF CPU@ 3.80 GHz, and the GPU processor is NVIDIA GeForce RTX 3060 Ti. Fine-tuning the multi-modal model is at the core of the whole framework, and we set the number of training epochs to 50 and the batch size to 8 to better balance the training efficiency and the convergence effects. The learning rate is set to 1e-4 for the encoder (CNN), and 4e-4 for the decoder (LSTM), and this differential learning strategy will better balance the optimisation process of the two components. In addition, to effectively improve the generalisation ability of the model, we also set up a BLEU-4-based early stopping mechanism. If the effectiveness of the validation set does not improve in 10 consecutive epochs, the training stop will be triggered. This early stopping strategy can prevent the model from overfitting during training.

3.2 Results and Performance Evaluation

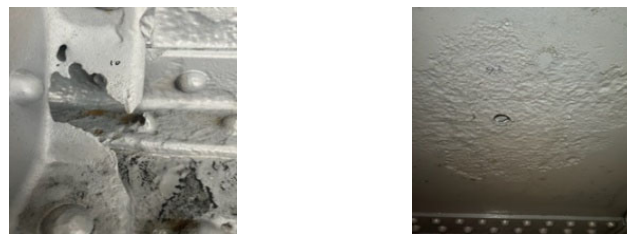
(1) Semantic-Enriched Multi-modal Model Fine-tuning Result

We first trained the multi-modal model and evaluated the fine-tuning results on the test set using the BLEU family of metrics mentioned in 3.1. The evaluation results include comparing the generated text with the reference text from BLEU-1 to BLEU-4, i.e., at the 1-gram level to the 4-gram level, as shown in Table 1.

Table 1. The assessment metrics for multimodal model

BLEU-1	BLEU-2	BLEU-3	BLEU-4
0.822	0.793	0.770	0.749

As shown in Table 1, the BLEU scores decrease as the n-gram in the assessment increases. These scores indicate that the model is most accurate at the word-level assessment dimension and can predict individual words accurately. As the assessment dimension becomes longer, moving to longer dimensions of phrases and sentence structures, the match between the generated text and the reference text decreases. This decline is because longer dimensions are evaluated with more demands on sentence fluency and semantic integrity. BLEU-4 achieves a score of 0.749, indicating that the model can generate text more consistent with the reference descriptions, meaning it could effectively capture semantic information about the context of the image in the feature space. Therefore, the encoder part of the model can be effectively used in the subsequent feature indexing and image retrieval stages. In addition, we provide some examples of generated descriptions in Figure. 3 to facilitate a more intuitive demonstration of the effectiveness of the multi-modal model.



Ref: section loss at the lower crank angle of stiffener
 Gen: section loss at the lower crank angle of stiffener
 Ref: hole at the deck plate
 Gen: hole at the deck plate

Figure 3. Generation description of multi-modal model (Ref: Reference text, Gen: Generated text)

(2) Evaluation of Retrieval Performance

Before performing image retrieval, the fine-tuned model encoder is first used to extract the image features on the test set. The extracted high-dimensional vectors need to be saved into the FAISS vector database for feature indexing. Then, mAP and Recall, as mentioned in 3.1 are used as the primary evaluation metrics in the image retrieval stage to assess the similarity retrieval performance of the images. Meanwhile, to better demonstrate the advancement of the SEIR-Net framework in retrieval performance, we compare it with the benchmark models of two CNN encoders (VGG-16 and ResNet-50) that are widely used in the current image retrieval field (Fang et al., 2022; Wang et al., 2023). The specific results are shown in Table 2.

Table 2. The comparative experiments on retrieval performance

Method	mAP	Recall@1	Recall@2	Recall@4
VGG-16	0.270	0.256	0.279	0.302
ResNet-50	0.628	0.581	0.663	0.724
SEIR-Net	0.642	0.628	0.686	0.744

As can be seen from Table 2, SEIR-Net outperforms VGG-16 and ResNet-50 in both mAP and Recall@K (K=1, 2, 4). This comparison fully demonstrates the advantages of the encoder fine-tuned with the multi-modal model. It can better capture high-level semantic information while extracting the low-level visual feature of the image, thus making the retrieved image more compatible with the query image in terms of visual and semantic content. In contrast, VGG-16 and ResNet-50 rely only on the extracted low-level visual features for image matching. They cannot effectively utilise the semantic information present within the image, resulting in a lower retrieval effect. In addition, to demonstrate the effectiveness of the SEIR-Net framework more intuitively, we provide some retrieval examples in Figures 4 and 5.



Figure 4. Visualisation examples of image retrieval (Green: Relevant, Red: Irrelevant)

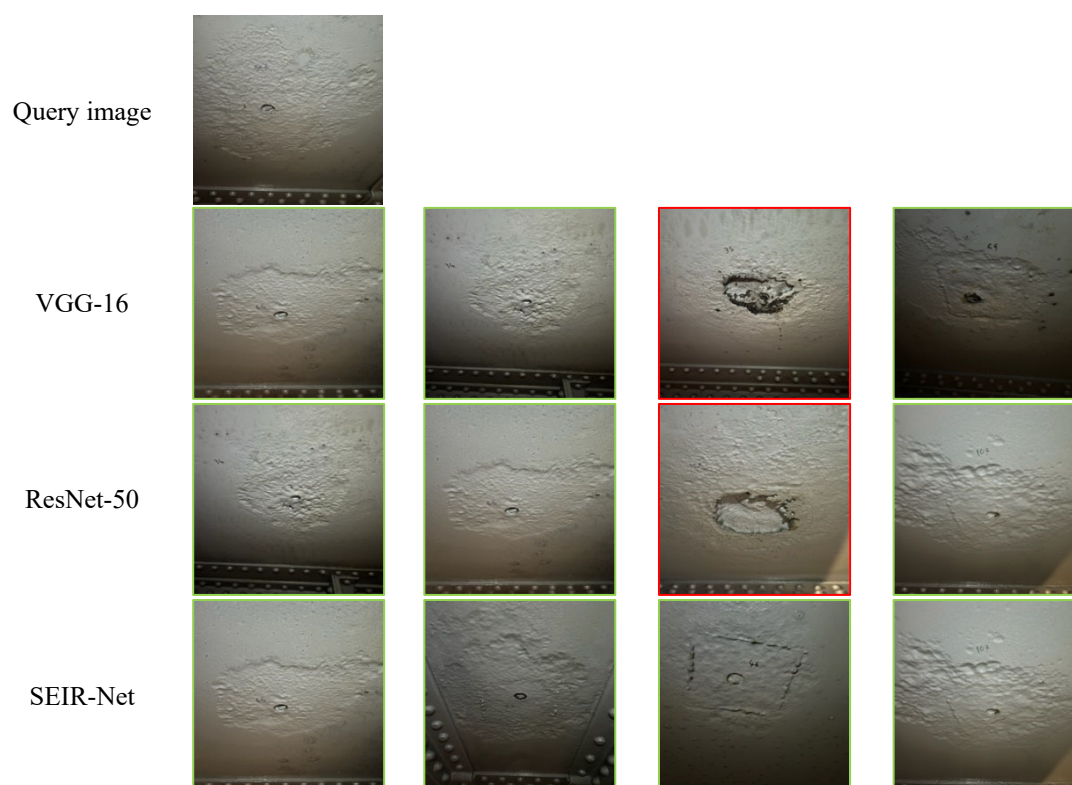


Figure 5. Visualisation examples of image retrieval (Green: Relevant, Red: Irrelevant)

These examples show that the retrieved images returned by the SEIR-Net framework have a higher correlation with the query image. It can effectively retrieve images from the historical damage database that are more consistent with the query image in terms of the semantic content of the damage. Thus, it has a higher utility in bridge damage assessment.

4. CONCLUSION

In this study, we propose a novel semantic-enriched image retrieval framework for bridge damage assessment. By incorporating semantic information into the retrieval model, we significantly enhance the relevance and accuracy of retrieval. Notably, the fine-tuning of the multi-modal image captioning model enables the image encoder to effectively capture the high-level semantic information embedded in the images while extracting visual low-level features. Comparative experiments on a damage dataset constructed based on the real-world bridge inspection report demonstrate that our proposed method outperforms the commonly used VGG-16 and ResNet-50 models on the mAP and Recall@K (K=1, 2, 4) metrics. It also shows that incorporating the semantic content of damage in image retrieval will be more beneficial for the damage assessment.

At the same time, this study has some limitations that need to be further addressed. Firstly, due to the lack of publicly available bridge damage datasets, the dataset we constructed is small and needs to be further expanded. Secondly, we need to try other deep learning models to explore better retrieval methods combined with multi-modal information to make the model more useful.

ACKNOWLEDGMENTS

This work is part of the Knowledge Transfer Partnerships (KTP) project, DIGIBRIDGE: BIM and Digital Twins in support of Smart Bridge Structural Surveying. The project receives funding from Innovate UK with reference number 10003208.

REFERENCES

- Ali, R., Chuah, J.H., Talip, M.S.A., Mokhtar, N. and Shoaib, M.A., 2022. Structural crack detection using deep convolutional neural networks. *Automation in Construction*, 133, pp.103989.
- ASCE, (2021). U.S. bridge report card. Retrieved from: <https://infrastructurereportcard.org/cat-item/bridges-infrastructure>.
- Brilakis, I. and Soibelman, L., 2005. Content-based search engines for construction image databases. *Automation in Construction*, 14(4), pp.537-550.
- Brilakis, I. and Soibelman, L., 2006. Multimodal image retrieval from construction databases and model-based systems. *Journal of Construction Engineering and Management*, 132(7), pp.777-785.
- Brilakis, I. and Soibelman, L., 2008. Shape-based retrieval of construction site photographs. *Journal of Computing in Civil Engineering*, 22(1), pp.14-20.
- Cha, Y.J., Ali, R., Lewis, J. and Büyüköztürk, O., 2024. Deep learning-based structural health monitoring. *Automation in Construction*, 161, p.105328.
- Douze, M., Guzhva, A., Deng, C., Johnson, J., Szilvasy, G., Mazaré, P.E., Lomeli, M., Hosseini, L. and Jégou, H., 2024. The faiss library. arXiv preprint arXiv:2401.08281.
- Dubey, S.R., 2021. A decade survey of content based image retrieval using deep learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(5), pp.2687-2704.
- Fang, W., Love, P.E., Luo, H. and Xu, S., 2022. A deep learning fusion approach to retrieve images of People's unsafe behavior from construction sites. *Developments in the Built Environment*, 12, p.100085.
- Ha, I., Kim, H., Park, S. and Kim, H., 2018. Image retrieval using BIM and features from pretrained VGG network for indoor localization. *Building and Environment*, 140, pp.23-31.
- Lee, S., Masoud, M., Balaji, J., Belkasim, S., Sunderraman, R. and Moon, S.J., 2017. A survey of tag-based information retrieval. *International Journal of Multimedia Information Retrieval*, 6, pp.99-113.
- Li, N., Li, Q., Liu, Y.S., Lu, W. and Wang, W., 2020. BIMSeek++: Retrieving BIM components using similarity measurement of attributes. *Computers in Industry*, 116, p.103186.
- Li, X., Yang, J. and Ma, J., 2021. Recent developments of content-based image retrieval (CBIR). *Neurocomputing*, 452, pp.675-689.
- Ma, J.W., Czerniawski, T. and Leite, F., 2021. An application of metadata-based image retrieval system for facility management. *Advanced Engineering Informatics*, 50, p.101417.
- Papineni, K., Roukos, S., Ward, T. and Zhu, W.J., 2002, July. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics* (pp. 311-318).
- Spencer Jr, B.F., Hoskere, V. and Narazaki, Y., 2019. Advances in computer vision-based civil infrastructure inspection and monitoring. *Engineering*, 5(2), pp.199-222.
- Sun, A., Bhowmick, S.S., Nam Nguyen, K.T. and Bai, G., 2011. Tag - based social image retrieval: An empirical evaluation. *Journal of the American Society for Information Science and Technology*, 62(12), pp.2364-2381.
- Wan, J., Wang, D., Hoi, S.C.H., Wu, P., Zhu, J., Zhang, Y. and Li, J., 2014, November. Deep learning for content-based image retrieval: A comprehensive study. In *Proceedings of the 22nd ACM international conference on Multimedia* (pp. 157-166).

- Wang, Y., Xiao, B., Bouferguene, A., Al-Hussein, M. and Li, H., 2023. Content-based image retrieval for construction site images: leveraging deep learning-based object detection. *Journal of Computing in Civil Engineering*, 37(6), p.04023035.
- Wogen, B.E., Choi, J., Zhang, X., Liu, X., Iturburu, L. and Dyke, S.J., 2024. Automated Bridge Inspection Image Retrieval Based on Deep Similarity Learning and GPS. *Journal of Structural Engineering*, 150(3), p.04023238.
- Zohourianshahzadi, Z. and Kalita, J.K., 2022. Neural attention for image captioning: review of outstanding methods. *Artificial Intelligence Review*, 55(5), pp.3833-3862.

ADVANCING BLOCKCHAIN AND SMART CONTRACTS FROM ACADEMIA TO INDUSTRY: INSIGHTS INTO THE CONSTRUCTION SECTOR

Xuling Ye¹, Liu Liu², and Markus König³

1) Ph.D. Candidate, Computing in Engineering, Faculty of Civil and Environmental Engineering, Ruhr University Bochum, Bochum, Germany. Email: xuling.ye@rub.de

2) Ph.D. Candidate, Computing in Engineering, Faculty of Civil and Environmental Engineering, Ruhr University Bochum, Bochum, Germany. Email: liu.liu-m6r@rub.de

3) Dr. -Ing., Prof., Chair of Computing in Engineering, Faculty of Civil and Environmental Engineering, Ruhr University Bochum, Bochum, Germany. Email: koenig@inf.bi.rub.de

Abstract: Blockchain (BC) and smart contract (SC), known for their decentralized and secure frameworks, have been successfully applied across industries to enhance security, efficiency, and transparency in technical and engineering aspects such as automated transactions, data management, and supply chain monitoring. In recent years, the construction industry has seen significant advancements in adopting BC and SC. This review paper examines the latest BC and SC prototypes, case studies, and technical applications published over the past four years, focusing on how these academic developments can be transitioned into industry practice. By categorizing the collected studies using Technology Readiness Levels (TRLs) and conducting an in-depth analysis, the paper seeks to identify strategies for accelerating the adoption of these technologies in the construction sector. Reviewing recent advancements, this paper uncovers key trends, benefits, and challenges associated with implementing BC and SC in construction. The findings suggest that these technologies can significantly enhance project efficiency, contract management, and supply chain transparency while fostering trust among stakeholders. While these technologies offer significant potential, critical challenges, including scalability, privacy, and integration complexities, hinder their widespread adoption. Moreover, specific limitations, such as the difficulty in establishing regulatory compliance and technical barriers in data management, remain unresolved. By providing actionable insights and highlighting pathways to bridge the adoption gaps, the paper aims to accelerate the integration of BC and SC into the construction industry, driving innovation, transparency, and operational efficiency.

Keywords: Blockchain, smart contract, construction industry, technology readiness level, technology adoption, review

1. INTRODUCTION

Blockchain (BC) and smart contracts (SC) are increasingly recognized for their ability to transform industries through decentralized, secure, and transparent frameworks (Casino et al., 2019). BC, a distributed ledger technology, ensures tamper-proof data storage and process transparency, while SC automates contractual agreements using self-executing code that enforces the terms without intermediaries (Christidis & Devetsikiotis, 2016). Initially discussed as conceptual innovations, BC and SC have evolved over the past decade into tangible applications and technical prototypes. Industries such as finance and healthcare have already leveraged these technologies for enhanced security, efficiency, and transparency (Casino et al., 2019). Similarly, the construction industry—traditionally burdened with complex workflows, fragmented projects, and contractual inefficiencies—has begun to explore the potential of BC and SC over the past decade.

Several review papers have highlighted the advancements in BC and SC research specific to the construction industry. For instance, J. Li and Kassem (2021) provides a comprehensive analysis of BC and SC applications, including improving information management, enhancing transparency, and integrating with technologies like Building Information Modeling (BIM) and Internet of Things (IoT). Similarly, Rathnayake et al. (2022) focuses on SC applications for automating payment systems and contract execution, identifying strategies to overcome adoption barriers and enhance efficiency in construction projects. These studies collectively emphasize the increasing maturity of BC and SC applications in addressing the longstanding challenges of construction management. Particularly in the last five years, there has been a growing body of literature examining BC and SC applications in construction, with a noticeable shift from conceptual discussions to practical case studies and technical implementations (Sun et al., 2023). Recent research reviews the use of BC and SC for purposes such as automating contract management, streamlining payments, and enhancing supply chain transparency (B. G. Celik et al., 2024). The progression from concept to application suggests that these technologies have the potential to improve project delivery and stakeholder collaboration in construction.

However, the adoption of BC and SC technologies in construction is still in its early stages, and several barriers remain—ranging from scalability and privacy concerns to the integration with existing systems. These challenges underscore the need for a structured approach to assess the technological maturity of BC and SC innovations, which is where the Technology Readiness Levels (TRLs) framework comes into play. TRLs offer a comprehensive and objective measure of technological maturity, allowing stakeholders to gauge which innovations are ready for large-scale adoption and which require further development. Building upon our previous work (Ye

et al., 2022), which provided a comprehensive review of BC-based SC applications in construction, this study seeks to extend the understanding of these technologies by examining their readiness for industrial adoption through the lens of the TRL framework.

Therefore, this paper aims to provide a review of BC and SC technologies in construction with a particular focus on their technical implementation and to investigate how BC and SC can be adapted to the construction industry. By categorizing recent studies on BC and SC based on TRLs, this paper seeks to identify which construction sub-domains derived from Ye et al. (2022) have made the most progress in adopting BC and SC, and where further development based on the technical metrics is needed to bring these technologies into mainstream industrial use. The key research questions guiding this review are: 1) Which construction sub-domains represent the greatest potential for the adoption of BC and SC technologies? 2) What aspects of the technical metrics are at the center of the studies that have achieved high TRL? 3) How could the technical solutions be adapted to the industry considering the benefits and barriers?

To address these questions, TRL introduces a basic framework for categorizing studies in Section 2. Section 3 outlines the literature review process. Section 4 contains the descriptive analysis and in-depth analysis, offering insights into the construction industrial adoption of BC and SC. Section 5 explores the possibility of the technical solutions adaption, followed by the conclusion in Section 6.

2. TRL FOR LITERATURE ANALYSIS

To assess the development maturity of BC and SC technologies in the construction industry, the TRL framework, as defined by Mankins (1995) for NASA, provides a structured and systematic approach. This framework assesses the maturity of a technology throughout its development process. It spans nine levels, ranging from basic research and concept development (TRL 1) to fully operational and deployed technologies (TRL 9). Each level is differentiated based on the achievement of technical functions and the fulfillment of validation requirements. At the lowest level, TRL 1 focuses on observing and reporting on basic scientific research (e.g., basic BC and SC technologies), which begins the basis of designing a proof of concept and the transition to more applied research, such as its potential applications in the construction industry. Until TRL 3, the focus is on “analytical” and “experimental” approaches to prove the feasibility of a concept. From TRL 4 to TRL 6, technological components are developed and validated in controlled laboratory environments. Up to TRL 7, a system prototype is built and validated in an operational environment.

To streamline the classification of the research studies analyzed in this review, three overarching categories have been defined, derived from the overview of the TRL scale by Mankins (2009): 1) Research (TRL 1-3): This phase focuses on establishing the basic concepts of BC and SC technologies and analyzing their core ideas; 2) Development (TRL 4-6): During this phase, the BC and SC technologies are further developed, progressing from small-scale prototypes to testing in relevant operational environments; 3) Deployment (TRL 7-9): The final phase focuses on demonstrating the effectiveness of the technologies in real-world environments in preparation for full-scale system deployment. These categories allow for a simplified and systematic assignment of each research study to an appropriate TRL, thus facilitating the overall analysis of technological maturity. Even when the three overarching categories are used, effective application of TRL assessment can be challenging due to the complexity of selecting the appropriate metrics (Mankins, 2009). This is particularly true in the construction industry, which is undergoing significant digital transformation. Therefore, to determine and analyze in depth the research studies on the maturity of BC and SC technologies in this review paper, several technical aspects server as metrics are defined in the Methodology section.

3. RESEARCH METHODOLOGY

To conduct a literature review of BC and SC technical adoptions in the construction industry, the overall process of the research methodology is illustrated in Figure 1. It contains three steps: data collection, data analysis, and discussion. The design of this study borrows some aspects from the systematic review set out by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (Moher et al., 2010). Based on the flow diagram of PRISMA, Scopus was used as the search engine and source in the data collection step to retrieve literature addressing BC and SC in the construction or more broadly in the Architecture, Engineering and Construction (AEC) industry. The papers were retrieved as a preliminary process through advanced search in the article title/abstract/keywords field combining keywords: (blockchain OR “smart contract”) AND (prototype OR implement OR case) AND (construction OR AEC) for the construction industry. The data used in this review were retrieved in September 2024, and the search yielded 587 studies.

After the literature search, a series of inclusion and exclusion criteria were defined to guide the extraction of the most relevant papers. To be a follow-up study of the previous review (Ye et al., 2022), only journal articles in the last 4 years (i.e., from 2021 to 2024) focusing on technical application research are considered. Papers were excluded for the following reasons: 1) papers not available in English, 2) papers without full availability, 3) papers not as a journal article, and 4) papers not in the target fields, i.e., in the construction or AEC field. After the inclusion and exclusion process, 180 papers in construction were left. Then each of the papers was reviewed based

on three quality evaluation questions: 1) Is the paper primarily focused on BC or SC technologies, rather than just mentioning them? 2) Does the paper propose a solution/framework, rather than an opinion paper based on reviews, surveys, interviews, questionnaires, and/or focus groups? 3) Does the paper provide technical insights with applications, prototypes, or implementation details? After the quality evaluation, there were 44 papers in total.

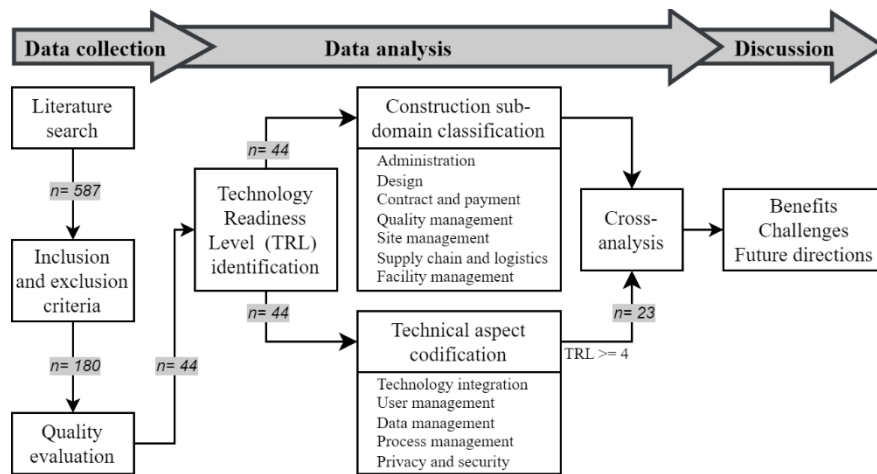


Figure 1. Overview of research methodology

In the data analysis step, the collected papers were analyzed based on three steps, namely TRL identification, construction sub-domain classification and technical aspect codification, which serves the cross-analysis. Figure 2 is drawn based on the TRL introduced in Section 2 as a guidance workflow of evaluating the TRL of BC and SC adoption readiness in each paper. It presents a logical sequence of TRLs reached and adopted that allows researchers to evaluate each BC and SC application, thus establishing its maturity.

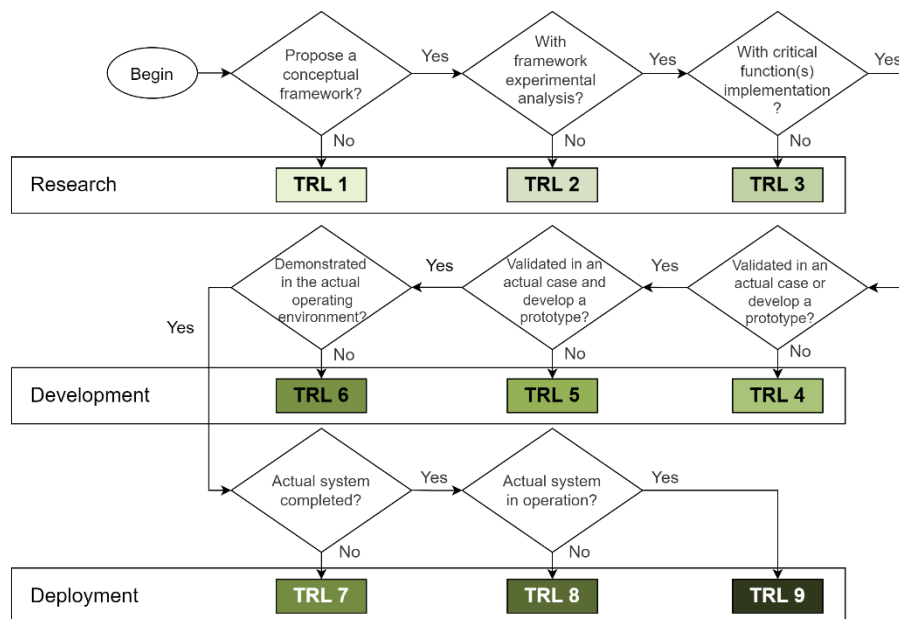


Figure 2. TRL evaluation workflow designed in this paper

As follow-up research, the classification of the construction sub-domains follows the research domain codification by Ye et al. (2022), as shown in Figure 1. To aggregate technical findings across studies, a comprehensive list of first- and second-level codes is presented in Table 1, serving as the foundation for technical aspect codification. The first-level codes are derived from frequently mentioned technical terms in the collected journal articles. These first-level codes are subsequently organized into more specific second-level codes. Through this systematic codification process, five key technical aspects for the adoption of BC and SC in the construction field have been identified from the 44 collected articles: 1) Technology Integration, 2) User Management, 3) Data Management, 4) Process Management, and 5) Privacy and Security. In the end, a comprehensive discussion can be

developed by cross-analyzing the articles from multiple perspectives, including TRLs, construction sub-domains, and technical aspects.

Table 1. Technical aspect codification

First-level code	Second-level code
Internet of Things (IoT), Building Information Modeling (BIM), fog computing, electric power transmission networks, Artificial Intelligence (AI), Virtual Reality (VR), embedded systems, edge computing, deep learning, learning systems	Technology integration
human, user, stakeholder, multiple stakeholders, worker, trust, behavioral research, accountability, access control, authentication	User management
information management, digital storage, data storage, data sharing, information sharing, cloud computing, big data, data integrity, data handling, cloud storage, data sensitivity, transparency, traceability, large-sized information, off-chain data	Data management
decision making, automation, computer-aided instruction, optimization, process model, computational modeling, Business Process Model and Notation (BPMN)	Process management
network security, cryptography, security, privacy, zero-knowledge proof, permissioned, privacy-preserving, privacy protection, computer security, public key cryptography	Privacy and security

4. CROSS-ANALYSIS AND RESULTS

4.1 Descriptive Analysis

The distribution of collected publications from 2021 to September 2024 is illustrated in Figure 3. The noticeable annual increase in the number of publications signifies ongoing technical advancements of this research in this sector. Additionally, the TRL of these publications shows a slight annual increase. The significant rise in TRLs observed in 2024 highlights the growing recognition of BC and SC's potential, their technical implementation, and their application in real-world construction case studies. This trend suggests that this research may continue to grow in the near future.

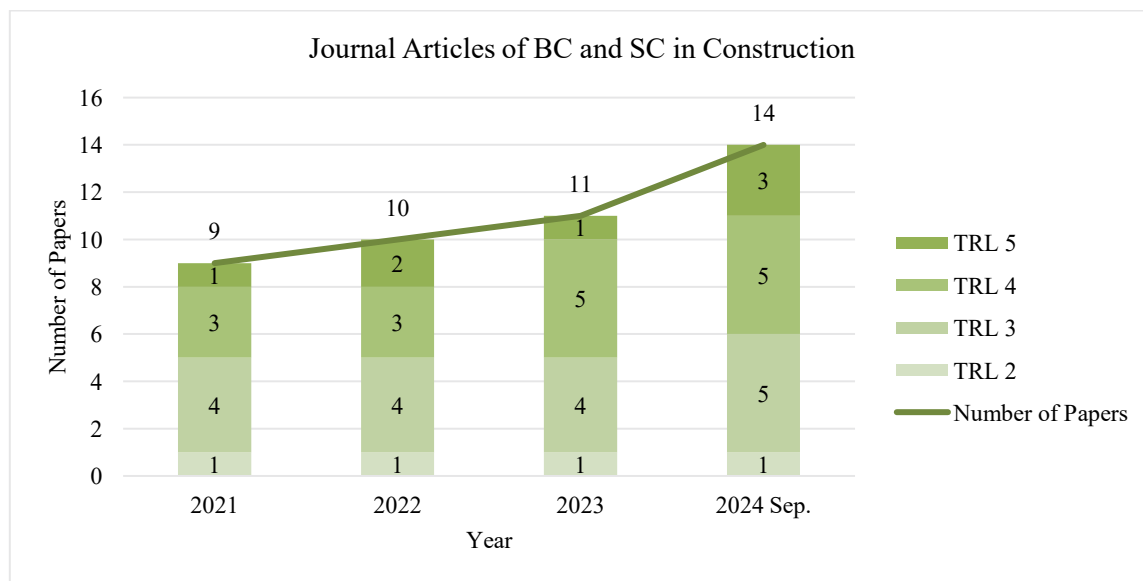


Figure 3. Year and TRL distribution of the collected journal articles in construction

Figure 4 presents the distribution of the collected publications by year, TRL, and construction sub-domain. The data highlights that supply chain, site management, payment, and contract are the most researched areas in BC and SC studies within the construction sector, particularly about technological adoption. Notably, publications on-site management are characterized by higher TRLs, indicating more advanced, ready-to-implement solutions. Both supply chain and site management stand out as the domains with the highest TRLs, reflecting their maturity for practical application. In contrast, technical advancements in facility management and administration are limited, while research on quality management remains predominantly theoretical.

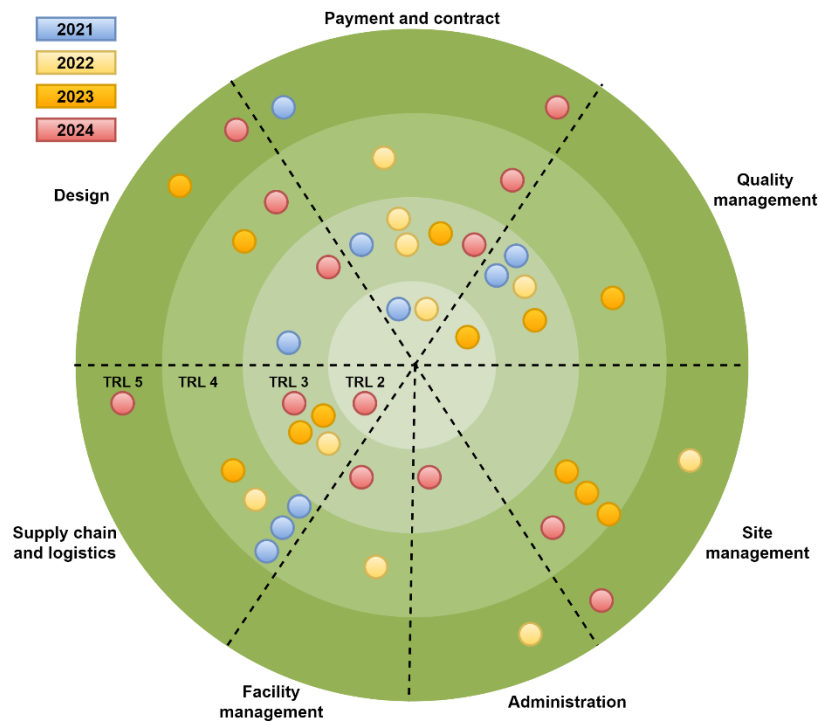


Figure 4. Year, TRL, and construction sub-domain distribution of the collected journal articles in construction

4.2 Detail analysis and results

To provide a targeted analysis of the current state of technical adoption and derive insights from academic research within the construction sector, this section focuses on the detail analyzing publication achieved at least TRL 4 to discover the technical status of this research. In total, 23 articles were analyzed (15 at TRL 4 and 8 at TRL 5), with their findings summarized in Table 2. These articles are evaluated based on their specific construction sub-domain, the selected BC platform, whether a prototype is developed or a case study is included, and the extent to which each technical aspect is thoroughly addressed.

Table 2. Comparative analysis of papers with TRL 4 and 5 according to the technical aspects

Reference	Construction sub-domain	BC platform	Case study /Prototype	Technical aspects					
				TI	UM	DM	PM	P&S	Sum
TRL 4									
Li et al. (2021)	SC&L	HLF	Both	○	◐	◐	○	○	1
Lu et al. (2021)	SC&L	HLF	Both	◐	◐	◐	○	○	1.5
Tezel et al. (2021)	SC&L	Ethereum	Prototype	○	●	○	○	●	2
Li et al. (2022)	SC&L	HLF	Both	◐	◐	◐	○	○	1.5
van Groesen and Pauwels (2022)	Facility management	Ethereum	None	◐	◐	◐	○	○	1.5
L. Wu et al. (2022)	Payment and contract	HLF	Prototype	○	◐	◐	◐	○	1.5
Y. Celik et al. (2023)	Design	Ethereum	Prototype	◐	◐	◐	○	◐	2
Chen et al. (2023)	Site management	Ethereum	Both	◐	○	◐	◐	○	1.5
Lu et al. (2023)	SC&L	HLF	Both	○	◐	○	◐	◐	1.5
Naderi et al. (2023)	Site management	Ethereum	Prototype	◐	●	◐	○	○	2
H. Wu et al. (2023)	Quality management	HLF	Prototype	◐	◐	◐	◐	○	2
Xiao et al. (2023)	Site management	Ethereum	None	◐	◐	◐	○	○	1.5
Bahnas et al. (2024)	Site management	Ethereum	Prototype	◐	○	◐	◐	○	1.5
Liu et al. (2024)	Design	HLF	Prototype	◐	◐	◐	◐	○	2

Pham et al. (2024)	Payment and contract	Ethereum	Case study	●	●	●	●	○	2
Percentages of technical aspects in TRL 4 (%)				36.67	50.00	43.33	23.33	13.33	33.33
TRL 5									
Sigalov et al. (2021)	Payment and contract	GoQuorum	Both	●	●	●	○	●	2.5
Adel et al. (2022)	Administration	IBM BC	Prototype	●	●	●	○	○	2.5
Yang et al. (2022)	Site management	HLF	Case study	●	●	●	○	●	3
Tao et al. (2023)	Design	HLF	Prototype	●	●	●	○	●	2.5
Bao et al. (2024)	Site management	Ethereum	Prototype	●	●	●	●	○	3
Ding et al. (2024)	SC&L	HLF	Prototype	●	●	●	○	●	2.5
Lu and Wu (2024)	Design	HLF	Both	●	●	●	○	●	2.5
Ye et al. (2024)	Payment and contract	Ethereum	Prototype	●	●	●	●	○	3.5
Percentages of technical aspects in TRL 5 (%)				68.75	56.25	93.75	18.75	37.50	55.00

SC&L = Supply chain and logistics; BC = Blockchain; HLF = Hyperledger Fabric; TI = Technology integration, UM = User management, DM = Data management, PM = Process management, P&S = Privacy and security; ○ = Not addressed, ● = Theoretically/Partially addressed, ● = Technically/Fully addressed.

As shown in Table 2, for the TRL-4 articles, 5 out of 15 focus on “Supply Chain and Logistics”, while 4 are centered on “Site Management”. In the TRL 5 papers, 2 out of 8 cover topics related to “Design”, “Site Management”, and “Payment and Contract”, respectively. Regarding BC platforms, Hyperledger Fabric (HLF) and Ethereum are the most frequently used. Specifically, all the TRL-4 articles use either HLF or Ethereum. In TRL 5, 50% of articles use HLF, and 25% use Ethereum. Other BC platforms, such as GoQuorum and IBM Blockchain, are occasionally used in TRL 5 studies, indicating a shift toward broader platform exploration at higher TRLs. When evaluating the technical aspects of each article, a full solid circle (●) represents that the target technical aspect is fully addressed in the article (counted as 1), a half solid circle (◐) denotes partial or theoretical consideration (counted as 0.5), and an empty circle (○) means the aspect is not addressed (count as 0). Papers at TRL 4 typically score between 1 and 2 across all five technical aspects, while those at TRL 5 score between 2.5 and 3.5. In terms of practical implementation, 12 of the 15 TRL-4 articles have developed a prototype, and 6 have tested their framework in a use case. Only two TRL-4 articles neither developed a prototype nor tested their frameworks. At TRL 5, 7 out of 8 papers have developed a prototype, with 3 having tested their frameworks in real-world case studies. According to the percentages of technical aspects, most of the aspects in TRL 5 are higher than TRL 4, except for the process management aspect. In the following sub-sections, articles that fully address the technical aspects are analyzed in detail aspect by aspect to assess the current state of technical development and to extract key insights from their findings.

4.2.1 Technology Integration

Technology integration is a critical aspect of advancing the construction sector through the adoption of BC and SC. In TRL 4, numerous articles have explored the potential of integrating BC and SC with other advanced technologies such as BIM, IoT, Artificial Intelligence (AI), and digital twins. However, it is noteworthy that none of these articles have provided technical or practical validation of their proposed concepts. This indicates a gap between theoretical exploration and practical implementation at this TRL, highlighting the need for more empirical research and real-world applications.

Conversely, in TRL 5, the discussion around integrating BC and SC with other technologies is more advanced, with seven studies fully addressing the aspect. For instance, Adel et al. (2022) proposed a blockchain-based chatbot system for information management in construction firms, specifically targeting work progress tracking in construction projects. The chatbot leverages IBM Watson Assistant, and the blockchain network is built using the IBM Blockchain Platform, while the connection between the chatbot and the network is via IBM Cloud Functions and IBM Cloudant Databases. This solution reduces the technical difficulty of using blockchain for construction stakeholders. Bao et al. (2024) proposed a token-based incentive framework for Virtual Reality (VR) safety training and tested it via prototype development (uploading the VR model to the cloud, and safety scenario metadata to the blockchain). Ye et al. (2024) developed a generation tool that converts Business Process Model and Notation (BPMN) to Solidity, which is open-sourced on GitHub (2024c) and implemented a decentralized application (DApp) to visualize a real-time payment process via a BPMN diagram. These examples demonstrate the practical applications and benefits of integrating blockchain with other technologies in the construction sector, paving the way for more innovative solutions.

4.2.2 User management

User management refers to the control of user roles, user collaboration, permissions, and access in BC and SC systems. It is frequently discussed in the collected articles, although it often lacks technical depth. At TRL 4, Tezel et al. (2021) proposed three BC and SC-based models for supply chain management, which are project bank accounts (PBAs) for payments, reverse auction-based tendering for bidding, and asset tokenization for project financing with role configuration and access control. The source code of these three models is available as open source on GitHub (2024b). Naderi et al. (2023) proposed a method to combine computer vision, InterPlanetary File System (IPFS) with BC and SC, and tested the user management component through the SafetyCredit DApp, also open-sourced on GitHub (2024a). At TRL 5, Ye et al. (2024) advanced user management by integrating BPMN with SC to control the user role for task execution with source code available on GitHub (2024c). These advancements underline the importance of secure and scalable user control mechanisms in BC and SC-based applications within the construction industry.

4.2.3 Data management

Data management involves the organization, storage, and handling of data within blockchain networks, including on- and off-chain strategies. In the reviewed papers, it is primarily addressed through two approaches: 1) on- and off-chain data handling, and 2) BC network configuration and SC coding for data management. At TRL 4, Adel et al. (2022) proposed a chatbot system that integrates on-chain and off-chain data management for tracking construction progress, simplifying BC and SC adoption for construction stakeholders. The chatbot is built using IBM Watson Assistant, and the blockchain network is constructed using the IBM Blockchain Platform. This solution simplifies the technical complexities associated with blockchain adoption for construction stakeholders.

In TRL 5, most of the studies have provided technical insights into the data management aspect. The BIMcontracts approach proposed by Sigalov et al. (2021) integrates blockchain and smart contracts with Common Data Environment (CDE) and payment services for payment automation and contract management. In this approach, the smart contract is initialized firstly by a billing container using the Information Container for Linked Document Delivery (ICDD) technology, then configured in JSON format, and deployed to the blockchain. Billing information is securely stored on the blockchain, enhancing payment accuracy and transparency. Tao et al. (2023) manage to store a BIM model fingerprint, user ID, and process status on the HLF blockchain to handle BIM data sharing and issue management by implementing corresponding smart contracts. Bao et al. (2024) proposed a token-based incentive framework for VR safety training and tested it via a prototype development with a configuration of Avalanche primary network and subnet for different data storage. Ding et al. (2024) focused on information sharing by channel controlling and data access control via smart contracts in HLF blockchain and IPFS. Lu and Wu (2024) proposed a blockchain-based framework to protect building design intellectual property rights in the AEC industry. Each BIM design version is hashed as a Non-Fungible Token (NFT), with data stored in CouchDB and IPFS. These approaches demonstrate innovative solutions for secure and efficient data management in the construction sector.

4.2.4 Process management

Process management refers to the coordination, monitoring, and execution of tasks and workflows in construction projects, particularly how they are governed through blockchain and smart contracts. It remains a challenging but vital aspect of the construction sector. At TRL 4, seven papers address the topic theoretically, but none provide significant technical insights, signaling a substantial gap in practical process management solutions at this stage. This gap highlights the need for more research and development to bring theory into practice in process management. At TRL 5, only two papers address process management, with just one offering technical insights. Ye et al. (2024) developed a solution for real-time process monitoring by integrating BPMN with smart contracts, providing a practical method for managing complex construction processes in real time. Process management remains a challenging topic in construction due to the complexity and uniqueness of each project. Future research could focus on developing more robust and practical solutions for process management in the construction industry.

4.2.5 Privacy and security

Privacy and security in blockchain systems for construction focus on enhancing data protection (both on-chain and off-chain) and ensuring secure access control to different sensitive levels of construction data. Although less frequently discussed than data management in general, each TRL (4 and 5) includes at least one paper with technical insights in this area. In TRL 4, Tezel et al. (2021) proposed models for supply chain management that incorporate role configuration and access control to enhance data privacy and security. These models are available as open-source resources on GitHub (2024b), encouraging further exploration and development. In TRL 5, Yang et al. (2022) developed a JavaScript and Chaincode-based solution to verify scaffolding work by comparing parameters such as location and quantity throughout various stages—ordering, procurement, and installation. This approach enhances the accuracy and security of scaffolding processes, ensuring the reliability of construction projects. These advancements underscore the importance of robust privacy and security measures in BC and SC-based

construction applications, ensuring the system access control and the integrity of construction processes.

5. DISCUSSION

The reviewed articles show diverse applications of BC and SC at different TRLs. Based on the analysis and results, the three research questions proposed in the introduction can be answered. The construction sub-domains “Supply chain and logistics” and “Payment and contract” showcase the greatest potential for the adoption of BC and SC technologies. At TRL 4, most studies focus on supply chain and logistics or site management, while TRL 5 broadens to include design and payment management. HLF and Ethereum are the dominant platforms in TRL 4, but TRL 5 studies explore other platforms like GoQuorum and IBM Blockchain, reflecting increased technical diversity. TRL 5 papers generally address technical aspects more thoroughly, scoring higher in areas like data handling and process management. While TRL 4 studies remain more theoretical, both levels demonstrate practical progress, with most TRL 5 studies developing prototypes and testing them in real-world scenarios. This shift from theory to practice highlights the growing maturity of BC applications in construction. The integration of BC and SC presents numerous advantages, such as increased trust between parties, streamlined payment processes, and improved supply chain management. BC and SC can enhance contract execution by automating tasks and enforcing compliance through coded agreements, reducing disputes and delays. Additionally, these technologies improve data integrity and transparency across project lifecycles, enabling better decision-making and reducing inefficiencies. The integration of BC and SC presents numerous advantages, such as increased trust between parties, streamlined payment processes, and improved supply chain management. BC and SC can enhance contract execution by automating tasks and enforcing compliance through coded agreements, reducing disputes and delays. Additionally, these technologies improve data integrity and transparency across project lifecycles, enabling better decision-making and reducing inefficiencies.

Despite the potential benefits, there are several barriers to the widespread adoption of BC and SC in the construction industry. A key issue is the lack of sufficient technical insights in many studies. For instance, while many studies offer comprehensive solutions addressing various construction engineering aspects, a key limitation remains the lack of focus on fundamental challenges such as further improving privacy and security to fit construction practices in BC and SC applications. This issue is particularly critical for high-potential areas like the construction sub-domain “Payment and contract”, where secure data handling (especially for off-chain data) is essential. Addressing this gap requires advanced development in BC and SC technology to ensure robust privacy and security measures are in place, enhancing their reliability and adoption in the construction industry. While BC and SC show great promise, there is still significant room for improvement in the depth of technical understanding and practical applications. Current solutions tend to lack realism and are not sufficiently developed to address the industry’s specific needs, limiting their practical utility. The adoption of BC and SC in the construction industry is further hindered by industry fragmentation, insufficient digital infrastructure, and resistance to technological change. This fragmentation limits collaboration among industry players, making it challenging to develop standards and share valuable information that could facilitate widespread implementation. Achieving standardization across the construction sector is crucial for the scalable adoption of these technologies. Additionally, transitioning from academic research to practical industry applications presents significant challenges, including the need for regulatory alignment and the development of industry-specific, scalable solutions.

Looking forward, the integration of BC and SC with emerging technologies such as AI, IoT, and digital twins is anticipated to play a pivotal role in overcoming these barriers. These integrations could enable advanced predictive analytics, enhance real-time data sharing, and improve operational efficiencies across construction projects. Future trends also suggest a growing focus on decentralization, with applications in autonomous supply chain operations and enhanced stakeholder collaboration through distributed decision-making frameworks. Furthermore, as the construction industry increasingly adopts green building practices, blockchain could support sustainability efforts by offering transparent systems for tracking carbon footprints and verifying eco-friendly materials. Process management, an area where technical insights are currently limited due to the complexity and uniqueness of construction workflows, represents another promising avenue for future research and development. Addressing these challenges while capitalizing on emerging trends will be essential for unlocking the full potential of BC and SC in the construction sector. While BC and SC offer significant potential for transforming the construction industry, especially for supply chain, payment, and contracts, there is a clear need for deeper technical exploration, improved collaboration, and practical solutions tailored to industry challenges. Addressing these barriers will be critical for enabling the successful adoption of these technologies in the construction sector.

6. CONCLUSIONS

This paper provides a comprehensive review of blockchain and smart contract (BC and SC) implementations in the construction industry, exploring their potential benefits and challenges. By analyzing 44 journal articles published between 2021 and 2024, this study identifies key technical aspects essential for BC and SC adoption, including technology integration, user management, data management, process management, and privacy and se-

curity. The findings reveal that while BC and SC hold significant promise for transforming the construction industry, much of the research remains in early stages, with many solutions still conceptual or experimental. Advanced technological maturity is evident in areas like supply chain, site management, and payment systems, but other domains such as quality and facility management remain underexplored. This indicates a need for more targeted, industry-specific solutions that address construction's unique challenges, including process complexity and data security.

Future research should focus on bridging the gap between theoretical models and real-world applications of BC and SC in the construction industry. One critical area is the further development of practical, industry-specific solutions that can address issues like the complexity of construction processes and off-chain data security. More detailed investigations into process management could help translate theoretical advancements into scalable, implementable systems. Furthermore, exploring cross-sector collaborations to develop unified standards and protocols for BC and SC in construction will help promote broader industry adoption. These efforts should focus on creating interoperable systems that can streamline operations across diverse project environments. Another promising avenue for future development is incorporating emerging technologies like BIM, IoT, and AI with BC and SC. This integration can improve automation, transparency, and efficiency in construction processes. However, challenges around scalability, standardization, and implementation must be addressed before these technologies can be widely adopted. While the potential of these integrations is often overstated, a strategic combination of BC, SC, and these tools offers the possibility of more streamlined, secure, and data-driven construction workflows. Advancing privacy and security measures within BC and SC systems is crucial, especially as construction projects involve large-scale data with complex hierarchy and multiple stakeholders. Addressing these issues will enable the development of robust, practical applications that move beyond theoretical discussions, positioning BC and SC as key enablers of digital transformation in the construction sector.

REFERENCES

- Adel, K., Elhakeem, A., and Marzouk, M. (2022). Chatbot for construction firms using scalable blockchain network. *Automation in Construction*, 141, 104390.
- Bahnas, N., Adel, K., Khallaf, R., and Elhakeem, A. (2024). Monitoring and controlling engineering projects with blockchain-based critical chain project management. *Automation in Construction*, 165, 105484.
- Bao, Q. L., Tran, S. V.-T., Yang, J., Pedro, A., Pham, H. C., and Park, C. (2024). Token incentive framework for virtual-reality-based construction safety training. *Automation in Construction*, 158, 105167.
- Casino, F., Dasaklis, T. K., and Patsakis, C. (2019). A systematic literature review of blockchain-based applications: Current status, classification and open issues. *Telematics and Informatics*, 36, 55–81.
- Celik, B. G., Abraham, Y. S., and Attaran, M. (2024). Unlocking Blockchain in Construction: A Systematic Review of Applications and Barriers. *Buildings*, 14(6), 1600.
- Celik, Y., Petri, I., and Barati, M. (2023). Blockchain supported BIM data provenance for construction projects. *Computers in Industry*, 144, 103768.
- Chen, G., Liu, M., Zhang, Y., Wang, Z., Hsiang, S. M., and He, C. (2023). Using Images to Detect, Plan, Analyze, and Coordinate a Smart Contract in Construction. *Journal of Management in Engineering*, 39(2), 4023002.
- Christidis, K., and Devetsikiotis, M. (2016). Blockchains and Smart Contracts for the Internet of Things. *IEEE Access*, 4, 2292–2303.
- Ding, S., Hu, H., Xu, F., Chai, Z., and Wang, W. (2024). Blockchain-based security-minded information-sharing in precast construction supply chain management with scalability, efficiency and privacy improvements. *Automation in Construction*, 168, 105698.
- GitHub. (2024a). *h-naderi/SafetyCredit-Dapp: A decentralized application aiming to boost construction safety through rewarding tokens based on safety helmet use in construction sites*. Retrieved from GitHub website: <https://github.com/h-naderi/SafetyCredit-Dapp>
- GitHub. (2024b). *huddersfield-uni-smart-contracts*. Retrieved from GitHub website: <https://github.com/huddersfield-uni-smart-contracts>
- GitHub. (2024c). *XulingYe/SmartProcess_SmartContractGenerator: Generate Solidity Smart Contracts from YAWL and BPMN*. Retrieved from GitHub website: https://github.com/XulingYe/SmartProcess_SmartContractGenerator
- Li, J., and Kassem, M. (2021). Applications of distributed ledger technology (DLT) and Blockchain-enabled smart contracts in construction. *Automation in Construction*, 132, 103955.
- Li, X., Lu, W., Xue, F., Wu, L., Zhao, R., Lou, J., and Xu, J. (2022). Blockchain-Enabled IoT-BIM Plat-

- form for Supply Chain Management in Modular Construction. *Journal of Construction Engineering and Management*, 148(2), 4021195.
- Li, X., Wu, L., Zhao, R., Lu, W., and Xue, F. (2021). Two-layer Adaptive Blockchain-based Supervision model for off-site modular housing production. *Computers in Industry*, 128.
- Liu, Y., Tao, X., Das, M., Gong, X., Liu, H., Xu, Y., Xie, A., and Cheng, J. C. (2024). Blockchain-enabled platform-as-a-service for production management in off-site construction design using openBIM standards. *Automation in Construction*, 164.
- Lu, W., Li, X., Xue, F., Zhao, R., Wu, L., and Yeh, A. (2021). Exploring smart construction objects as blockchain oracles in construction supply chain management. *Automation in Construction*, 129.
- Lu, W., and Wu, L. (2024). A blockchain-based deployment framework for protecting building design intellectual property rights in collaborative digital environments. *Computers in Industry*, 159-160, 104098.
- Lu, W., Wu, L., and Zhao, R. (2023). Rebuilding trust in the construction industry: a blockchain-based deployment framework. *International Journal of Construction Management*, 23(8), 1405–1416.
- Mankins, J. C. (1995). Technology readiness levels. Retrieved from Mankins, John C. website: https://aiaa.kavi.com/apps/group_public/download.php/2212/trls_mankins%20paper_1995.pdf
- Moher, D., Liberati, A., Tetzlaff, J., and Altman, D. G. (2010). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *International Journal of Surgery (London, England)*, 8(5), 336–341.
- Naderi, H., Shojaei, A., and Ly, R. (2023). Autonomous construction safety incentive mechanism using blockchain-enabled tokens and vision-based techniques. *Automation in Construction*, 153, 104959.
- Pham, V., Vo, T. T., and Dang, N. (2024). Applying blockchain technology in smart contracts for construction payment: a comprehensive solution for lumpsum contracts. *Asian Journal of Civil Engineering*, 25(4), 3549–3564.
- Rathnayake, I., Wedawatta, G., and Tezel, A. (2022). Smart Contracts in the Construction Industry: A Systematic Review. *Buildings*, 12(12), 2082.
- Sigalov, K., Ye, X., König, M., Hagedorn, P., Blum, F., Severin, B., Hettmer, M., Hückinghaus, P., Wölcklerling, J., and Groß, D. (2021). Automated payment and contract management in the construction industry by integrating building information modeling and blockchain-based smart contracts. *Applied Sciences (Switzerland)*, 11(16).
- Sun, W., Antwi-Afari, M. F., Mehmood, I., Anwer, S., and Umer, W. (2023). Critical success factors for implementing blockchain technology in construction. *Automation in Construction*, 156, 105135.
- Tao, X., Das, M., Zheng, C., Liu, Y., Wong, P. K.-Y., Xu, Y., Liu, H., Gong, X., and Cheng, J. C. P. (2023). Enhancing BIM security in emergency construction projects using lightweight blockchain-as-a-service. *Automation in Construction*, 150, 104846.
- Tezel, A., Febrero, P., Papadonikolaki, E., and Yitmen, I. (2021). Insights into Blockchain Implementation in Construction: Models for Supply Chain Management. *Journal of Management in Engineering*, 37(4), 4021038.
- van Groesen, W., and Pauwels, P. (2022). Tracking prefabricated assets and compliance using quick response (QR) codes, blockchain and smart contract technology. *Automation in Construction*, 141.
- Wu, H., Li, H., Luo, X., and Jiang, S. (2023). Blockchain-Based Onsite Activity Management for Smart Construction Process Quality Traceability. *IEEE Internet of Things Journal*, 10(24), 21554–21565.
- Wu, L., Lu, W., and Xu, J. (2022). Blockchain-based smart contract for smart payment in construction: A focus on the payment freezing and disbursement cycle. *Frontiers of Engineering Management*, 1–19.
- Xiao, J., Zhang, W., and Zhong, R. Y. (2023). Blockchain-enabled cyber-physical system for construction site management: A pilot implementation. *Advanced Engineering Informatics*, 57, 102102.
- Yang, J., Lee, D., Baek, C., Park, C., and Lan, B. Q. (2022). Leveraging Blockchain for Scaffolding Work Management in Construction. *IEEE Access*, 10, 39220–39238.
- Ye, X., Zeng, N., and König, M. (2022). Systematic literature review on smart contracts in the construction industry: Potentials, benefits, and challenges. *Frontiers of Engineering Management*, 9(2), 196–213.
- Ye, X., Zeng, N., Tao, X., Han, D., and König, M. (2024). Smart Contract Generation and Visualization for Construction Business Process Collaboration and Automation: Upgraded Workflow Engine. *Journal of Computing in Civil Engineering*, 38(6), 4024030.

A Study on the Performance Comparison of Sediment Movement Detectors Using Unet-based Models and the Sliding Partitioning Method

Xiaosong Liu¹, Makoto Oda², Kei Kawamura³ and Tsuyoshi Wakatsuki⁴

1) Graduate School of Sciences and Technology for Innovation, Yamaguchi University, Yamaguchi, Japan. Email: e115vgv@yamaguchi-u.ac.jp

2) Graduate School of Sciences and Technology for Innovation, Yamaguchi University, Yamaguchi, Japan. Email: e092vgv@yamaguchi-u.ac.jp

3) D. Eng., Prof., Graduate School of Sciences and Technology for Innovation, Yamaguchi University, Yamaguchi, Japan. Email: kay@yamaguchi-u.ac.jp

4) Principal Investigator, the Storm, Flood and Landslide Research Division, National Research Institute for Earth Science and Disaster Resilience (NIED), Japan. Email: waka@bosai.go.jp

Abstract: Japan is covered by mountainous and hilly terrain and thus is prone to disasters such as landslides caused by earthquakes and heavy rainfall. Currently, the identification of sediment disaster areas is conducted through visual interpretation of aerial photographs taken after these disasters. To improve the efficiency of this process, research has been advancing in the automatic detection of sediment movement areas using deep learning techniques applied to aerial photographs. In this study, we compare the performance of sediment movement detectors using Unet and Unet++, in order to investigate the impact of changes in the loss function parameters during training on detection performance. Additionally, we evaluate the performance improvement when the sliding partitioning method is used for input images to the detectors. Results show that using sliding partitioning and adjusting the value of the loss function coefficient β can improve detection performance.

Keywords: Debris flow, Deep learning, Image segmentation, Unet, Unet++.

1. INTRODUCTION

Japan frequently experiences earthquakes, typhoons, and heavy rains, which increase the likelihood of sediment-related disasters such as debris flows and landslides. To analyze and develop countermeasures for such sediment disasters, sediment movement distribution maps are created by visually interpreting sediment movement areas based on aerial photographs taken after disasters, allowing for a broad understanding of slope movement (Yang Erhong et al., 2007). However, creating these sediment movement distribution maps is time-consuming and labor-intensive. Due to these challenges, previous researches of this study into the automatic detection of sediment movement areas is currently being conducted, with a particular focus on applying various deep learning framework which have rapidly advanced in recent years (Liu et al., 2015; Tadatakashi ISHII et al., 2020). We have also been conducting research on the automatic detection of sediment movement areas from aerial photographs, including the construction of detectors using the learning models Unet and Unet++, and investigating the impact of adjusting loss function parameters during training on detection performance. Furthermore, they have proposed a method that uses slide-split images as input during inference. However, since these studies were conducted separately, this study aims to analyze the impact on detection performance when these methods are combined to better clarify the characteristics of the detectors.

2. Proposed Method in Comparison to Existing Research

2.1 Sediment Movement Detection Process

In existing detection processes, the system receives an image (aerial photograph) of the area where sediment movement is detected and sediment movement is calculated by the sediment movement detector. Our study, however, leverages Unet and its improved model Unet++ as detectors while adjusting the loss function parameters during training to investigate the effect on relative detection performance (Kei KAWAMURA et al., 2023). We then verified the impact on detection performance when using slide-split images for the input image to the detector by applying either "AND" or "OR" processing to the output images (Kei KAWAMURA et al., 2018).

2.2 Detector Model

(1) Features of Unet

Unet is a network developed for pixel-level region detection in images that employs a feature detector algorithm capable of learning the identifying characteristics present of the images. Using this feature detector, it can detect the image characteristics from novel images. The frameworks of Unet is as shown in Figure 1 (Ronneberger et al., 2015).

(2) Features of Unet++

Unet++ incorporates multiple decoders, forming four branch networks which are used to detect sediment

movement. During training, errors from the four branch networks are calculated and averaged along with the errors from the encoder. This averaged value is then used as the error to train the neural networks that constitute the encoders. The frameworks of Unet++ is as shown in Figure 2(Zhou et al., 2018).

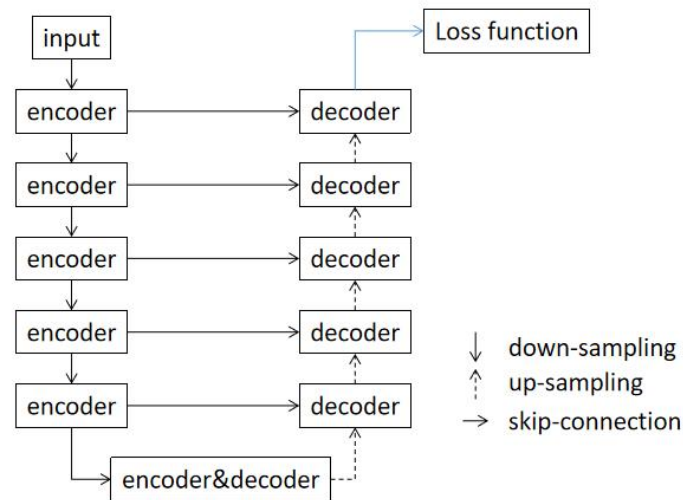


Figure 1. structure of Unet

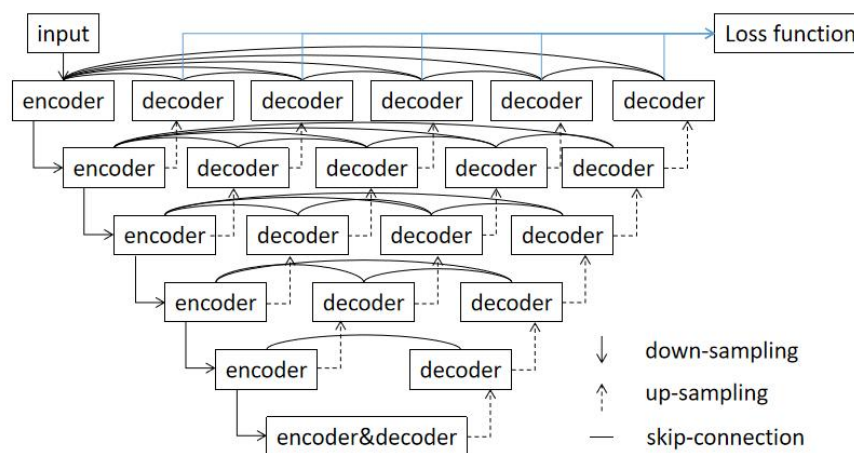


Figure 2. structure of Unet++

2.3 Processing of Input Images for the Detector

The size of the input images for the Unet and Unet++ detector models in this study is 224×224 pixels. Therefore, the training images or target region images for sediment detection (test images, etc.) must be divided into 224×224 pixel sections from the top-left to the bottom-right. In this study, any leftover regions smaller than 224×224 pixels that may occur on the right or bottom edges after dividing the aerial photographs were not considered as target areas for detection (see Figure 3 below). This division can be achieved using one of the following methods:

(1) Simple Division Method

The simple division method divides a large target image into predefined image sizes from the top-left to the bottom-right without overlapping regions for sediment detection using aerial photographs.

(2) Slide-Split(overlapping) Method

The slide-split method divides the target image into predefined image sizes from the top-left to the bottom-right while sliding it horizontally and vertically by a certain width during the division process. However, the slide width is set smaller than the width and height of the input image in order to better analyze the division boundaries. In this experiment, the slide width was set to 112 pixels. The advantage of the slide-split method is that adjacent split images have overlapping regions, resulting in more accurate predictions.

2.4 Combining Output Images from the Detector

When combining slide-split images, two methods were compared for handling the overlapping regions: OR/AND processing. In OR processing, when any pixel in the overlapping region is judged as sediment in the split images, the result is output as sediment for that pixel. With AND processing, it is only output as sediment if the pixel is judged as sediment in all overlapping split images.

2.5 Training Process

For the training of the detector, we used Equation (1) as the loss function (equivalent to the Dice coefficient). Additionally, in this study, we analyzed the effect of changing the loss function coefficient β on detection performance.

$$F_{\beta} = (1 + \beta^2) \frac{\text{recall} * \text{precision}}{\beta^2 * \text{recall} + \text{precision}} \quad (1)$$

Here,

$$\text{recall} = \frac{TP}{TP + FN} \quad (2)$$

$$\text{precision} = \frac{TP}{TP + FP} \quad (3)$$

The symbols in the equation refer to recall (true positive rate), precision, true positives (TP), false negatives (FN), and false positives (FP) (Hand, D. & Christen, P., 2018).

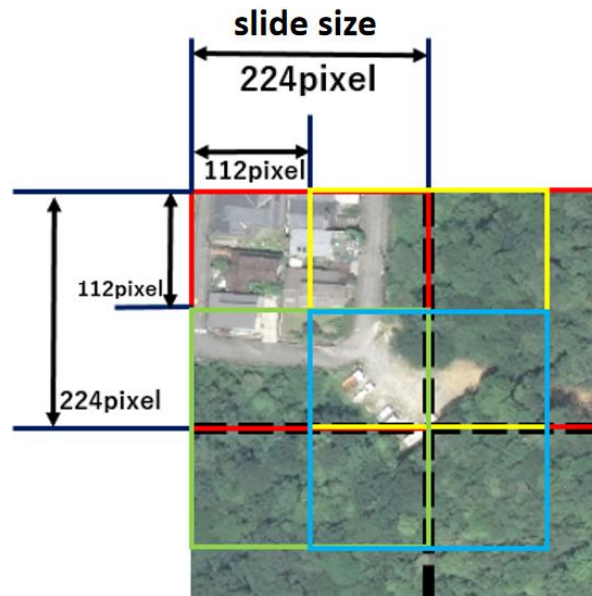


Figure 3. The simple segmentation method and the sliding segmentation method

3. Automatic Sediment Detection Methods and Performance Comparison Experiments

3.1 Experiment Flow

In this study, aerial photographs are input into the classifier using two image division methods: a simple division and the slide division method, which will be explained later. For the classifier, we used Unet and its improved model, Unet++, aiming to improve classifier performance. Additionally, by adjusting the loss function, we examined how to balance recall and precision better than in previous studies. We further performed performance comparisons using three result images: one using the simple division method combined output, and two others using the slide division method combined with "AND" and "OR" logic.

Three experiments in total were conducted using aerial photographs taken in Hiroshima after the 2018 sediment disaster and targets interpreted by workers. The first experiment focused on changing the value of the loss function's β parameter and assessing the impact of this change on the F-score using Unet and Unet++ (referred to as Experiment 1). The second experiment compared the evaluation values of the conventional simple division method and the newly introduced slide division method for the test images (referred to as Experiment 2). The third experiment examined the effect on recall, precision, and F-score when using the β value obtained from Experiment 1 in conjunction with the division methods from Experiment 2 (referred to as Experiment 3).

3.2 Experiment Method

This experiment used orthophotos of Hiroshima Sakacho Area (taken on July 9 and 11) after the 2018 heavy rainfall in Western Japan as its primary source data. To create the training data, we selected twelve pairs

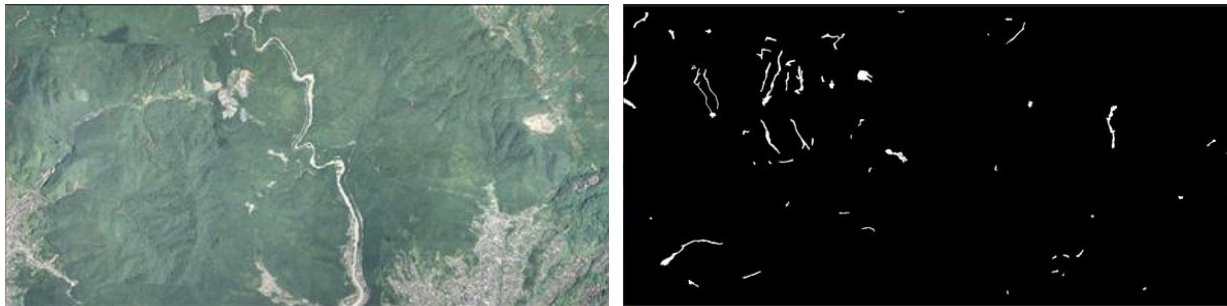
of regions (Training 1-12) from the aerial photos taken after the sediment disaster where we had interpreted the sediment movement areas by visual judgment, which were subsequently divided into 244×244 pixel (px) sections. One aerial photograph covers an area of $12,560 \times 6,087$ px, approximately 7.4×3.6 km. The actual size of each divided image is about $82\text{m} \times 82\text{m}$. The targets are of the same size as the aerial photographs. In total, 9,264 pairs of training data were created, of which 80% (7,408 pairs) were used for training while the remaining 20% (1,856 pairs) were used for validation.

Following this, the created data was input into the network for training, in which one training cycle on the entire data set was considered a single epoch. At the end of each epoch, the weights of the trained network were saved, and prediction accuracy was calculated using both the training and validation datasets based on the error function. The epoch with the highest prediction accuracy for the validation data set was then identified, and its weights were used for the classifier. In this study, the Dice coefficient was used as a measure of prediction accuracy. The Dice-score, which measures the similarity between two sets, is determined using the following formula: if the two sets are identical, the value is 1, and if they are completely different, the value is 0.

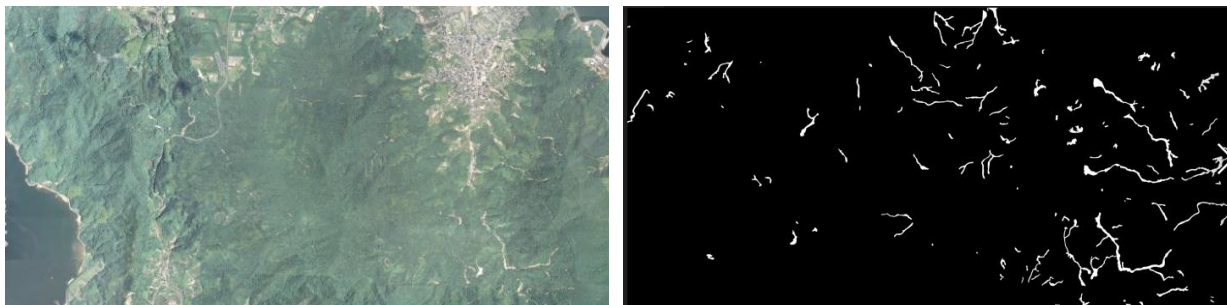
$$(\text{Dice} - \text{score}) = \frac{2 \times |X \cap Y|}{|X| + |Y|} \quad (4)$$

Here, X is the predicted data, and Y is the target data. To evaluate the generalization ability of the sediment classifier, we used four test regions (Test 1 to Test 4) as shown in Figure 4, which located at different sites within the same areas as the training regions. The size of each image is the same as the training data, $12,560 \times 6,087$ px, covering an area of approximately $3.6 \text{ km} \times 7.4 \text{ km}$. The top of the image corresponds to the north, the bottom to the south, the right side to the east, and the left side to the west. These images are from regions other than the twelve pairs used in the training data. Furthermore, the test data was not used during the training phase. The four test datasets were created by dividing these four test images using the method described earlier. The characteristics of each test region are as follows: Test 1 contains a large river in the center of the image; Test 2 features an urban area in the upper right corner; Test 3 has urban areas in the lower right and upper left corners; Test 4 contains many sediment movement areas in a mountainous region.

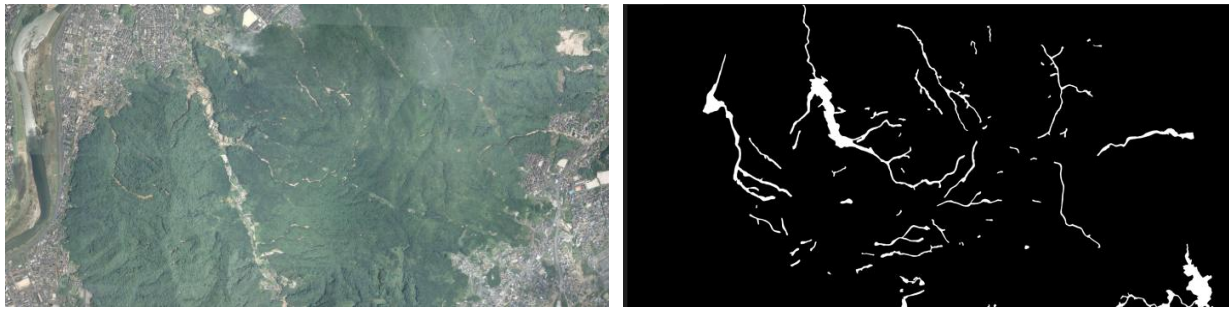
The characteristics of the four test regions are as follows: Test 1 has a large river in the center of the image, Test 2 features an urban area in the upper right corner, Test 3 has urban areas in the lower right and upper left corners, and Test 4 is a mountainous area with many sediment movement sections.



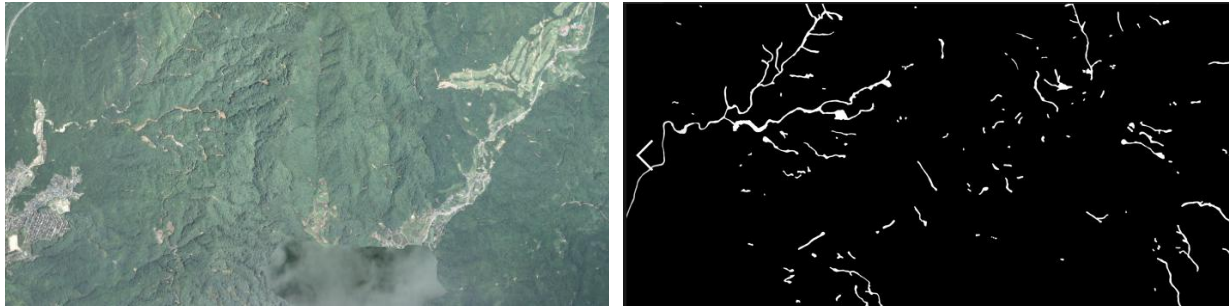
(a) Aerial photograph and target in Test 1.



(b) Aerial photograph and target in Test 2.



(c) Aerial photograph and target in Test 3.



(d) Aerial photograph and target in Test 4.

Figure 4. Aerial photograph and target in the test data.

3.3 Evaluation Metrics

To evaluate the detection results against the targets, each pixel is classified into one of four categories: True Positive (TP), False Positive (FP), False Negative (FN), or True Negative (TN), as shown in the table. Then, recall, precision, and the weighted F-score are calculated using the formulas below and are used for evaluation.

$$recall = \frac{TP}{TP+FN} \quad (5)$$

$$precision = \frac{TP}{TP+FP} \quad (6)$$

$$F_{\beta} - score = (1 + \beta^2) \frac{recall * precision}{\beta^2 * recall + precision} \quad (7)$$

Here, “recall” refers to the percentage of sediment areas in the target that were correctly identified as sediment by deep learning. Precision refers to the percentage of all areas identified as sediment by deep learning that were actual sediment areas. The F-score is a performance metric calculated based on recall and precision. The best value for each metric is 1, and higher values indicate better accuracy. Table 1 shows the experimental conditions.

Table 1. experimental conditions.

Network Used	Unet	Unet++
Image Size	224×224(pixel)	
Loss Function	Fβ(β=0.3,0.4,0.5,1,2,3,10,20)	
Number of Training Data	7408 pairs	
Number of validation Data	1856pairs	
Number of Training Iterations	100 times	300times
Batch Size	16	
Optimization Method	Adam(learning rate=1.0×10-4)	

4. RESULTS

4.1 Experiment 1

Figure 5 shows the changes in the loss function values for Unet and Unet++, respectively, as the β values were varied, as listed in the tables. Figures 6 and 7 show the difference images for $\beta = 0.5, 1, 3$, and 20. From Figure 5, it can be observed that recall decreases with an increase in β , while precision increases with an increase in β . Thus, it is possible to adjust recall and precision by modifying the β values. Since recall and

precision have a trade-off relationship, the F-value does not change significantly with varying β . However, there is a slight change in the maximum F-value observed at $\beta = 2$ for this test image. Figures 6 and 7 show that both Unet and Unet++ generated false positives for areas with similar hues to the sediment, such as sandy areas.

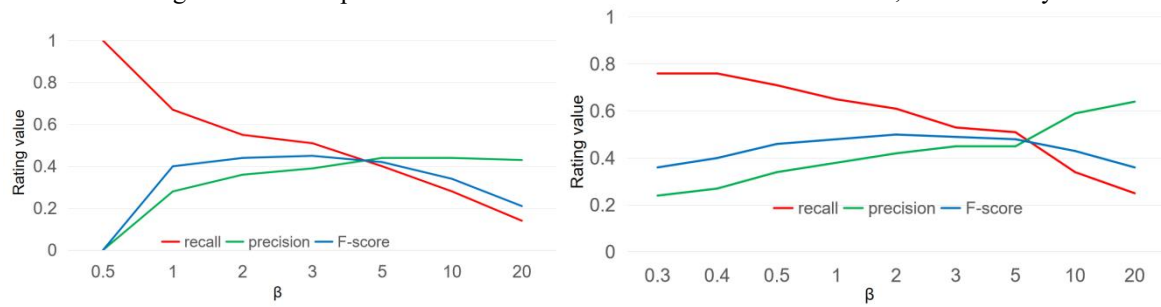


Figure 5. Comparison of evaluations when modifying the Dice coefficient β for Unet (left) and Unet++ (right).

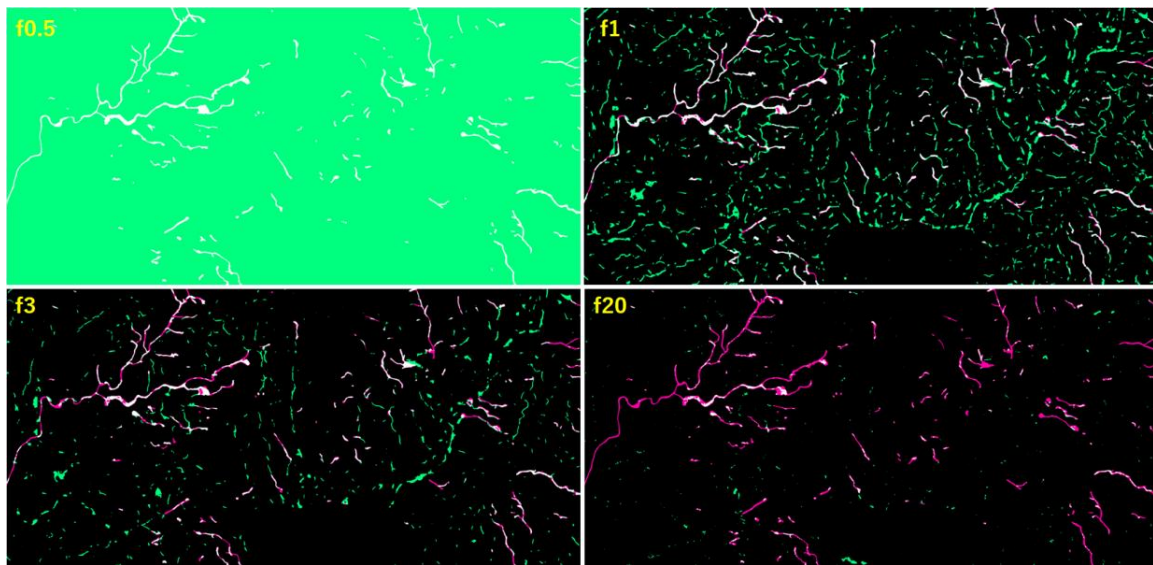


Figure 6. Difference images when varying the Dice coefficient β in Unet (Test 4).

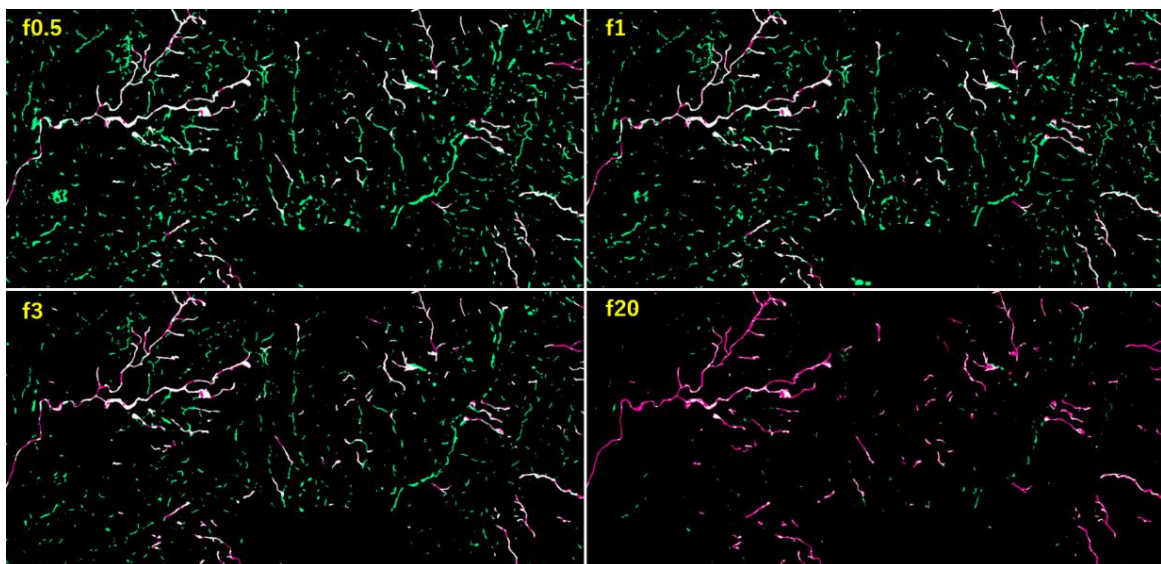


Figure 7. Difference images when varying the Dice coefficient β in Unet++ (Test 4).

4.2 Experiment 2

Table 2 presents the difference in F-scores between the results obtained by applying the sliding segmentation method and the simple segmentation method. Unet++ was trained using the loss function coefficient patterns β from Table 1, and the sliding segmentation method was applied to Test 4. The F-scores (values in parentheses in Table 2) were calculated after applying either "and" or "or" processing to the output

from the detector. The difference values were then derived by subtracting the F-scores of the simple segmentation method from those of the sliding segmentation method.

For $\beta = 1, 2$, and 3 , the F-score difference values (AND) are positive, indicating that the sliding segmentation method outperforms the simple segmentation method. Similarly, for $\beta = 10$, the F-score difference values (OR) are positive, further demonstrating the reliability of the sliding segmentation method over the simple segmentation method. These results suggest that the sliding segmentation method can enhance the performance of the detector as shown in the image comparison in Figure 8.

Table 2. The difference in F-score between the OR/AND combination and the simple combinations.

β	1	2	3	5	10
AND combination's F-score difference value (F-score)	0.05 (0.53)	0.04 (0.55)	0.03 (0.54)	-0.02 (0.52)	-0.04 (0.43)
OR combination's F-score difference value (F-score)	-0.05 (0.43)	-0.05 (0.46)	-0.03 (0.49)	-0.02 (0.52)	0.04 (0.51)

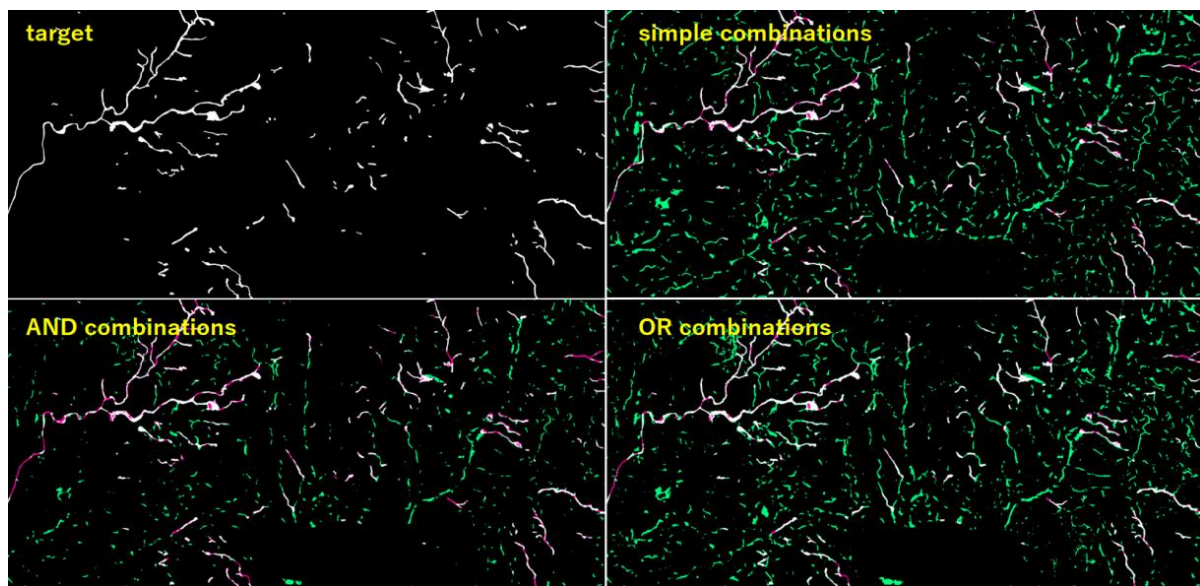


Figure 8. Difference images using different combinations (Test 4).

4.3 Experiment 3

Figure 9 shows the transitions in Unet's performance using AND and OR combinations for different β values. Figures 10 and 11 show the difference images for $\beta = 0.5, 1, 2$, and 3 . From these, we can see that recall and precision follow the same trends for simple division, AND combination, and OR combination. As in Experiment 1, recall and precision have a trade-off relationship, so the F-value does not change significantly with varying β . In Experiment 2, as with the AND combination, recall for the slide combination was lower than that for simple division. However, precision improved because only the detected areas were output. The recall for the OR combination improved since it outputs areas detected in any of the images, outperforming simple division, although precision decreased compared to simple division. When increasing β with the AND combination, the precision further improved compared to the AND combination alone. These results suggest that classifier performance is influenced by changes in β , even when using the slide combination.

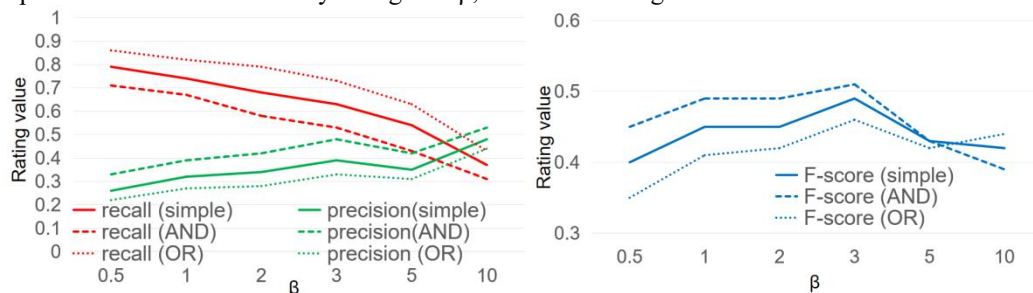


Figure 9. Comparison of precision, recall (left) and F-score (right) when varying the Dice coefficient β using the sliding fusion method.

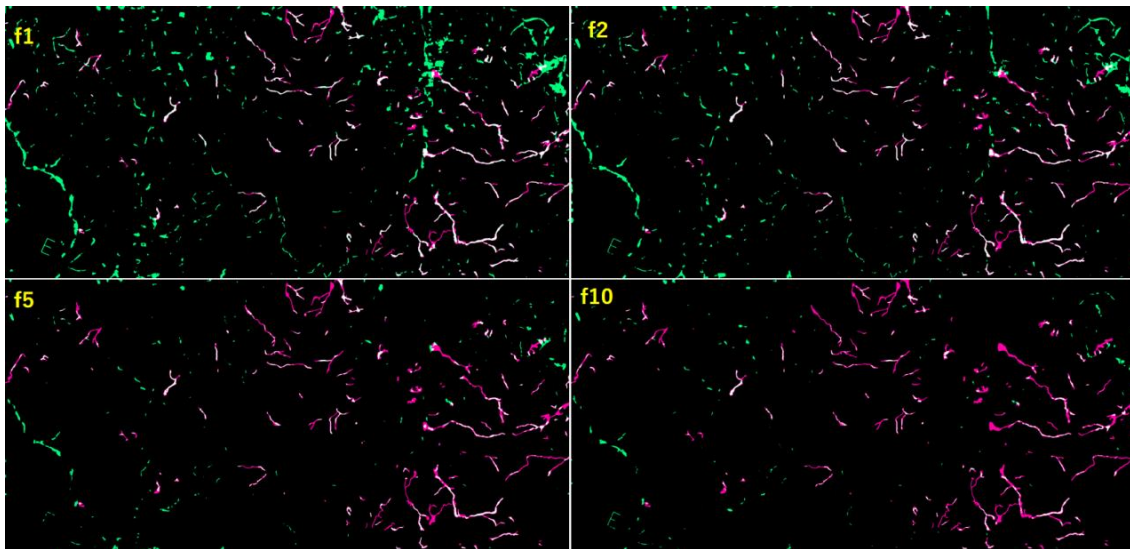


Figure 10. Difference images using AND fusion with varying Dice coefficient β (Test 2).

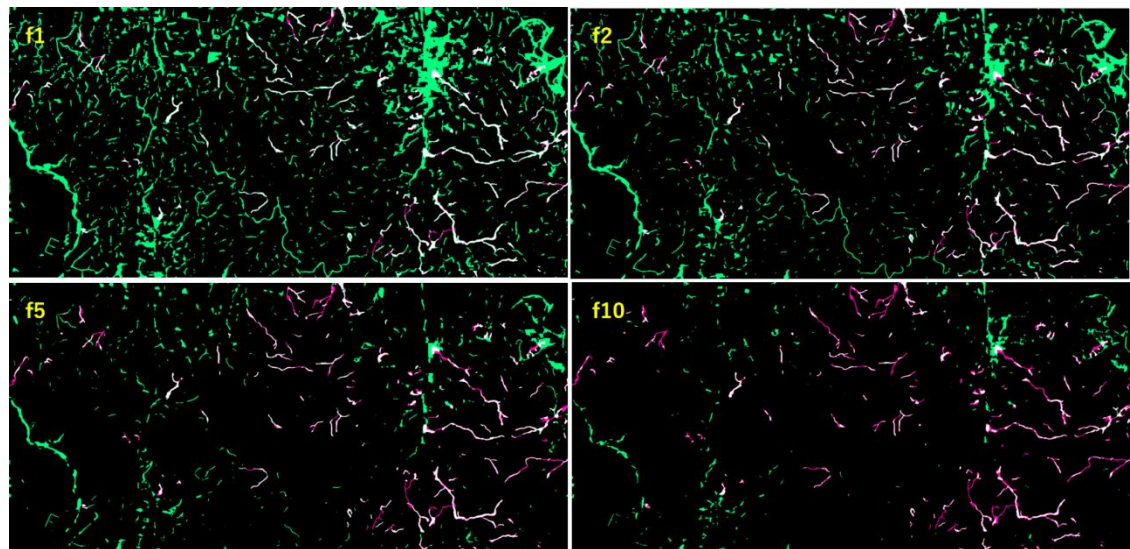


Figure 11. Difference images using AND fusion with varying Dice coefficient β (Test 2).

5. CONCLUSION

This study aimed to suppress false-positive detections that occurred during the analysis of sediment areas using the Unet classifier via the introduction of the Unet++ classifier, adjusting the loss function coefficient β , and adopting the sliding partitioning method. Experimental results suggested that the detection performance of sediment areas improved with the Unet++ classifier, as the F-score increased compared to the Unet classifier. By changing the loss function coefficient β , it is possible to adjust the balance between recall and precision without significantly affecting the F-score. Moreover, the AND combination method in the sliding partitioning technique was able to enhance precision while suppressing the decline in recall. As a result, the F-score improved, leading to an overall enhancement in classifier performance.

When developing deep learning software, it is necessary to design a classifier tailored to the preferences of the user performing the interpretation tasks. In this study, the Unet++ classifier, using sliding partitioning, showed an improvement in precision and an increase in the F-score compared to the Unet classifier. Additionally, by adjusting the value of the loss function coefficient β , it is possible to emphasize either recall or precision without changing the F-score. However, for different training datasets, the optimal combination method and the corresponding β value may vary, which warrants further investigation in future studies.

It remains essential to further refine the combination methods in the sliding partitioning technique to enhance precision without reducing recall. Although we compared detection performance using the Unet++ model, there are other deep learning methods such as pix2pix that are utilized for image segmentation. In the future, it will be necessary to conduct performance comparisons with models other than the Unet system and explore ways to further improve detection performance.

REFERENCES

- Erhong, Y., Zhihua, W., & Shaohui, J. (2007). The application of remote sensing digital technique in the investigate of the tiantai landslide in Xuanhan County. *Journal of Jilin University (Earth Science Edition)*, 37(3), 557-563.
- Liu, Y., & Wu, L. (2016). Geological disaster recognition on optical remote sensing images using deep learning. *Procedia Computer Science*, 91, 566-575.
- Tadataka ISHII, Michiyuki HIROKANE, Kazumasa KURAMOTO, & Naoki NSHIHARA. (2020). CONSIDERING SOCIAL FACTOR IN SETTING OF SEDIMENT DISASTER WARNING AREA USING DEEP LEARNING. *Japanese Journal of JSCE F6*, 76(2), I_193-I_199.
- Kei KAWAMURA, Yushi NAKAMURA, Tsuyoshi WAKATSUKI, & Toshikazu SAMURA. (2018). A STUDY ON THE AUTOMATIC DETECTION OF LANDSLIDE FROM AERIAL PHOTOGRAPHS USING DEEP LEARNING. *Japanese Journal of JSCE F3*, 74(2), I_132-I_143.
- Kei KAWAMURA, Shunsuke SUGAHARA, Syoki RYU, & Tsuyoshi WAKATSUKI. (2023). A STUDY ON THE AUTOMATIC DETECTION OF LANDSLIDE FROM AERIALPHOTOGRAPHS USING UNet AND UNet+. *Japanese Journal of JSCE*, 79(22), 22-22026.
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention-MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18* (pp. 234-241).
- Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., & Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4* (pp. 3-11).
- Hand, D., & Christen, P. (2018). A note on using the F-measure for evaluating record linkage algorithms. *Statistics and Computing*, 28, 539-547.

Development of 'Voices of Bridges' for Infrastructure Management

Yota Chinen¹, Kei Kawamura², and Shuji Sawamura³

1) Undergraduate School of Sciences and Technology for Innovation, Yamaguchi University, Yamaguchi, Japan. Email: b701ffv@yamaguchi-u.ac.jp

2) D. Eng., Prof., Graduate School of Sciences and Technology for Innovation, Yamaguchi University, Yamaguchi, Japan. Email: kay@yamaguchi-u.ac.jp

3) Department of construction, Yamaguchi Prefecture Government, Yamaguchi, Japan. Email: sawamura.shiyuuji@pref.yamaguchi.lg.jp

Abstract: In Yamaguchi Prefecture, periodic inspections of social infrastructure facilities, such as bridges, have become mandatory every five years due to revisions to the Road Act. However, a significant challenge arises from the lack of personnel available to conduct these inspections, especially in rural municipalities. This shortage of staff, combined with the rapid aging of infrastructure, has heightened the risk of accidents and disasters. This paper introduces the "Voices of Bridges" (VoB) system, which leverages citizen participation to address bridge maintenance through a user-friendly, interactive platform. The system's architecture integrates modern web technologies such as Node.js, MySQL, and Google Maps API to allow efficient data handling and visualization. By utilizing open data from Yamaguchi Prefecture and employing crowdsourcing techniques, VoB enables users to report bridge defects while simultaneously augmenting the efforts of local governments in maintaining infrastructure through community engagement.

Keywords: Crowdsourcing, Infrastructure Maintenance, Bridge Inspection, Citizen Participation, Open Data

1. INTRODUCTION

In Japan, aging infrastructure has become a pressing issue. Following the Sasago Tunnel ceiling collapse in 2012, the government has enacted strict regulations mandating periodic inspections of public facilities every five years. While this policy aims to prevent accidents by ensuring infrastructure safety, local governments in rural areas face significant challenges related to the large number of bridges requiring inspection and the limited availability of civil engineering personnel. Yamaguchi Prefecture is no exception, with a growing number of bridges that require ongoing maintenance. The need for a scalable solution to address the aging infrastructure is critical, and traditional methods of inspection are both time-consuming and costly.

To mitigate these challenges, citizen participation in infrastructure maintenance is a model currently being considered as a possible solution. Systems such as "Chiba Repo" and "Native Ube" have demonstrated the potential for crowdsourced data collection to facilitate maintenance and repairs; however, these systems focus on urban maintenance tasks like reporting road damage and illegal dumping, and thus do not address specific needs of bridge management. This paper presents "Voices of Bridges" (VoB), a web-based system designed to harness local community involvement in bridge maintenance where citizens can report defects on bridges using a simple interface, improving information sharing and aiding local governments in managing their aging infrastructure.

2. RELATED WORKS AND TECHNOLOGIES

2.1 Overview

This section reviews existing research and related technologies that contribute to the foundation of VoB. Notable systems such as "Native Ube" and "Chiba Repo" are citizen-driven platforms that enable local governments to respond to issues reported by the public. These systems allow users to upload reports with images and location data, promoting quick action by municipal authorities. VoB draws from these concepts but focuses on a more specialized domains: bridge health and infrastructure management.

2.2 Similar Systems

(1) Native Ube

Native Ube is a community-based reporting system developed by Ube City, designed to allow residents to easily report local issues such as road damage, illegal dumping, and concerns about public facilities (Gougai NET Ube City & Sanyo Onoda). The system integrates user-friendly features such as geolocation and photo submission, allowing citizens to provide accurate and detailed reports (Ube City Official Website). By attaching photos and using location data, users can precisely convey the nature and exact location of the problems they encounter, facilitating a quicker response from the city administration.

Despite its effectiveness in handling general civic issues, however, Native Ube's scope is limited when compared to systems like VoB in the domain of infrastructure management. While VoB is explicitly designed to facilitate the inspection, maintenance, and management of bridges, focusing on critical infrastructure, Native Ube does not provide specialized tools or categories for reporting infrastructure-related issues such as bridge defects (Yinyongdong Ma et al., 2024). Its primary focus is on general urban maintenance, making it less suited for addressing the complex needs of infrastructure monitoring and safety assessments essential for bridge management.

In terms of citizen engagement, both Native Ube and VoB share a common philosophy: empowering citizens to contribute to the maintenance of their city through digital tools. The photo and location submission features in Native Ube mirror those found in VoB, demonstrating the similarity in the participatory approach. However, while VoB emphasizes a collaborative model between experts, facility managers, and citizens to ensure the accuracy and reliability of reports for critical infrastructure, Native Ube primarily focuses on general community issues without the involvement of specialized experts. This difference highlights VoB's targeted approach to bridge maintenance in contrast to Native Ube's more general scope (Sorabatake2024).

In addition, VoB provides a structured framework for handling reports by integrating bridge-specific data from Yamaguchi Prefecture's Open Data repository. This makes VoB more suitable for long-term infrastructure management, as it allows maintenance experts to analyze defect patterns and ensure the safety of critical structures over a longer timeframe. Native Ube, on the other hand, does not possess this specialized integration with infrastructure datasets, limiting its use for in-depth infrastructure assessment.

(2) Chiba Repo

Chiba Repo, developed by the City of Chiba, offers functionality similar to Native Ube, with a focus on reporting urban issues such as broken streetlights, damaged roads, and graffiti. The platform is widely used among Chiba residents, who can report issues through the system and monitor the status of their reports. Chiba Repo integrates open data to provide transparency in the way city officials handle reports, which allows citizens to see how their concerns are being addressed and ensures accountability in local governance (City Chiba).

A notable feature of Chiba Repo is an ability to provide real-time updates of status of reports that encourages ongoing citizen engagement by showing users the progress made in addressing their submissions. This transparency resembles VoB's reporting status features, where users can track the resolution of bridge defects and other infrastructure concerns. Both systems rely on open data to foster trust and collaboration between the public and local government, but their specific focus areas differ significantly.

While Chiba Repo is an effective platform for general urban maintenance, it is not tailored for the inspection and upkeep of bridges. Like Native Ube, Chiba Repo lacks the specialized tools and categories required for reporting structural defects in bridges or other critical infrastructure. In contrast, VoB is designed specifically to support infrastructure safety, allowing for detailed reporting on bridge conditions, including crack size, structural integrity, and material degradation. VoB integrates professional input from Maintenance Experts (ME) to ensure reports are reviewed by individuals with expertise in assessing infrastructure risk, something Chiba Repo does not provide.

Another difference is the geospatial mapping capability. While Chiba Repo allows users to report issues with geolocation, VoB goes a step further by providing an interactive Google Maps API-powered interface that highlights all bridges in a given area, allowing users to visualize the distribution of bridge conditions and defects. This spatial awareness is vital for infrastructure maintenance, as it allows facility managers to prioritize areas that may require immediate attention based on the concentration of reported issues.

Finally, VoB's multifaceted user base, which includes both general citizens and experts, sets it apart from Chiba Repo's citizen-focused model. By incorporating engineering expertise alongside public input, VoB ensures a more streamlined and thorough process for evaluating the condition of bridges, making it a more robust tool for infrastructure management than Chiba Repo.

2.3 Related technologies

(1) Open Data

VoB leverages open data provided by Yamaguchi Prefecture, which includes detailed information about bridges such as location, construction date, and maintenance history. By utilizing this open data, VoB can automatically populate bridge information and provide accurate context for user reports (Public Transport Open Data Council 2024).

(2) Node.js

Node.js is employed in the VoB system as the backend framework due to its scalability and non-blocking I/O model. The non-blocking I/O model allows the system to process multiple tasks at the same time without waiting for one task to finish before starting another. For example, when a user submits a bridge report, the system can immediately start processing it while still handling other user requests, such as displaying a map or searching for bridge information. This means that even if many users are interacting with the system simultaneously, it can continue responding quickly without delays. This approach ensures that the VoB system remains responsive and can handle heavy traffic efficiently, making it ideal for real-time applications like infrastructure management.

(3) Express

Express is a lightweight web application framework for Node.js used in VoB to manage routing and other server-side functionality. It facilitates the smooth processing of user requests, such as submitting reports or retrieving bridge data from external databases.

(6) EJS (Embedded JavaScript)

EJS is the templating engine used to dynamically generate HTML pages. By embedding JavaScript

directly into HTML, EJS enables VoB to provide real-time, dynamic content to users such as bridge information and recent reports.

(4) MySQL

VoB employs MySQL as its database architecture to store bridge information, user reports, and login history. The relational structure of MySQL ensures data is efficiently stored and accessible, allowing quick retrieval of search queries or user interactions.

3. POSITIONING AND SYSTEM DESIGN

3.1 Overview

This section outlines the system's design and development, highlighting its role as an innovative solution for enhancing the efficiency of bridge maintenance by involving a broad range of users, from experts to the general public. By leveraging modern web technologies and integrating open data, VoB provides a collaborative platform where citizens and professionals can work together to ensure the safety and longevity of bridges.

Infrastructure management is a pressing issue for local governments, particularly in regions like Yamaguchi Prefecture, where a significant number of bridges require regular inspection. Traditional methods of inspection rely heavily on professional engineers and municipal staff, but these resources are often stretched thin due to budget constraints and staffing shortages. VoB bridges this gap by incorporating a crowdsourcing approach, allowing both citizens and experts to participate in the reporting and maintenance processes. The system can handle a wide range of user interactions while providing tools for data collection, inspection, and decision-making — all within a scalable, cloud-based environment.

3.2 Positioning of the Research

The VoB system is positioned as a complementary tool to existing bridge inspection methodologies. Rather than replacing professional inspections, VoB enhances them by introducing a community-driven element that increases the frequency and granularity of monitoring. In traditional settings, bridges are inspected periodically by professional engineers, but this can result in long intervals between inspections that can lead to delays in identifying damage or wear.

With VoB, both ME and general citizens can contribute to the ongoing assessment of bridge conditions. Citizens act as first-line observers, reporting visible defects or concerns as they encounter them. This ensures that potential issues are flagged early, well before scheduled inspections. Experts, in turn, review these reports and provide technical assessments, making the system a collaborative environment where different user groups complement each other's efforts.

The system is particularly effective in addressing the challenges posed by the aging infrastructure in rural and suburban areas. By encouraging citizen engagement, VoB increases the capacity for monitoring, while still preserving the expert oversight necessary to ensure that reports are properly evaluated and prioritized. This combination of community participation and professional expertise makes VoB a scalable and sustainable solution for infrastructure management.

3.3 Development Environment

VoB was developed in a cloud-based environment, utilizing Amazon Web Services (AWS) to ensure scalability, security, and high availability. The choice of AWS as the hosting platform allows VoB to handle increasing amounts of data and traffic as more users engage with the system. The flexible infrastructure provided by AWS supports the system's rapid growth without compromising performance.

On the backend, Node.js is employed for its event-driven, non-blocking architecture, which makes it ideal for handling multiple simultaneous requests from users reporting defects or retrieving bridge information. Express, a web application framework for Node.js, is used to streamline the development of APIs and manage server-side logic, ensuring that user requests are processed efficiently.

The system stores user-generated data, bridge information, and inspection reports in a MySQL relational database. MySQL was chosen for its ability to handle large datasets efficiently and its strong support for complex queries, which are necessary for retrieving detailed bridge reports or user activity logs. The integration of MySQL ensures that all data is stored in a structured and secure manner, facilitating quick access and analysis by system administrators and experts.

Google Maps API is integrated into VoB to provide geospatial services, allowing users to view the location of bridges on an interactive map and to report defects directly from the map interface. This map-based interface simplifies the user experience, making it easy for non-experts to locate bridges and contribute to the maintenance process.

3.4 System Architecture

The system architecture of "Voices of Bridges" (VoB) is designed to support interactions between various user roles, including general citizens, inspection experts, and facility managers, as well as external data sources such as Yamaguchi Prefecture Open Data and Google Maps API. Figure 1 illustrates the flow of information between these users and systems, highlighting key interactions that enable efficient bridge maintenance reporting and management.

The flow of information between the Service Link Platform (SLP), the Yamaguchi Prefecture Open Data, and VoB demonstrates how external data is integrated into the system. Google Maps API is used to provide location-based services, such as mapping bridge locations and reporting defects. The system supports three main user roles: general citizens who submit reports and view bridge data, inspectors (Experts), who provide detailed assessments of reported defects, and facility managers who oversee the entire maintenance process.

3.5 Technologies Used

The VoB system integrates multiple web technologies to deliver a seamless and intuitive user experience. Node.js and Express form the backbone of the system, managing user requests and ensuring fast response times. These technologies were chosen for their ability to handle the real-time nature of the platform, where multiple users may be submitting reports or viewing data simultaneously.

EJS (Embedded JavaScript) is used on the front-end (user end) to dynamically generate HTML pages. EJS allows VoB to deliver personalized content to users based on their roles and actions, such as displaying different dashboard options for general users versus administrators. This dynamic content generation is crucial for ensuring that users only see the information and tools relevant to their role.

The system relies on MySQL for data management, storing everything from user accounts to detailed bridge inspection logs. The relational structure of MySQL makes it easy to manage complex relationships between entities in the system, such as linking user reports to specific bridges and tracking the status of each report.

Google Maps API plays a central role in the visual representation of bridges, enabling users to locate and report defects. The map interface is fully interactive, allowing users to zoom in on specific regions and click on individual bridges to view more detailed information or submit reports.

3.6 Expanded System Architecture and Data Flow

The VoB system integrates data from multiple sources and facilitates seamless interaction between users and system components.

Yamaguchi Prefecture Open Data: VoB automatically imports bridge data from the Yamaguchi Prefecture Open Data repository. This data includes detailed information about each bridge's location, construction history, and maintenance status, ensuring that users have access to the most up-to-date information when submitting reports or reviewing data.

Google Maps API: The Google Maps API provides geospatial functionality, allowing users to interact with maps that display the locations of bridges. This integration simplifies the process of locating a bridge, making it easier for users to submit reports with accurate location data.

Service Link Platform (SLP): SLP is a platform developed by our research lab, designed to provide access to open data via API. It serves as an intermediary layer, offering data in API format, such as bridge-related information, for seamless integration into systems like VoB. SLP ensures secure and efficient data transfer between external data sources and internal system components, supporting real-time updates and interactions.

User Roles: Each user interacts with the system in a different way, contributing to a multi-directional flow of data. General citizens provide real-time reports of defects, experts offer detailed assessments, and administrators manage the system, ensuring that data is processed and acted upon effectively. This interaction model ensures that infrastructure maintenance is more responsive and efficient, leveraging the strengths of both professional and public input.

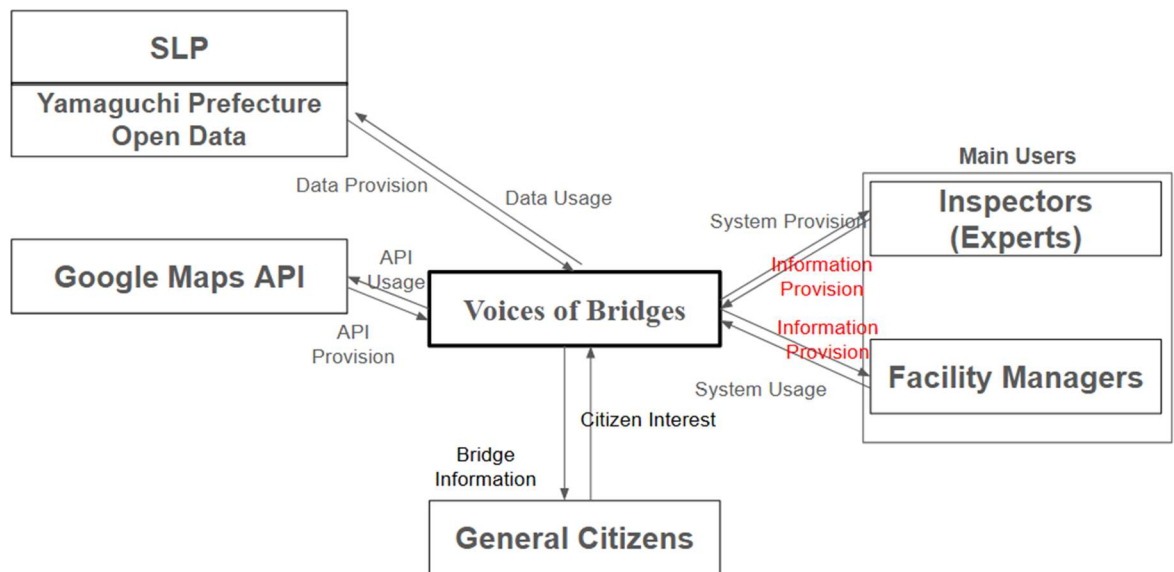


Figure 1. System Architecture of Voices of Bridges (VoB)

4. IMPLEMENTATION ARCHITECTURE

4.1 Overview

This chapter discusses the key implementation results of the "Voices of Bridges" (VoB) system, focusing on the main features and functionalities that facilitate bridge maintenance through citizen engagement. By combining open data, user input, and modern web technologies, the system enables efficient monitoring and sharing of bridge conditions. Below, we break down the core features of VoB and how they contribute to infrastructure management.

4.2 Structure and Features of Voices of Bridges

(1) User Access and Roles

VoB is designed to cater to multiple user types and accordingly adjusts its interface and available tools based on the role of the user. General citizens are primarily focused on reporting defects, while experts are tasked with providing detailed assessments. Facility managers oversee the entire process, ensuring reports are reviewed and acted upon appropriately. Upon accessing the system, users are presented with an interface that guides them through login or registration as necessary. Once logged in, the dashboard is customized to the user's role, showing relevant tasks such as viewing recent reports, accessing bridge information, or managing inspection requests.

(2) Bridge Search and Filtering

VoB provides an intuitive search function that allows any site visitor to quickly find bridges based on location or name. Using the search bar, users can input specific criteria to filter the bridges displayed. This feature is useful for facility managers who may need to locate specific bridges for maintenance scheduling or for general users who want to report defects on bridges in their area. Search results include key information such as the bridge's name, location, and any existing reports or inspections.

(3) Interactive Map and Bridge Data Visualization

One of the primary features of VoB is the interactive map powered by the Google Maps API. This map provides a visual representation of bridge locations across Yamaguchi Prefecture. Users can click on map markers representing individual bridges to access detailed information, including the bridge's construction date, dimensions, and any ongoing maintenance activities. The map also highlights bridges that have recently been reported by users, making it easy for facility managers to identify which bridges may require immediate attention (see Image 1).

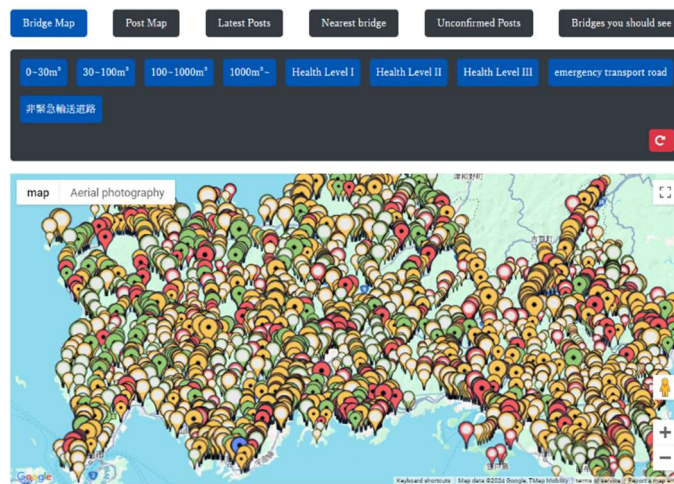


Image 1. Bridge Map

In addition to displaying bridge information, the system features a Post Map that visualizes all defect reports submitted by users. Each report is represented by a marker on the map, color-coded based on its status (e.g., "Unconfirmed" or "Confirmed"). This helps facility managers and other users quickly identify reported issues and assess their distribution across the region. Users can click on individual markers to view the details of a report, such as photos, descriptions, and any follow-up actions taken. This not only aids in identifying problem areas but also enhances user engagement by allowing them to see the impact of their contributions (see Image 2).

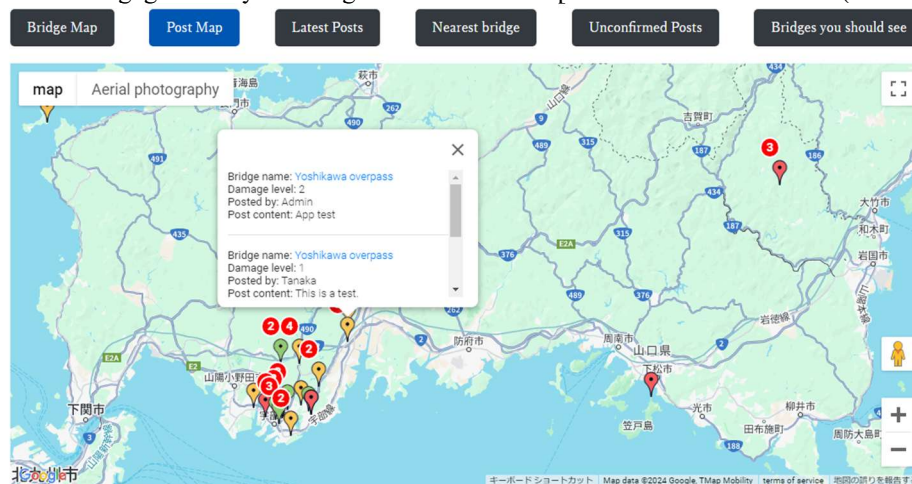


Image 2. Post Map

Furthermore, the Bridge Map provides an overview of all bridges in the prefecture, allowing users to filter and search for specific structures based on various criteria, such as location or inspection status. The interactive map makes it easy to pinpoint bridges that may need urgent inspection or maintenance, thereby streamlining the prioritization process. Additionally, integrating the Google Maps API enables dynamic map updates, ensuring that information remains current as users move or as new data is submitted.

By integrating both the Post Map and Bridge Map with the system's reporting and visualization features, VoB simplifies the process of geographically identifying problem areas and prioritizing bridge inspections. The visual nature of these maps not only helps general citizens understand the location of bridges and the relevance of their reports within the broader infrastructure network but also supports facility managers in making informed maintenance decisions based on real-time data.

(4) Report Submission and Defect Tracking

The report submission feature is at the heart of VoB's user interaction. Citizens and experts alike can submit reports about bridge defects or concerns directly through the platform. The report form allows users to upload photos, add comments, and mark the location of the bridge using GPS data. This process ensures that detailed and accurate information is captured for facility managers to review. Once a report is submitted, it is displayed in the system along with other existing reports, allowing facility managers to track all reports in a single location. They can mark reports as reviewed, in progress, or resolved, helping to maintain transparency and

accountability in the maintenance process. Users who submit reports are notified when their report's status changes, encouraging continued participation by keeping them informed of the progress of their contributions.

(5) Engagement Features: Likes and Follow-Up

To encourage continued user engagement, VoB incorporates a "Like" feature for reports. Users can show support or agreement with a report by liking it, signaling to facility managers that certain defects are particularly concerning to the public. This feature is especially valuable in cases where multiple users report the same issue or where a particular bridge is frequently mentioned. Additionally, VoB allows users to follow up on previously submitted reports by adding new photos or comments if the situation changes or worsens. This dynamic reporting system ensures that bridge conditions are continuously updated, giving facility managers the most up-to-date information possible.

(6) Nearest Bridge Identification Feature

One of the key interactive features of the *Voices of Bridges* (VoB) system is the Nearest Bridge Identification functionality (see Image 3). This feature, available to both general users and experts, leverages geolocation data to show users the closest bridges based on their current location. When users enable location services on their device, the system automatically identifies their current position and displays a list of nearby bridges sorted by proximity. This feature, powered by the Google Maps API, not only displays bridges on a map but also allows users to interact with them. The map dynamically updates as users move, ensuring that the list of nearby bridges remains accurate and up-to-date. The system calculates the distance between the user's current location and each bridge using the Haversine formula, providing an accurate measurement of proximity.

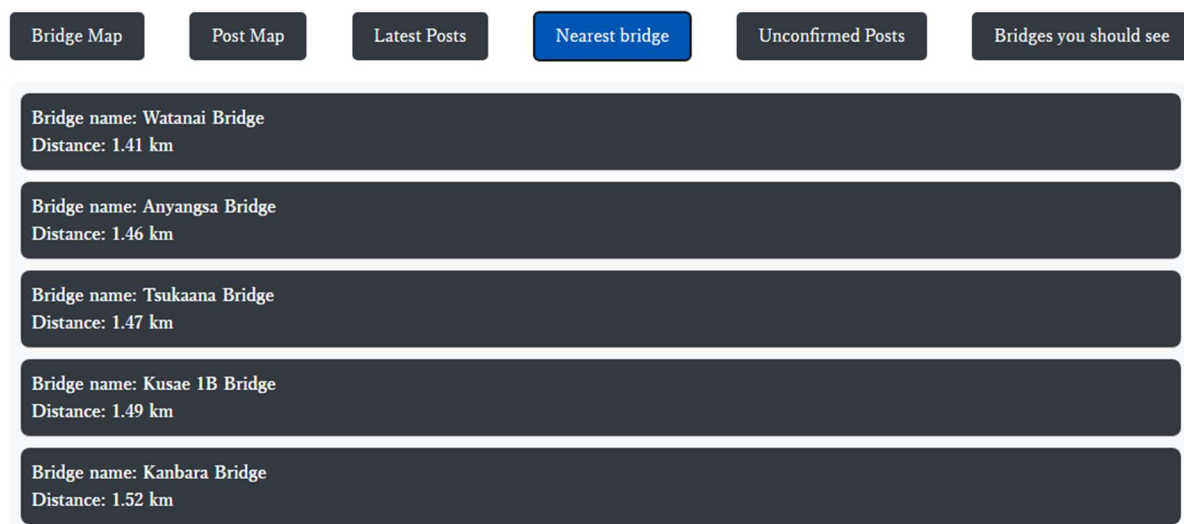


Image 3. Nearest bridge

Once nearby bridges have been identified, VoB presents the user with detailed information for each bridge, including the bridge's name, its condition, and any recent defect reports. This information is shown in a visually appealing format with an animation effect that enhances the user experience. Each bridge on the list is clickable, allowing users to view detailed reports or submit new defect information directly from the nearest bridge display. This interactive approach not only helps in reporting but also promotes public engagement in bridge monitoring.

(7) Unconfirmed Posts

The "Unconfirmed Posts" feature of the "Voices of Bridges" (VoB) system provides an organized way to manage and review bridge defect reports submitted by users (see Image 4). This section categorizes reports as one of two statuses: "Unconfirmed" or "Confirmed." When a user submits a report about a bridge, it is initially marked as "Unconfirmed." Facility managers and maintenance experts review the content of the report, and after it has been verified and processed, they change its status to "Confirmed." Reports marked as "Confirmed" automatically disappear from the list of unconfirmed posts to streamline the workflow.

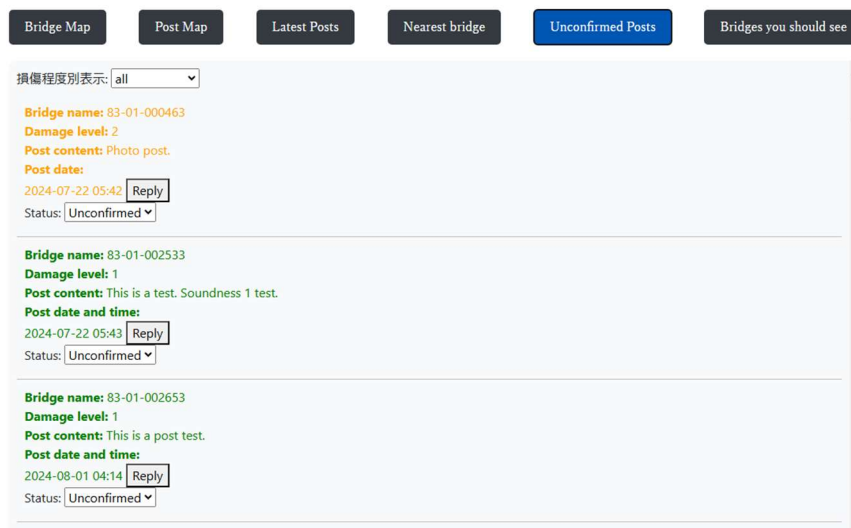


Image 4. Unconfirmed Posts

This feature ensures that unaddressed issues remain visible and prioritized until appropriate action is taken, while completed reports are hidden from view to prevent clutter. The mechanism also promotes efficiency and transparency, allowing facility managers to track the progress of all defect reports and focus on unresolved cases. By keeping the user engaged with status updates and ensuring that reports are handled systematically, this feature helps maintain the integrity of the overall infrastructure management process.

4.3 Conclusion of Implementation Results

The implementation of VoB successfully demonstrates how a web-based platform can integrate open data with user-generated content to improve the efficiency of bridge maintenance. Through features such as the interactive map, defect reporting system, and role-based interfaces, VoB facilitates collaboration between experts, facility managers, and the general public. The system's real-time data sharing and reporting capabilities contribute to more informed decision-making and timely maintenance actions, ensuring that critical infrastructure in Yamaguchi Prefecture is managed effectively.

5. PERFORMANCE VALIDATION

This chapter focuses on validating the effectiveness and reliability of the Voices of Bridges (VoB) system through various methods, including user feedback, system performance testing, and real-world case studies. The aim of this validation process is to assess how well the system meets its objectives of enhancing bridge maintenance efficiency, improving user engagement, and ensuring scalability in handling both technical and user demands.

5.1 User Feedback

User feedback plays a crucial role in evaluating the usability and practicality of the VoB system. In its initial phase, the system was tested with a select group of Maintenance Experts (ME) in Yamaguchi Prefecture. These experts were tasked with using the platform to identify, report, and assess defects on local bridges. Their feedback was overwhelmingly positive, particularly regarding the system's ability to streamline the defect identification process.

The map-based interface was frequently highlighted as one of the system's most effective features. The integration of the Google Maps API allowed experts to visualize the locations of bridges in real time, making it easier to pinpoint and report issues. Additionally, the ability to view previous reports and overlay current bridge conditions provided context and historical insight, which was invaluable for assessing the urgency of new reports.

Experts also appreciated the report submission workflow, which allowed them to attach photos, add detailed comments, and categorize defects based on severity. This structured approach to reporting made it easier for facility managers to prioritize maintenance tasks and respond to urgent issues more efficiently. Overall, users reported that the system not only saved time but also reduced the likelihood of missing critical defects between inspections.

VoB's usability was also tested among general citizens, who were encouraged to report visible defects or concerns about bridges in their local communities. This group provided feedback on the accessibility and intuitiveness of the platform. Despite initial concerns about non-experts using a technical system, the general public found VoB easy to navigate. Many users cited the simplicity of submitting reports via mobile devices,

which enabled them to contribute to infrastructure maintenance without requiring specialized knowledge. The like feature for reports also helped to encourage participation by allowing users to endorse reports, thus helping prioritize issues that were of broader public concern.

5.2 Real-World Case Studies

To assess the system's usability and functionality, VoB was tested in real-world scenarios in Yamaguchi Prefecture, where bridge maintenance is a key concern. These tests focused on verifying whether the system could operate correctly, rather than collecting actual defect reports. Both experts and general users participated in evaluating the platform's ease of use, testing functions such as posting reports, and checking whether features like locating nearby bridges and current location tracking were working properly.

Instead of receiving actual defect reports, the testing involved ensuring that users could easily navigate the platform, submit reports correctly, and that the system could accurately display the nearest bridges based on the user's location. The system's map-based interface was also tested to confirm that it could provide real-time location data, ensuring that users could find relevant bridges in their vicinity.

These tests highlighted VoB's potential to enhance user interaction through its intuitive design, and the system successfully demonstrated its ability to handle key functionalities such as data submission and geolocation. While no real maintenance actions were triggered, the trials confirmed that VoB can serve as a reliable platform for future citizen-driven infrastructure monitoring and reporting.

5.3 Lessons Learned and Areas for Improvement

While the initial validation of VoB has been highly positive, several areas for improvement have been identified. One key lesson learned is the importance of providing continuous user support. Some general users initially struggled with the technical aspects of submitting reports, particularly when attaching photos or using the geolocation features. To address this, future iterations of VoB will include enhanced user guides and tutorials, as well as a simplified submission process to further lower the barrier for citizen participation.

Another area for improvement is the integration of predictive analytics. While VoB is effective at collecting and displaying current defect data, there is potential to enhance the system by incorporating AI-driven analytics. These features could predict future defects based on historical data trends and environmental factors, allowing facility managers to proactively schedule maintenance before issues arise. Such predictive capabilities would elevate VoB from a reactive tool to a more proactive infrastructure management system.

6. CONCLUSION

This paper introduced and examined the "Voices of Bridges" (VoB) system, a novel approach to addressing the growing challenge of infrastructure maintenance in Yamaguchi Prefecture. VoB was developed as a web-based platform that enables citizen participation, aiming to complement traditional maintenance practices by engaging the community in reporting defects or concerns about local bridges. This system serves as a response to the critical issue of aging infrastructure, where local governments face significant difficulties due to resource constraints and manpower shortages.

By integrating open data provided by Yamaguchi Prefecture with modern web technologies such as Node.js, MySQL, and the Google Maps API, VoB has successfully created a scalable and efficient solution. The platform's key strength lies in its ability to merge expert knowledge with the power of crowdsourcing, allowing citizens, maintenance experts, and facility managers to collaborate and contribute to the ongoing upkeep of critical infrastructure. The system's user-friendly interface, combined with its robust backend structure, ensures that defect reports are collected, processed, and managed in real time, improving the responsiveness of local governments to infrastructure issues.

Citizen participation is a core pillar of the VoB system, and it has been shown that by enabling residents to report issues such as cracks, rust, or other potential defects in bridges, local authorities can more effectively prioritize their maintenance efforts. The engagement features, such as report "likes" and the ability for users to track the progress of their reports, help sustain community involvement and encourage transparency between the public and local governments. This participatory approach not only helps alleviate the workload of civil engineering departments but also fosters a greater sense of responsibility and awareness among the general public regarding the state of their local infrastructure.

Looking ahead, several opportunities for further development of VoB have been identified. One promising avenue is the integration of artificial intelligence (AI) and machine learning technologies to enhance defect detection. By analyzing historical data from reports and using predictive modeling, the system could potentially identify early warning signs of structural issues before they become severe, enabling preventive maintenance. Additionally, AI algorithms could be employed to automatically assess the severity of reported defects, further reducing the reliance on manual inspections and allowing experts to focus on high-priority cases.

Another potential development is expanding the system's capabilities to include automated notifications and alerts for both citizens and facility managers. For example, users could be notified when a report they

submitted has been resolved, or when a nearby bridge requires attention. Such features would not only improve the overall user experience but also ensure that maintenance activities remain timely and efficient.

In conclusion, the VoB system represents a powerful tool for modernizing the way local governments manage their aging infrastructure. By leveraging the collective knowledge and engagement of citizens and experts alike, VoB provides a more efficient and scalable approach to ensuring the safety and longevity of bridges. With future enhancements such as AI-driven analysis and expanded infrastructure coverage, the system holds the potential to transform infrastructure management in Yamaguchi Prefecture and beyond, contributing to a safer and more resilient society.

ACKNOWLEDGMENTS

This work was supported by Council for Science, Technology and Innovation (CSTI), Cross ministerial Strategic Innovation Program (SIP), Building a Smart Infrastructure Management System (Funding agency: Public Works Research Institute).

REFERENCES

- City Chiba. "Chiba Repo (My City Report)." Website:
<https://www.city.chiba.jp/sogoseisaku/shichokoshitsu/kohokocho/chibarepo.html>
- Sorabatake. "Delivering Citizen Issues Directly to Chiba City's Administration via 'Chiba Repo'." Website:
<https://sorabatake.jp/17419/>
- Ube City Official Website. "Overview and Features of Native Ube." Website:
<https://www.city.ube.yamaguchi.jp/>
- Gougai NET Ube City & Sanyo Onoda. "Ube City Launches 'Native Ube,' an App to Assist Citizens with Daily Life." Website: <https://ube-sanyoonoda.goguy.net.jp/>
- Public Transport Open Data Challenge 2024. Website: <https://challenge2024.odpt.org>
- Yinyongdong Ma, Kei Kawamura, Junha Hwang, Shuji Sawamura and Hisao Emoto. "PROMOTING FLEXIBLE USE OF OPEN DATA THROUGH SERVICE LINK PLATFORM FOR INFRASTRUCTURE MANAGEMENT"

CB-YOLOV5: STREETLIGHT DETECTION BASED ON LOW-LIGHT IMAGES IN HIGH-INTERFERENCE ENVIRONMENT

Shiqi ZHANG¹, Jingyuan TANG², Jingke HONG³

1) MPhil, School of Management Science and Real Estate, Chongqing University, Chongqing, China. Email: 1983153240@qq.com

2) Ph.D., Ass. Prof., School of Management Science and Real Estate, Chongqing University, Chongqing, China. Email: jtangap@cqu.edu.cn

3) Ph.D., Prof., School of Management Science and Real Estate, Chongqing University, Chongqing, China. Email: hongjingke@cqu.edu.cn

Abstract: The monitoring and operation maintenance (O&M) of urban streetlights is crucial for traffic safety and socio-economic development. However, how to accurately and robustly detect streetlights in low-light and high-interference environments is still a problem that concerns researchers. In recent years, deep learning has made remarkable progress in the field of object detection, among which the single-stage detection algorithm represented by You Only Look Once (YOLO) shows a satisfactory detection effect. It brings a new opportunity to detect streetlights based on images collected in a complicated street environment. Therefore, this study proposes an improved YOLOv5 model, as CB-YOLOv5, to accurately and robustly detect streetlights based on low-light images with high interferences. This proposed model integrates a Convolutional Block Attention Module (CBAM) and Bidirectional Feature Pyramid Network (BiFPN) to enhance its learning ability of spatial and channel dimension feature information, promote information fusion and transfer between multi-scale objects. Experimental results show that compared with the standard YOLOv5 algorithm, the proposed CB-YOLOv5 model can achieve significant improvement in accuracy and ability of interference resistance in streetlight detection tasks. The mAP_{0.5} reached 0.968, which is 23.5% higher than that of the standard YOLOv5 algorithm. In general, the CB-YOLOv5 model provides a new method to detect small objects in low-light and complex scenes. The developed method is also expected to provide a theoretical basis for automated monitoring and operation maintenance of urban lighting facilities.

Keywords: Streetlight detection, Low-light environment, YOLOv5, CBAM, BiFPN.

1. INTRODUCTION

With the acceleration of urbanization, the effective management and maintenance of lighting infrastructure, as a key factor in urban safety, residents' quality of life and energy consumption, has become increasingly important. As a core component of the urban lighting system, streetlights are not only related to pedestrian and vehicle safety at night but also directly affect the energy efficiency and environmental protection of the city. However, the traditional manual inspection method is inefficient and cannot achieve real-time monitoring and rapid response to the status of streetlights (Tosti et al., 2021). Therefore, the development of an efficient and accurate automatic streetlight recognition technology is of great significance for improving the intelligent level of urban management.

In recent years, the rapid development of deep learning technology, especially the progress in the field of target detection, has provided new solutions for streetlight recognition (T. Zhang, Dai, Song, Zhao, & Zhang, 2023). The YOLO series (You Only Look Once), as a leader in the field of real-time object detection, shows great potential in multiple application scenarios with its efficiency and accuracy (Y. Zhang et al., 2022). YOLOv5, as the latest generation, further improves detection speed and accuracy by optimizing the network structure and loss function (Wu, Wang, & Liu, 2021). However, in the face of complex and changeable urban environments, such as light change, uneven light, easy to be blocked, easy to confuse and other challenges, a single object detection model is often difficult to achieve the ideal recognition effect.

To overcome these challenges, in this study, the Convolutional Block Attention Module (CBAM) and Bidirectional Feature Pyramid Network (BiFPN) are introduced. CBAM enhances the feature representation through two dimensions of channel and space, enabling the model to focus on key information and ignore irrelevant background, so that the streetlight spots are not easy to be confused with other similar objects around (such as camera spots, Traffic light spots, etc.), thus improving the detection accuracy. BiFPN effectively solves the detection problem caused by the change of target scale through cross-scale feature fusion and enhances the detection ability of the model for small targets (such as distant streetlights). At the same time, combined with the efficient detection framework of YOLOv5, this study aims to build a high-performance streetlight recognition model to achieve accurate and rapid recognition of urban streetlights.

The main objectives of this study include: 1) For urban lighting infrastructure with streetlights as an example, a lightweight lighting facility recognition algorithm with both accuracy and stability is developed by using dashcams in unfavorable visual environments such as insufficient light, uneven light, shape occlusion and background change, and a streetlight recognition model based on CB-YOLOv5 is designed and implemented.

Realize real-time troubleshooting of lighting facility fault points; 2) The recognition performance of the model under low-light and high-interference environments was verified by experiments; 3) Compare and analyze the baseline YOLOv5 model to evaluate the improvement effect and practical application value of the model.

2. METHOD

2.1 Network Architecture of CB-YOLOv5

In the YOLOv5 model, to enhance the accuracy and robustness of the model's recognition of streetlights in complex and changeable environments, especially in the case of high occlusion and high interference, we took the following two improvement measures: 1) The CBAM module was introduced in the Backbone part. We chose to embed CBAM in the Backbone part of YOLOv5. By combining channel attention and spatial attention, CBAM module helps the model focus on the more critical feature information for street light detection, thus improving the accuracy and robustness of detection. Especially in the complex and changeable environment, the introduction of CBAM makes the model more effective in dealing with various interference factors, such as lighting changes, occlusion, etc., to accurately recognize streetlights. 2) The BiFPN structure is used for the Neck part. To make full use of multi-scale feature information and enhance the model's adaptability to complex environments, we innovatively introduced the BiFPN structure into the Neck part of YOLOv5. BiFPN is known for its excellent feature fusion and transfer capabilities, which can help the Neck part more effectively integrate feature maps from different levels and scales. With the introduction of BiFPN, the model can more accurately capture and identify streetlights of different sizes, thus significantly improving the detection performance. Figure 1 is the overall framework of the proposed method.

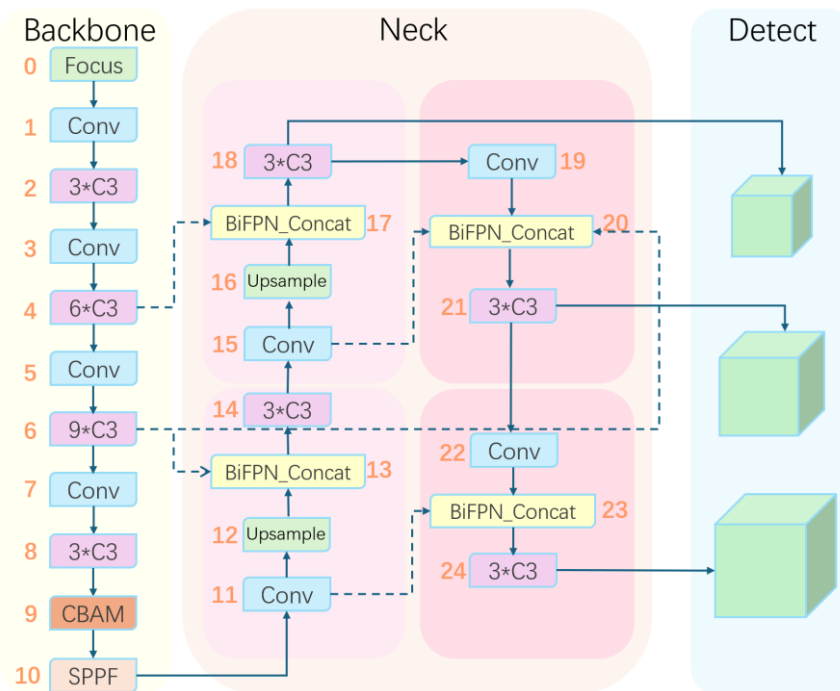


Figure 1. CB-YOLOv5 network structure

2.2 Bidirectional Feature Pyramid Network (BiFPN)

To solve the problem of multi-scale feature fusion in target detection tasks, we propose a novel Bidirectional Feature Pyramid Network (BiFPN). BiFPN's design was inspired by the bidirectional Feature Pyramid (FPN) and Path Aggregation Network (PANet), but has significantly improved upon it. The structure of BiFPN is shown in Figure 2.

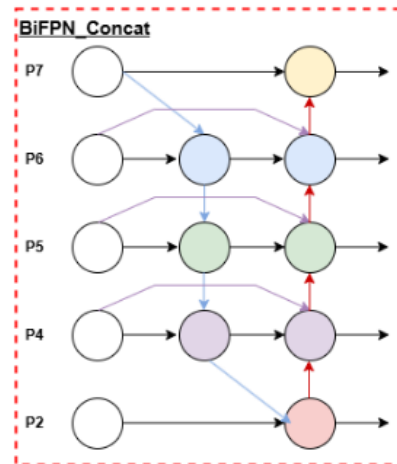


Figure 2. BiFPN structure

Bidirectional cross-scale connection

BiFPN retains the top-down and bottom-up paths in FPN and PANet for more comprehensive feature information transfer. This bidirectional connection allows the network to efficiently capture context information from different scales, thereby enhancing feature representation.

Weighted feature fusion

BiFPN uses weighted feature fusion to fuse input feature layers of different resolutions. These weights are trained by the neural network and used to adjust the importance of different features. In this way, the network can adaptively select and fuse the most relevant features, thus improving the effectiveness and accuracy of feature fusion.

Studies have shown that BiFPN can improve the efficiency and performance of models (M., R., & Q.V., 2020), especially for object detection, instance segmentation, and other related computer vision tasks (Wang, Pan, Lu, & Zhang, 2023). Since there may be various interference factors (such as light changes, shadows, occlusion, etc.) in street view in low-light environment, traditional recognition methods may be greatly affected. BiFPN, through its powerful feature fusion and weight allocation capabilities, can better deal with these interference factors and improve the robustness of identification.

2.3 Convolutional Block Attention Module (CBAM)

CBAM is a simple but effective attention module. It is designed to be flexible, easily integrated into various existing CNN architectures, and is widely used in a variety of visual recognition tasks. This dual attention mechanism enables the model to focus on key channels and spatial positions in input features. This adaptive recalibration mechanism helps the model to focus on key features better and improve recognition performance (Zhu, Lyu, Wang, & Zhao, 2021). Experiments show that CBAM is integrated into different models, and the performance of the model is greatly improved on different classification and detection data sets, which proves the effectiveness of this module (Xudong, Shuai, & Chaoqun, 2022). In the street view we obtained, there are elements that are easy to confuse the streetlight spots, such as camera light spots. The occlusion of leaves and signs, as well as the fuzzy imaging caused by low-light environments, make the recognition task difficult. The CBAM attention module is introduced into the model to help CB-YOLOv5 resist confusing information and focus on useful target objects. The CBAM module is shown in Figure 3.

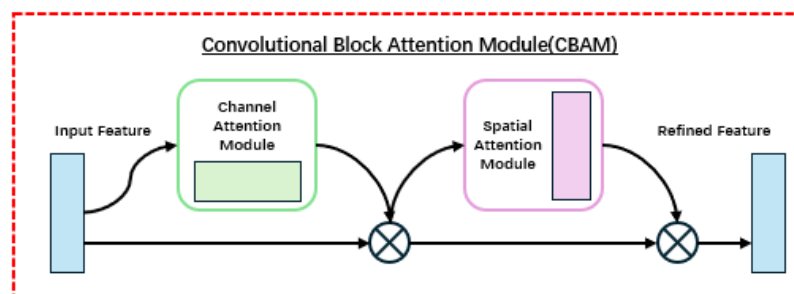


Figure 3. CBAM structure

3. EXPERIMENTS

3.1 Datasets

The experimental dataset includes 200 images which were extracted by a video of a real streetlight monitoring task in Chongqing, China. This video was captured in low-light environment with complex backgrounds. The data was divided into a training set and a validation set in a 7:1 ratio.

3.2 Details

Our experiment was conducted under the Pytorch 2.0.0 framework of Python 3.11.9, using a personal computer with the NVIDIA GeForce RTX 4070 GPU for training and testing. The 640×640 RGB images were used as an input. In the training phase, we use part of the pre-trained model from yolov5s, because CB-YOLOv5 and YOLOv5 share some parts of the Backbone and the Neck, there are many weights that can be transferred from YOLOv5s to CB-YOLOv5, by using these weights we can save a lot of training time.

3.3 Evaluation Indicators

In this paper, precision rate, recall rate, average precision and mean average precision are used as evaluation indexes. The following equations define these indices:

$$P = \frac{TP}{TP + FP} \quad (1)$$

$$R = \frac{TP}{TP + FN} \quad (2)$$

$$AP = \int_0^1 P(R) dR \quad (3)$$

Where P, R, and AP represent precision rate, recall rate, and average precision, respectively. TP, FP, and FN define true positives, false positives, and false negatives. The precision rate reflects the proportion of the real number of targets predicted by the model in all detected targets, the calculation formula is formula (1). Recall rate reflects the proportion of the number of real targets predicted by the model in all real detection targets, the calculation formula is formula (2). Average precision reflects the overall detection precision of the model by calculating the area enclosed by the precision-recall (PR) curves, the calculation formula is formula (3). mAP (Mean Average Precision) is a common index used to measure the overall performance of an object detection model. It considers the Recall and Precision of the model. The higher the mAP value, the better the model performance. mAP_{0.5} indicates the mAP calculated when the IOU threshold is 0.5. This means that a prediction is considered correct when the overlap between the prediction box and the real box is 50% or more.

3.4 Results

We used the pre-trained weights of the YOLOv5s model on the COCO dataset. The learning rate was set to 0.01, and the Stochastic Gradient Descent (SGD) method was used to iterate. In the training period, batch size was set to 16 and run 300 epochs. The training results are shown in Figure 4.

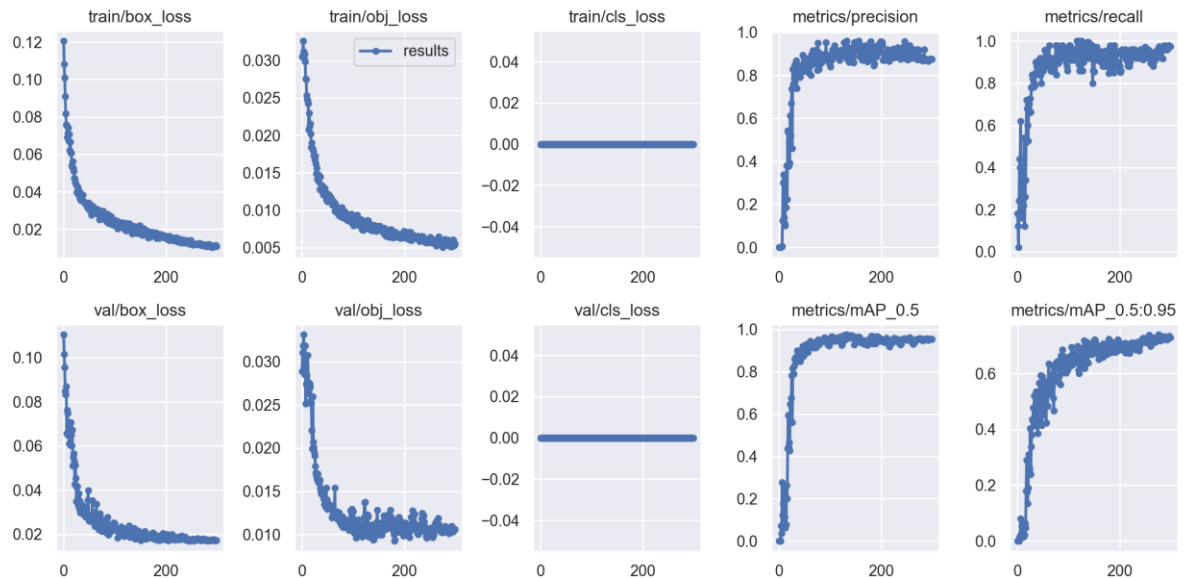


Figure 4. Training results of our model

The CB-YOLOv5 model has strong resistance to occlusion and high interference. It can be seen from the

test results of some streetlights in Figure 5 that even the streetlights shielded by leaves with only a small part of the light spots exposed can be identified, the camera spots very similar to the streetlights can be excluded, and the white light spots on the surrounding buildings can also be avoided. At the same time, in the low-light environment, due to the existence of the CBAM attention mechanism, the CB-YOLOv5 model can also overcome the influence of the harsh environment and focus on the characteristics of streetlights, to achieve accurate recognition.



Figure 5. Some streetlight detection results in complex scenarios.

Comparative tests and evaluations were conducted on different network structures, and the results are shown in Table 1. As shown in Table 1, adding a CBAM module to the backbone for standard YOLOv5s can increase $mAP_{0.5}$ by 16.3%. Based on adding the CBAM module, the BiFPN structure can further improve $mAP_{0.5}$ by 7.2%.

Table 1. Detection results of different models

Models	$mAP_{0.5}$	P	R
YOLOv5s	0.733	0.713	0.709
YOLOv5s+CBAM	0.896	0.898	0.814
YOLOv5s+CBAM+BiFPN	0.968	0.921	0.962

4. CONCLUSIONS

In this paper, we proposed an improved YOLOv5 model to improve the efficiency and performance of streetlight recognition monitoring by combining CBAM and BiFPN. By utilizing CBAM, the loss in feature extraction is reduced, ensuring that more relevant and distinct features are retained. In addition, BiFPN facilitates multi-scale feature fusion and propagation, which is critical for accurately detecting streetlights of different sizes and distances.

Results obtained on our customized dataset show a significant improvement in the detection accuracy of the CB-YOLOv5 model, with $mAP_{0.5}$ exceeding the baseline YOLOv5 model by 23.5%. In this study, the model will be applied to streets with different lighting conditions and road environments to verify its practicability and generalization ability. In addition, the CB-YOLOv5 model will be further applied to the recognition of streetlight poles in future research, and then the recognition performance of the model will be verified again, to generalize it for the recognition of other small objects under complex backgrounds and harsh environment conditions. Based on the two identification results of the streetlight spot and the streetlight pole, the fault point of the lighting facility is investigated, and the brightness threshold can be set subsequently to classify the working performance of the streetlight with high detail.

In addition, this study will continue to explore the model lightweight strategy in the future, in order to ensure recognition accuracy, reduce the consumption of computing resources, and facilitate deployment on edge devices.

ACKNOWLEDGMENTS

The authors are grateful for the Fundamental Research Funds for the Central Universities of Ministry of Education of China (No. 2023CDJKYJH023).

REFERENCES

- M., T., R., P., & Q.V., L. (2020). EfficientDet: Scalable and efficient object detection %J Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 10778-10787.
- Tosti, F., Gagliardi, V., Ciampoli, L. B., Benedetto, A., Threader, S., & Alani, A. M. (2021). *Integration of Remote Sensing and Ground-Based Non-Destructive Methods in Transport Infrastructure Monitoring: Advances, Challenges and Perspectives*. Paper presented at the 2021 IEEE Asia-Pacific Conference on Geoscience, Electronics and Remote Sensing Technology (AGERS).
- Wang, J., Pan, Q., Lu, D., & Zhang, Y. (2023). An Efficient Ship-Detection Algorithm Based on the Improved YOLOv5. *Electronics*, 12(17). doi:10.3390/electronics12173600
- Wu, T.-H., Wang, T.-W., & Liu, Y.-Q. (2021). *Real-Time Vehicle and Distance Detection Based on Improved Yolo v5 Network*. Paper presented at the 2021 3rd World Symposium on Artificial Intelligence (WSAI).
- Xudong, D., Shuai, Y., & Chaoqun, D. (2022). A lightweight vehicles detection network model based on YOLOv5 %J Engineering Applications of Artificial Intelligence. 113.
- Zhang, T., Dai, J., Song, W., Zhao, R., & Zhang, B. (2023). OSLPNet: A neural network model for street lamp post extraction from street view imagery. *Expert Systems with Applications*, 231. doi:10.1016/j.eswa.2023.120764
- Zhang, Y., Guo, Z., Wu, J., Tian, Y., Tang, H., & Guo, X. (2022). Real-Time Vehicle Detection Based on Improved YOLO v5. *Sustainability*, 14(19). doi:10.3390/su141912274
- Zhu, X., Lyu, S., Wang, X., & Zhao, Q. (2021). *TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios*. Paper presented at the Proceedings of the IEEE/CVF international conference on computer vision.

DEVELOPING A MARKOV DECISION PROCESS MODEL TO SUPPORT ESG DECISION-MAKING IN THE INTERNATIONAL CONSTRUCTION MARKET

Yuting Duan¹, Meng Ye^{2*}, Penglin Dai³

1) Postgraduate student. School of Economics and Management, Southwest Jiaotong University, Chengdu, Sichuan, China. Email: duanyt0731@163.com

2) Ph.D., Assistant Professor. School of Management Science and Real Estate, Chongqing University, Chongqing, China. Email: mengye@cqu.edu.cn

3) Ph.D., Associate Professor. School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu, Sichuan, China. Email: penglindai@swjtu.edu.cn

Abstract: International contractors face complex decision-making challenges under uncertain and rapidly changing institutional environments. This study addresses these challenges by proposing a Markov decision process (MDP) model to optimize Environmental, Social, and Governance (ESG) strategies. The proposed model consists of four key components: state, action, transfer probability, and reward function. Data collected through structured surveys on institutional pressures, ESG actions, and anticipated future pressures were used to measure the state, action, and transfer probability, respectively. ESG performance metrics, along with the associated costs and benefits for construction companies, were gathered to train a neural network model for representing the reward function. The value iteration method was used to derive optimal ESG decision-making strategies in the international construction market. The findings reveal that the MDP framework effectively identifies optimal ESG actions for contractors, providing practical insights to maximize rewards and adapt to dynamic institutional conditions. This approach supports contractors in achieving their operational and strategic goals in international markets.

Keywords: Institutional pressure, ESG decision-making, International contractors, Markov Decision Process (MDP)

1. INTRODUCTION

In recent years, the global landscape has been marked by increasing uncertainty and rapid changes (Worley & Jules, 2020), which posed significant challenges for businesses operating in volatile environments. International contractors, in particular, face heightened pressures due to the complex and dynamic nature of the international construction market, which involves large-scale investments, lengthy project timelines, and a variety of institutional environments (Jung et al., 2012). Operating across borders, these contractors are faced with evolving institutional pressures, making adaptability and resilience essential to project success (Hult et al., 2020; McCarthy et al., 2010). Unlike static environments, cross-border settings require international contractors to continuously respond to changing institutional demands, from local regulatory adjustments to shifting societal expectations. As these pressures evolve, they necessitate flexible and responsive management strategies that ensure both compliance and operational efficiency.

To address these complexities, international contractors are increasingly adopting Environmental, Social, and Governance (ESG) strategies as part of their broader approach to managing institutional pressures. ESG strategies not only help firms align with regulatory requirements and societal expectations but also enhance their legitimacy and long-term sustainability (Ramanujam, 2009). Legitimacy theory posits that organizations seek approval from society by demonstrating behaviors that align with socially constructed norms and values (Suchman & M., 1995). In the context of ESG, legitimacy implies that firms can enhance their corporate reputation and stakeholder trust by proactively addressing environmental and social concerns (Ye et al., 2022). Moreover, in highly regulated or publicly scrutinized sectors, ESG practices serve as a vital means of securing legitimacy and managing reputational risks, particularly in turbulent environments (Campbell, 2007; Wang et al., 2019).

However, managing ESG in a dynamic and uncertain institutional environment adds considerable complexity to the decision-making process. Institutional pressures are multifaceted, with both external and internal dimensions influencing firms' strategies. Traditional classifications describe institutional pressures as coercive, normative, and mimetic. Coercive pressures arise from formal regulations and compliance demands, normative pressures reflect societal values and professional standards, while mimetic pressures stem from the tendency of firms to imitate successful peers in uncertain environments (DiMaggio & Powell, 1983). Building on this, Borini et al. (2018) differentiate between external institutional pressures—such as the host country's regulatory and cultural demands—and internal ones, which stem from corporate culture and strategic objectives. This dual-layered pressure landscape further complicates the decision-making process for international contractors, therefore understanding how to navigate these pressures has become a crucial element of project management in cross-border settings (Ho et al., 2023).

This complex landscape has led to a growing recognition that traditional decision-making frameworks often fall short in addressing the uncertainties international contractors face. While current literature provides foundational insights into ESG decision-making (Brigham et al., 2023; Molin et al., 2023), it often focuses on general guidelines or reactive approaches to institutional pressures, offering limited support for proactive ESG decision-making in today's increasingly uncertain environment for international contractors. To address this need, this study introduces a

Markov Decision Process (MDP) framework as a tool for ESG decision-making in uncertain institutional contexts. Widely applied in sequential decision-making under uncertainty, the MDP model is particularly suited to complex environments, enabling firms to evaluate a range of potential actions and their consequences across different contexts (Alagoz et al., 2010; He & Jiang, 2018). By incorporating probabilities of state transitions and reward functions, the MDP approach provides a structured, adaptive model that supports continuous ESG strategy adjustments.

The purpose of this study is twofold: first, to develop a decision-making model that identifies the optimal ESG strategies for international contractors navigating periods of volatility; and second, to clarify the specific conditions under which certain actions maximize ESG outcomes amid varying institutional pressures. By introducing a proactive, adaptive ESG decision-making framework, this research moves beyond traditional reactive approaches, addressing critical gaps in the literature and offering actionable insights to guide contractors in complex, dynamic global environments.

2. THE PROPOSED MDP MODEL

2.1 Model Assumptions

There are three assumptions for the proposed MDP model shown as follows:

(1) Assumption 1: Markov property in ESG decision-making. According to the Markov property, a foundational principle of the MDP framework, it is assumed that the transition from one state of institutional pressure to another depends solely on the current state and the chosen ESG action, rather than the sequence of previous states. This assumption implies that the future state of institutional pressures faced by an international contractor is determined only by the current level of pressures and the ESG strategy implemented at that moment. This simplifies the decision-making process, enabling the use of MDP to model and optimize the contractor's ESG strategy in a dynamic environment.

(2) Assumption 2: The implementation of ESG practices by international contractors under varying institutional pressures is a decision problem. Confronted with diverse and complex institutional environments, contractors must determine how to implement ESG. It can be thus assumed that ESG implementation is a decision problem in which contractors select their response strategy. This assumption frames ESG implementation as an action, with the key focus on identifying the "best" course of action. In essence, contractors assess the potential outcomes of their decisions to guide their actions.

(3) Assumption 3: Observability of the external environment. The external environment is assumed to be fully observable to international contractors. This means the contractors' perception of the institutional environment aligns precisely with the actual external conditions, allowing for clear decision-making based on the current state information. Consequently, decisions can be made based on the known and observable state of the environment without requiring estimation or inference about hidden variables. This full observability simplifies the modeling process, enabling direct optimization of ESG strategies within the model.

2.2 Model Components

The MDP is a mathematical model for decision-making in uncertain environments, consisting of four components, State (S), Action (A), Transfer Probability (P), and Reward Function (R). The set of states is all possible situations faced by the decision maker, the set of actions is all possible actions that the decision maker can take in each state, the transfer probability is the probability of transferring to the next state after each action is taken in each state, and the reward function is the instantaneous reward obtained after each action is taken in each state. MDP can be solved for the optimal strategy, i.e., what action can be taken to maximize the desired reward in each state.

State (S) represents the level of internal (e.g., from employees and parent companies) and external (e.g., from competitors and society) institutional pressures faced by an international contractor. By capturing these current pressure levels, we ensure that the state contains all the necessary information for predicting future outcomes and guiding decisions, without the need to consider past states, which aligns with the Markov property.

To quantify these states, we map internal and external pressures onto a two-dimensional coordinate system, with internal pressures (IIP) on the X-axis and external pressures (EIP) on the Y-axis, where each coordinate represents the intensity of pressures faced by contractors. Then, the international contractors' state set, $S = \{s_1, s_2, \dots, s_9\}$, defines nine unique combinations of institutional pressure levels, categorizing the contractor's environment into distinct states.

Action (A) in this model represents the strategies an international contractor might use to address ESG responsibilities across various states. Actions are organized into three dimensions: Environmental (E), Social (S), and Governance (G), capturing a broad range of potential ESG behaviors.

To measure these actions, each ESG dimension is further divided into levels of intensity, from low to high, resulting in 27 unique actions, each representing a specific combination of ESG strategy intensities. This action set, denoted as $A = \{a_1, a_2, \dots, a_{27}\}$, enables analysis of how contractors adjust their ESG strategies in response to different institutional pressures.

Transfer Probability (P) in this model represents the likelihood that an international contractor will transition between states of institutional pressure after implementing a specific ESG action, capturing the uncertainty and dynamic nature of the environment. This model assumes transitions depend only on the current state and chosen

action, in line with the Markov property.

To estimate these probabilities, we conducted a survey among international contractors to collect data on their current state, implemented ESG actions, and anticipated state transitions. Using the Law of Large Numbers, we approximated transition probabilities by calculating the frequency of state changes. Specifically, $p_{ij}(a_k)$, the probability of moving from state s_i to s_j after action a_k , is determined by the ratio of respondents who transitioned from s_i to s_j following a_k to the total respondents who took a_k while in s_i .

Reward Function (R) assigns a value to each state-action pair, indicating the desirability of a particular action in a given state for an international contractor. This MDP model captures how decisions in response to current pressures influence future outcomes, making it suitable for optimizing ESG strategies in dynamic and uncertain environments.

To quantify R , we first calculate the Cost-Benefit Ratio (CBR) as shown in formula (1), for a sample of construction-related companies and combine these values with their ESG scores across Environmental, Social, and Governance dimensions. These combined CBR and ESG scores serve as input data to train a neural network model, establishing a predictive relationship between ESG scores and the reward R (i.e., the CBR).

$$CBR = \frac{Benefit}{Cost} \quad (1)$$

Once the CBR values are calculated, the next step involved processing this data and training the neural network model to establish a predictive relationship between the ESG scores and the reward R . The ESG scores for various companies, spanning the Environmental, Social, and Governance dimensions, were used as input variables (x_1, x_2, x_3) in the neural network model. The output variable (y) represented the reward R , which is quantified by the CBR. As indicated in formula (2), $f(\cdot)$ denotes the function learned by the neural network. Once trained and validated, the model is applied to predict the reward R for international contractors using ESG scores collected from survey responses. By feeding the survey-derived ESG scores into the model, the corresponding reward R values are generated, providing an estimate of the expected rewards that each contractor could achieve by implementing specific ESG strategies under varying institutional pressures.

$$y = f(x_1, x_2, x_3) \quad (2)$$

2.3 MDP Model Optimization and Policy Derivation

To evaluate the optimal decision-making process for international contractors concerning ESG, we implemented a MDP using value iteration. The value iteration algorithm was employed to calculate the optimal value function $V^*(s)$. This method iteratively updates the value of each state based on the Bellman optimality equation (referenced as formula (3)), where γ is the discount factor, s represents the current state, a is the action taken, and s' denotes the next state, and l , representing the number of pressure states, is 9. The process continues until the change in the value function, δ , is less than a small threshold, indicating convergence. In this study, γ was set to 0. This choice reflects the immediate nature of decision-making in a rapidly changing environment, where future rewards are considered less relevant, thereby placing emphasis on current outcomes. This assumption is aligned with the reality faced by international contractors operating in volatile markets where long-term planning is overshadowed by immediate pressures and uncertainties.

$$V^*(s) = \max_{a \in A} \sum_{i=1}^l P(s' | s, a) [(R(s, a, s') + \gamma V^*(s))] \quad (3)$$

After achieving convergence in the value iteration process, the optimal policy π^* was determined. The policy dictates the best action a to take in each state s , maximizing the expected reward. The policy was derived by selecting the action that maximizes the value function:

$$\pi^* = \arg \max \sum_{i=1}^l P(s' | s, a) [(R(s, a, s') + \gamma V^*(s))] \quad (4)$$

The results of the value iteration algorithm yielded the optimal value function V^* and the corresponding optimal policy π^* for each state.

3. DATA ANALYSIS, FINDINGS AND DISCUSSIONS

3.1 Data Collection

To support the calculation and derivation of the MDP components, especially for the transfer probability P and the reward function R as outlined in the previous stages, a comprehensive data collection process was undertaken.

3.1.1 Collection of S , A , and P for International Contractors

Specifically, for the determination of the state S , action A and transfer probability P , the data was gathered through a survey targeting international contractors. The snowball method was employed to conduct the questionnaire

survey from December 2023 to January 2024. A total of 235 questionnaires were distributed, with 162 responses received, yielding a response rate of 68.94%. Of these, 125 questionnaires were fully completed, resulting in an effective response rate of 77.16%. The survey was designed to be anonymous, and participation was voluntary.

3.1.2 Collection of CBR for Model Training Outputs (Y)

Data needed for calculating the reward function R was gathered from various sources, with the CBR serving as the key metric. Cost indicators were sourced from CSR reports and the CSMAR database, covering items such as environmental protection taxes, energy consumption, charitable contributions, audit fees, and R&D expenses, all converted into monetary values to assess ESG costs clearly. Benefit indicators were measured using non-financial metrics such as environmental honors, certifications (e.g., ISO9001, ISO14001), employee satisfaction, corporate reputation, innovation achievements, and governance outcomes, as detailed in Table 1. This approach provides a comprehensive CBR by balancing financial costs against qualitative benefits, enabling an effective evaluation of ESG strategies. The CBR then serves as the dependent variable in this analysis, offering a balanced assessment of ESG strategy success.

Table 1. Cost-benefit indicators

Indicator type	Indicators
Cost	Environmental protection tax
	Energy consumption
	Charitable input
	Audit fee
	Innovation research and development expenses
Benefit	Environmental honors or awards
	Whether passed ISO9001 and ISO14001 certification
	Employee satisfaction
	Corporate reputation
	Innovation achievement
	Governance achievement

3.1.3 Collection of ESG Data for Model Training Inputs (X)

To enhance the predictive modeling described in section 2.2, the dataset was further refined with comprehensive ESG scores from the Bloomberg database for various companies, which provided comprehensive ESG scores for various companies. This dataset was crucial for establishing the relationship between a company's ESG performance and the corresponding reward R . To ensure the relevance and accuracy of the analysis, a rigorous data screening process was applied. This process involved excluding financial firms, such as those in the banking, securities, and insurance sectors, due to their unique characteristics that could skew the analysis. Additionally, companies with special listing statuses (e.g., ST or *ST) were removed. The final dataset, focused on companies related to the construction industry, consisted of 60 listed firms in China for the year 2022. This dataset was then used to train a neural network model designed to predict the international contractors' reward R based on the collected ESG scores.

3.2 Data analysis and results of the components

3.2.1 Visualization and Analysis of S , A , and P for International Contractors

Using data from 125 completed questionnaires, we analyzed the relationships between states (S), actions (A), and transfer probabilities (P) for international contractors. Figure 1 provides a bubble plot illustrating the frequency and distribution of actions across nine institutional pressure states. The size of each bubble represents the relative frequency of a specific action within a particular state, while the color intensity indicates the magnitude of the corresponding values.

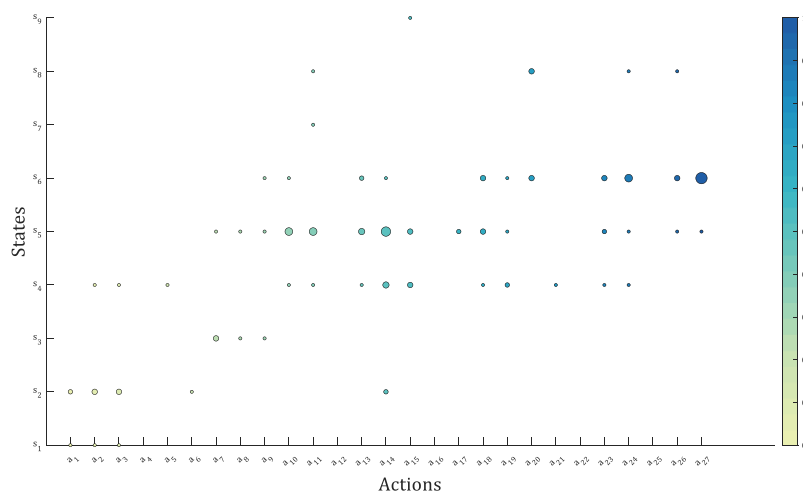


Figure 1. Distribution of ESG actions across institutional pressure states

Figure 1 reveals distinct patterns in the adoption of ESG strategies across different institutional states. For lower-pressure states (e.g., s_1 and s_2), fewer and less intensive actions are typically implemented, as evidenced by smaller and lighter-colored bubbles concentrated in the left portion of the action axis. This suggests a cautious approach to ESG implementation when institutional pressures are minimal. In contrast, states under medium to high institutional pressures (e.g., s_5 and s_6) exhibit larger bubbles across a broader range of actions, highlighting increased adoption and intensity of ESG strategies in response to elevated pressures. For instance, state s_6 demonstrates the most diverse action distribution, with actions a_{23} , a_{24} , and a_{27} showing particularly high frequencies.

Following this, the transfer probabilities were recalculated to better visualize the dynamics of state transitions in response to specific ESG actions. The updated state transition network is depicted in Figure 2, where the red nodes represent the nine institutional pressure states (s_1 through s_9), and the blue arrows indicate all possible transitions between these states based on the contractors' actions. The probability of each transition is labeled on the arrows alongside the corresponding ESG actions. The results showed that while ESG actions are necessary for managing institutional pressures, their effectiveness is closely tied to both the starting pressure state and the intensity of the actions taken. (1) Contractors in lower-pressure states may experience an increase in pressure if they implement low intensity ESG actions. For example, contractors in s_2 , despite being in a relatively low-pressure state, transitioned to higher-pressure states like s_9 after adopting weaker actions such as a_2 , indicating that insufficient ESG engagement in these contexts can lead to heightened external pressures. This suggests that contractors in low pressure environments face the risk of pressure escalation if they do not take robust actions to meet external expectations. (2) In high-pressure states, such as s_9 , the analysis reveals a distinct trend: even when contractors adopt relatively strong ESG actions, there is often little movement towards lower pressure states. For instance, contractors in s_9 who implemented higher-intensity actions like a_{15} frequently slightly transitioned to s_6 .

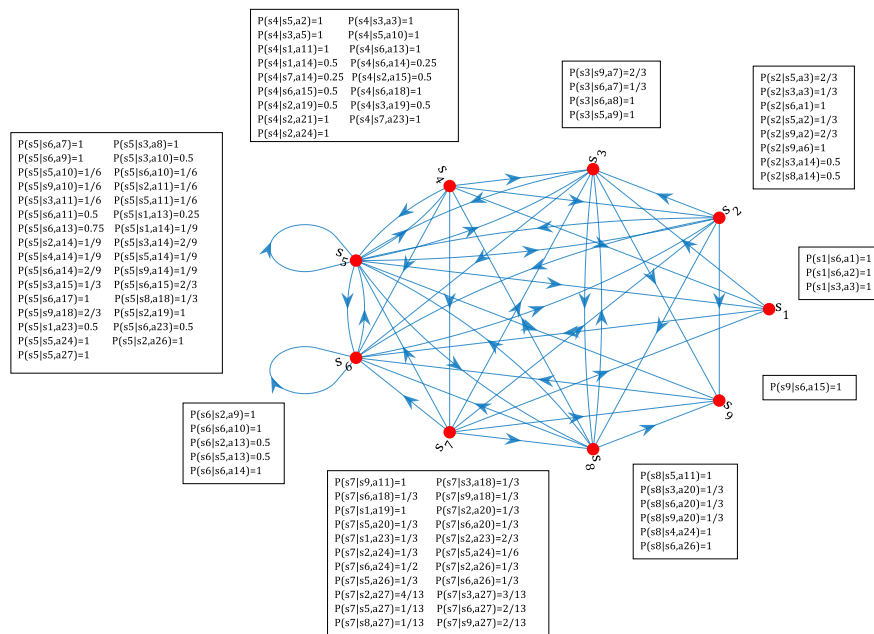


Figure 2. Transition visualization for international contractors' states

3.2.2 Neural Network Modeling for Reward Function Prediction

After calculating the transfer probabilities, the next step was to process and analyze the reward function R , modeled using a neural network. The input data consisted of x (ESG scores) and y (CBR values), which were used to train the network. As shown in Figure 3, the neural network architecture includes an input layer, a hidden layer, and an output layer. The input layer accepts three values corresponding to the ESG dimensions. These inputs are passed to the hidden layer, which contains 10 neurons. Each neuron in the hidden layer computes a weighted sum of the inputs using a weight matrix W and a bias b . The resulting output is passed through an activation function (represented by the curved symbol) to introduce non-linearity, allowing the network to learn more complex patterns. The weight matrix W in the hidden layer has dimensions of 3×10 , corresponding to the three inputs and ten neurons. After the hidden layer processes the data, the output is passed to the final output layer, which consists of a single neuron. The output layer applies its own weight matrix and bias before generating the final prediction.

The neural network was trained using the Levenberg-Marquardt backpropagation algorithm. During training,

the weights and biases were adjusted to minimize the error between the predicted and actual CBR values, optimizing the model's accuracy. This neural network structure was specifically designed to capture the relationship between ESG scores and CBR values.

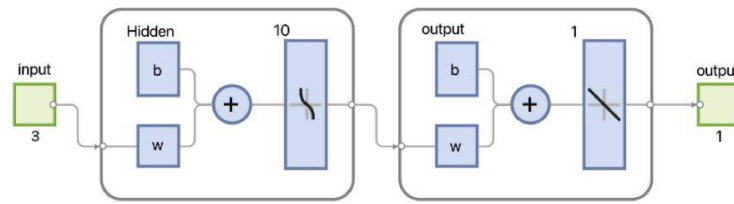


Figure 3. Two-layer feedforward network

3.2.3 Model Validation and Performance Evaluation

To ensure robust results, the data set is split into three train, validation, and test data sets with a ratio of [14:3:3]. Thus, the algorithm is trained in randomly chosen data from 42 companies. Then, the model is validated in the remaining randomly chosen 9 companies' data and tested in the remaining 9 companies' data. The results depicted in figure 4 strongly validate the effectiveness of this neural network model across various datasets. The model achieved a commendable correlation coefficient R of 0.765 on the training set, reflecting a robust alignment between the predicted and actual data. In this context, an R value above 0.7 generally indicates good predictive power (Bansal et al., 2024). This high R value indicates that the model has successfully internalized the underlying patterns within the data, affirming its capacity to accurately predict outcomes based on ESG scores. Although the validation set yielded a slightly lower R value of 0.590, it still represents a meaningful predictive capability, as values above 0.5 generally indicate moderate to strong correlation (Bansal et al., 2024), demonstrating that the model retains a significant degree of accuracy when applied to unseen data. The validation set is used to assess the model's performance on data it has not been trained on, thereby providing an estimate of its generalization ability. Importantly, the model's performance on the test set, with an R value of 0.736, underscores its strong generalizability, proving its ability to make reliable predictions even when confronted with entirely new data. When considering the combined data, the model achieves an overall R value of 0.706, further reinforcing its efficacy in capturing the intricate relationship between ESG scores and the cost-benefit ratio (CBR). These results collectively underscore the model's robustness and reliability.

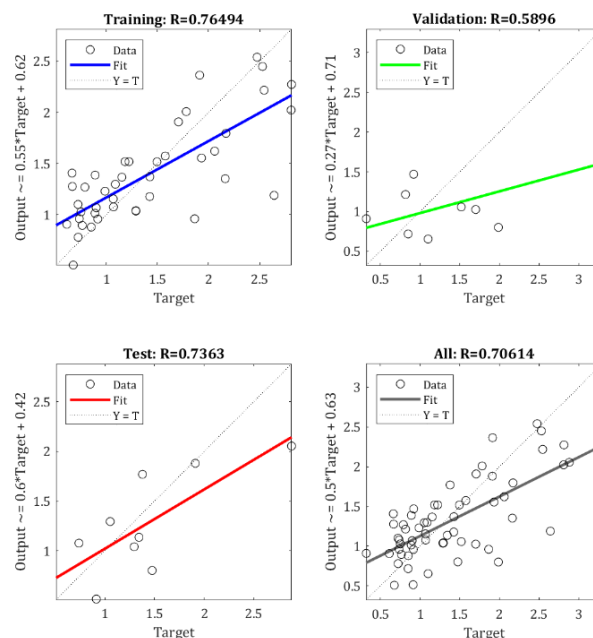


Figure 4. Regression fit for training, validation, test, and all data

3.3 MDP Model Results and Discussions

The contractors' rewards were calculated using the questionnaire data on ESG implementation and a trained neural network reward prediction model. These rewards, serving as key inputs, were incorporated into the MDP framework to optimize ESG decision-making strategies. The value iteration method was applied to solve the MDP, producing the results visualized in Figure 5. This figure represents state transitions, optimal actions, and their

associated rewards. States are depicted as nodes corresponding to different institutional pressure levels, while arrows indicate possible transitions resulting from specific ESG actions. The red arrows highlight the optimal strategies that lead to the highest rewards, with labels indicating the actions taken and their resulting outcomes.

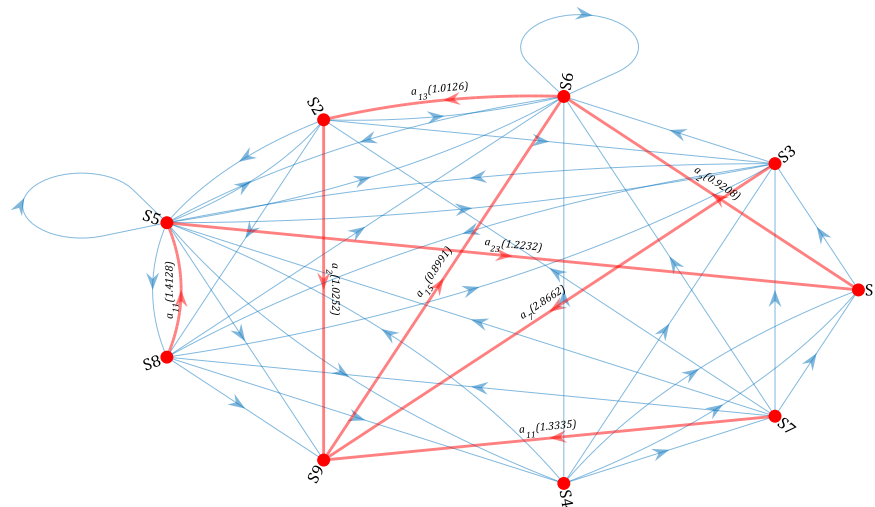


Figure 5. Policy visualization based on MDP

The analysis reveals distinct patterns in optimal ESG strategies across different institutional contexts. In high-pressure environments, such as s_9 , where both internal and external pressures are elevated, the optimal strategy involves significant investments across all ESG dimensions. This balanced approach helps contractors manage regulatory and societal expectations effectively. Conversely, in low-pressure environments like s_1 , optimal strategies minimize investments in social (S) and environmental (E) dimensions while maintaining steady governance (G) efforts. This strategy ensures compliance and stability with minimal resource expenditure.

Internal pressures significantly influence ESG decision-making. In states with low internal pressures, such as s_1 , s_2 , and s_3 , optimal strategies focus on governance stability while limiting environmental investments. For states with high internal pressures, such as s_7 , s_8 , and s_9 , stronger governance investments are required, alongside moderate environmental and carefully controlled ESG allocations. These strategies address the need for robust governance to maintain operational stability and stakeholder confidence in high-pressure scenarios.

External pressures also play a critical role. In low external pressure states, such as s_1 , s_4 , and s_7 , minimal investments in ESG are advised while maintaining governance as a baseline. In contrast, high external pressure states, such as s_3 , s_6 , and s_9 , require increased ESG investments to address heightened public scrutiny and regulatory demands. ESG emerges as the most flexible dimension, allowing contractors to scale investments based on external conditions while ensuring steady environmental and governance standards.

The model further identifies mid-level pressure states, such as s_5 , as intermediaries with diverse incoming and outgoing transitions. These states demand adaptable ESG strategies that reflect their dynamic nature. Governance remains a critical baseline across all states, but ESG investments exhibit greater flexibility, allowing contractors to adjust efforts based on external pressures and public scrutiny.

In summary, the MDP model demonstrates the importance of tailoring ESG strategies to specific institutional contexts. Low-pressure states benefit from governance stability and cost-efficient approaches, while high-pressure states require balanced, multifaceted ESG efforts. The model provides actionable insights into how international contractors can align their ESG initiatives with institutional pressures, optimizing resource allocation and enhancing long-term sustainability and performance.

4. CONCLUSION, IMPLICATIONS AND LIMITATIONS

This study developed and applied a Markov Decision Process (MDP) model to identify the optimal ESG strategies for international contractors, determining which actions maximize rewards under varying levels of institutional pressure. The findings suggest that baseline investments in environmental and governance responsibilities are essential across all scenarios for ensuring compliance and organizational stability, while ESG investments remain flexible, allowing contractors to adapt based on external scrutiny and regulatory demands. In low-pressure environments, contractors are advised to focus on operational efficiency with minimal ESG spending, particularly in social and environmental areas, while maintaining governance standards. In high-pressure contexts, however, a more balanced ESG investment across all dimensions is necessary to meet heightened regulatory expectations and sustain operations. This approach enables international contractors to balance compliance, public expectations, and cost-effectiveness, optimizing resource allocation across diverse institutional environments.

This research makes substantial contributions to the application of ESG strategy optimization, particularly under

varying institutional pressures, which are central to the decision-making environment in the international construction market. One major contribution is the introduction of an innovative, adaptable framework using a MDP model. This model not only evaluates ESG actions but also deeply integrates the critical influence of institutional pressures, with components like state transitions and reward calculations. This comprehensive approach enables international contractors to tailor their ESG investments to specific types and intensities of pressure, pushing beyond prior studies by emphasizing the role of regulatory and societal dynamics in shaping effective ESG practices. In addition, by uncovering the intricate relationship between ESG investments and resulting rewards, our findings underscore how specific action combinations yield optimal returns, thereby offering contractors a strategic approach to resource allocation that maximizes both efficiency and impact.

The study also offers practical implications for international contractors and businesses operating in complex institutional settings. By identifying the essential role of foundational investments in environmental and governance aspects across all pressure levels, this research provides actionable guidance for sustaining baseline ESG performance. At the same time, it highlights the flexibility of ESG investments, allowing contractors to strategically adjust this dimension based on external scrutiny and regulatory pressures. This structured and adaptable approach to ESG decision-making positions businesses to better navigate evolving ESG demands, optimize performance, and maintain competitiveness in a landscape increasingly shaped by stakeholder expectations and regulatory complexities.

Despite these contributions, this study has several limitations. First, the model assumes complete observability of the institutional environment, which may oversimplify the dynamic and often opaque nature of regulatory landscapes. Future research could address this by incorporating Partially Observable Markov Decision Process (POMDP) to better capture the incomplete or delayed information that often characterizes real-world regulatory dynamics. Second, the model primarily focused on quantitative measures of reward, such as the cost-benefit ratio. Future studies can broaden this scope by incorporating qualitative outcomes, such as reputational impact or stakeholder trust, which are crucial in evaluating ESG performance but more challenging to quantify. Third, although the model is tailored to international contractors, it has broader applicability across industries facing complex ESG demands. Future studies could extend this framework to other sectors, yielding valuable insights into how diverse industries manage ESG integration amid evolving institutional pressures.

ACKNOWLEDGMENTS

This work was supported by National Natural Science Foundation of China (NSFC) [Grant ID: 72301045], and Natural Science Foundation of Sichuan Province of China [Grant ID: 2023NSFSC1024].

REFERENCES

- Agudelo, M. A. L., Johannsdottir, L., & Davidsdottir, B. (2020). Drivers that motivate energy companies to be responsible. A systematic literature review of Corporate Social Responsibility in the energy sector [Review]. *Journal of Cleaner Production*, 247, Article 119094. <https://doi.org/10.1016/j.jclepro.2019.119094>
- Alagoz, O., Hsu, H., Schaefer, A. J., & Roberts, M. S. (2010). Markov Decision Processes: A Tool for Sequential Decision Making under Uncertainty [Article]. *Medical Decision Making*, 30(4), 474-483. <https://doi.org/10.1177/0272989x09353194>
- Anthonissen, J., Van den Bergh, W., & Braet, J. (2016). Review and environmental impact assessment of green technologies for base courses in bituminous pavements [Review]. *Environmental Impact Assessment Review*, 60, 139-147. <https://doi.org/10.1016/j.eiar.2016.04.005>
- Arminen, H., Puumalainen, K., Patari, S., & Fellnhöfer, K. (2018). Corporate social performance: Inter-industry and international differences [Article]. *Journal of Cleaner Production*, 177, 426-437. <https://doi.org/10.1016/j.jclepro.2017.12.250>
- Athanasopoulou, A., & Selsky, J. W. (2015). The Social Context of Corporate Social Responsibility: Enriching Research With Multiple Perspectives and Multiple Levels [Review]. *Business & Society*, 54(3), 322-364. <https://doi.org/10.1177/0007650312449260>
- Bansal, H., Chinagundi, B., Rana, P. S., & Kumar, N. (2024). Time series generative adversarial network for muscle force prognostication using statistical outlier detection [Article; Early Access]. *Expert Systems*. <https://doi.org/10.1111/exsy.13653>
- Borini, F. M., MacLennan, M. L. F., Pereira, R. M., Pavan, K. R., & Junior, F. H. (2018). Green and social certifications make up for home market underdeveloped institutional environment? Evidences from Brazilian subsidiaries. *Transnational Corporations Review*, 10.
- Boubakri, N., El Ghouli, S., Guedhami, O., & Wang, H. (2021). Corporate social responsibility in emerging market economies: Determinants, consequences, and future research directions [Article]. *Emerging Markets Review*, 46, Article 100758. <https://doi.org/10.1016/j.ememar.2020.100758>
- Brigham, M., Kiosse, P. V., & Otley, D. (2023). A structured framework to understand CSR decision-making: A case study of multiple rationales [Article]. *European Management Journal*, 41(3), 345-353. <https://doi.org/10.1016/j.emj.2022.04.001>
- Bundy, J., Vogel, R. M., & Zachary, M. A. (2018). Organization-stakeholder fit: A dynamic theory of cooperation,

- compromise, and conflict between an organization and its stakeholders [Article]. *Strategic Management Journal*, 39(2), 476-501. <https://doi.org/10.1002/smj.2736>
- Campbell, J. L. (2007). Why would corporations behave in socially responsible ways? an institutional theory of corporate social responsibility. *Academy of Management Review*, 32(3), 946-967. <https://doi.org/10.5465/amr.2007.25275684>
- DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, 147-160.
- Haji, A., & Abdifatah. (2013). Corporate social responsibility disclosures over time: evidence from Malaysia. *Managerial Auditing Journal*, 28(7), 647-676.
- He, Z., & Jiang, W. (2018). An evidential Markov decision making model [Article]. *Information Sciences*, 467, 357-372. <https://doi.org/10.1016/j.ins.2018.08.013>
- Ho, S. P., Dahal, R., & Tserng, H.-P. (2023). A Contingency Model of Strategic Responses to the Institutional Challenges in Emerging Countries: Evidence and Findings from Least Developed Countries [Article]. *Journal of Management in Engineering*, 39(4), Article 04023015. <https://doi.org/10.1061/jmenea.Meeng-5102>
- Hult, G. T. M., Gonzalez-Perez, M. A., & Lagerstrom, K. (2020). The theoretical evolution and use of the Uppsala Model of internationalization in the international business ecosystem [Article; Proceedings Paper]. *Journal of International Business Studies*, 51(1), 38-49. <https://doi.org/10.1057/s41267-019-00293-x>
- Jamali, D., & Mirshak, R. (2007). Corporate Social Responsibility (CSR): Theory and practice in a developing country context [Article]. *Journal of Business Ethics*, 72(3), 243-262. <https://doi.org/10.1007/s10551-006-9168-4>
- Jung, H.-J., & Kim, D.-O. (2016). Good Neighbors but Bad Employers: Two Faces of Corporate Social Responsibility Programs [Article]. *Journal of Business Ethics*, 138(2), 295-310. <https://doi.org/10.1007/s10551-015-2587-3>
- Jung, W., Han, S. H., Koo, B., & Jang, W. (2012). Which strategies are more effective for international contractors during boom and recession periods? [Article]. *Journal of Management in Engineering*, 28(3), 281-290. [https://doi.org/10.1061/\(ASCE\)ME.1943-5479.0000087](https://doi.org/10.1061/(ASCE)ME.1943-5479.0000087)
- McCarthy, I. P., Lawrence, T. B., Wixted, B., & Gordon, B. R. (2010). A MULTIDIMENSIONAL CONCEPTUALIZATION OF ENVIRONMENTAL VELOCITY [Article]. *Academy of Management Review*, 35(4), 604-626. <https://doi.org/10.5465/amr.2010.53503029>
- Molin, M., Pizzol, L., Pesce, M., Maura, A., Civiero, M., Gritti, E., Giotto, S., Ferri, A., Liguoro, L., Bagnoli, C., & Semenzin, E. (2023). An integrated decision-making framework for corporate sustainability [Article]. *Corporate Social Responsibility and Environmental Management*, 30(3), 1145-1160. <https://doi.org/10.1002/csr.2410>
- Park, B. I., & Cave, A. H. (2018). Corporate social responsibility in international joint ventures: Empirical examinations in South Korea [Article]. *International Business Review*, 27(6), 1213-1228. <https://doi.org/10.1016/j.ibusrev.2018.05.003>
- Prado-Lorenzo, J.-M., & Garcia-Sanchez, I.-M. (2010). The Role of the Board of Directors in Disseminating Relevant Information on Greenhouse Gases [Article]. *Journal of Business Ethics*, 97(3), 391-424. <https://doi.org/10.1007/s10551-010-0515-0>
- Suchman, & M., C. (1995). Managing Legitimacy: Strategic and Institutional Approaches. *Academy of Management Review*, 20(3), 571-610.
- Tolmie, C. R., Lehnert, K., & Zhao, H. (2020). Formal and informal institutional pressures on corporate social responsibility: A cross-country analysis [Article]. *Corporate Social Responsibility and Environmental Management*, 27(2), 786-802. <https://doi.org/10.1002/csr.1844>
- Udayasankar, K. (2008). Corporate Social Responsibility and Firm Size [Article]. *Journal of Business Ethics*, 83(2), 167-175. <https://doi.org/10.1007/s10551-007-9609-8>
- Villo, S., Halme, M., & Ritvala, T. (2020). Theorizing MNE-NGO conflicts in state-capitalist contexts: Insights from the Greenpeace, Gazprom and the Russian state dispute in the Arctic [Article]. *Journal of World Business*, 55(3), Article 101068. <https://doi.org/10.1016/j.jwb.2019.101068>
- Wang, S., Wang, H., & Wang, J. (2019). Exploring the effects of institutional pressures on the implementation of environmental management accounting: Do top management support and perceived benefit work? *Business Strategy and the Environment*, 28(1), 233-243. <https://doi.org/10.1002/bse.2252>
- Worley, C. G., & Jules, C. (2020). COVID-19's Uncomfortable Revelations About Agile and Sustainable Organizations in a VUCA World. *The Journal of Applied Behavioral Science*, 56(3), 279-283.
- Ye, M., Lu, W., & Xue, F. (2022). Impact of Institutional Distance on Environmental and Social Practices in Host Countries: Evidence from International Construction Companies. *Journal of construction engineering and management*(1), 148.
- Yin, J. (2017). Institutional Drivers for Corporate Social Responsibility in an Emerging Economy: A Mixed-Method Study of Chinese Business Executives [Article]. *Business & Society*, 56(5), 672-704. <https://doi.org/10.1177/0007650315592856>

INTELLIGENT DESIGN FOR COMPONENT SIZE OF REINFORCED CONCRETE FRAME STRUCTURES USING DIFFUSION MODELS

Yi Gu¹, Sizhong Qin², Wenjie Liao³, Siqi Chen⁴ and Xinzheng Lu^{5,*}

1) Ph.D. Student, Department of Civil Engineering, Tsinghua University, Beijing, China. Email: guy22@mails.tsinghua.edu.cn

2) Ph.D. Student, Department of Civil Engineering, Tsinghua University, Beijing, China. Email: qsz23@mails.tsinghua.edu.cn

3) Ph.D., Assoc. Res., School of Civil Engineering, Southwest Jiaotong University, Chengdu, China. Email: liaowj@swjtu.edu.cn

4) Ph.D. Student, Department of Civil Engineering, Tsinghua University, Beijing, China. Email: chen-sq24@mails.tsinghua.edu.cn

5) Ph.D., Prof., Department of Civil Engineering, Tsinghua University, Beijing, China. Email: luxz@tsinghua.edu.cn

Abstract: The dimension design of components in reinforced concrete frame structures heavily relies on engineering experience and iterative calculations, leading to significant inefficiencies. Existing intelligent design methods struggle to conduct component size design because it is challenging to accurately and densely represent information such as component layout, dimensions, and design conditions. This study proposes a method for intelligent component size design based on feature space for accurate and dense representation of design information, as well as a diffusion design process constrained by multi-channel masks. Firstly, the method substitutes the feature space for the traditional RGB space to represent component layout, dimensions, and design conditions, thereby enhancing data representation and neural network learning capabilities. Secondly, the study introduces an image-guided diffusion model with multi-channel mask tensors, and the corresponding training method is derived. Experimental results demonstrate that this model exhibits strong feature extraction capabilities and performs well in component dimension design tasks. Lastly, the study discusses the impact of parameters such as multi-channel masks and different dataset construction methods on the final prediction results.

Keywords: Generative AI design, Structural component size design, Diffusion model, RC frame structures, Image presentation

1. INTRODUCTION

Reinforced concrete (RC) frame structures are widely used in modern structural engineering, with the design of sectional dimensions being of significant importance for ensuring structural safety, cost-effectiveness, and environmental sustainability (Zhang et al., 2024). However, the traditional design process typically relies on the experience of structural engineers, leading to a high level of subjectivity and uncertainty in the design outcomes (Qin et al., 2024; Liao et al., 2024).

Traditional intelligent design methods for frame structure component dimension mainly consist of various optimization methods. Heuristic algorithms, such as genetic algorithms (Mariniello et al., 2022; Trapani et al., 2022; Xu and Gong, 2001), simulated annealing (Li et al., 2010), particle swarm optimization, artificial bee colony algorithm (Kaveh et al., 2020), are widely used in the optimization of RC frame component dimensions. These traditional optimization methods, especially in the application of complex structural design, are constrained by the design space scale and solution efficiency (Chang and Cheng, 2020). For example, heuristic algorithm-based optimization methods, while capable of providing effective solutions in certain scenarios, come with a significant computational burden, making it challenging to meet the design requirements of large RC frame structures (Peng et al., 2021). Additionally, the complexity of RC structures lies in the interactions between components and load effects, leading to highly nonlinear and coupled features in the design space (Trapani et al., 2022). The layout and section design of beams, columns, and slabs must comprehensively consider the influence of seismic loads, vertical loads, and other external factors (Xu et al., 2018). These complexities further compound the challenges posed by traditional design methods.

The emergence of models such as Generative Adversarial Networks (GANs) and Graph Neural Networks (GNNs) has introduced new methods for automated and efficient structural design. These algorithms, which learn from existing design drawings, can capture implicit knowledge in structural design. In comparison to traditional optimization algorithms, they have demonstrated significant improvements in both computational efficiency and design rationality. In recent years, the Diffusion Model has garnered attention in the field of structural engineering as an emerging generative model. The Diffusion Model operates by simulating data from high noise to low noise in a reverse process to gradually produce high-quality images (Ho et al., 2020). However, existing studies have primarily concentrated on shear wall layout design. To the best of the authors' knowledge, there is currently no method available for the dimension design of RC frame components based on the Diffusion Model.

In conclusion, the optimization of the dimension design of RC frame components encounters challenges stemming from the highly nonlinear space of traditional methods, the fitting limitations of GNN, and the sparse feature representation in GAN and Diffusion Models. Building on these identified challenges, this study seeks to investigate the utilization of the Diffusion Model in the dimension design of RC frame components. Through the

integration of multi-channel masks, a design method is proposed that densely considers the features of beam, slab, and column components. Specifically, the contributions of this study include the following points:

a) A dimension design method for RC frame components based on the Diffusion Model is proposed, which enhances the model's accuracy by introducing multi-channel masks.

b) A dataset construction method suitable for the dimension design of RC frame components is developed, taking into account story-related features such as story height and building-related features such as seismic design acceleration, and integrating multi-standard story information for the Diffusion Model learning.

The subsequent organization of this paper is as follows: Section 2 will provide an overview of the overall methodology, Section 3 will detail the construction method of the dataset; Section 4 will introduce the model architecture and its implementation details; Section 5 will discuss the impact of various factors on the generation effectiveness of the Diffusion Model. Finally, Section 6 will present the conclusion of this study.

2. THE DIFFUSION-BASED COMPONENT DIMENSION DESIGN METHOD FOR RC FRAME STRUCTURES

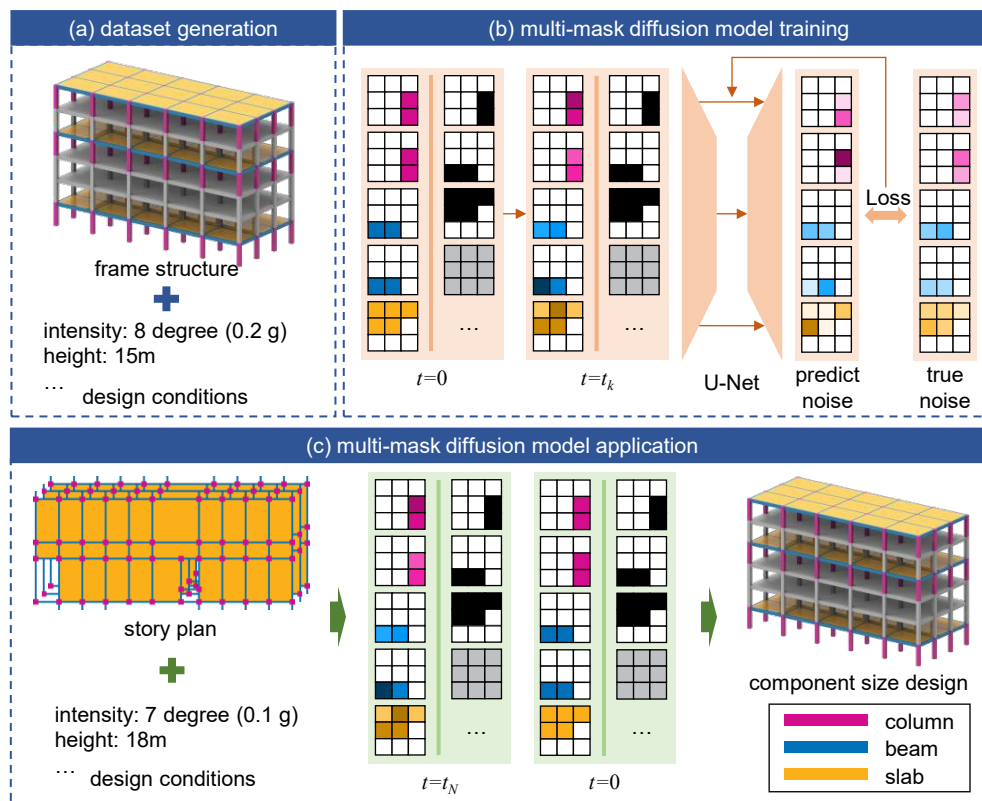


Figure 1. Diffusion-based design method for component size of RC frame structures. (a) dataset generation, (b) multi-mask diffusion model training, and (c) multi-mask diffusion model application.

This study introduces a dimension prediction method for RC frame components based on the diffusion process, named Frame-dimension-diffusion. As illustrated in Figure 1, the proposed method comprises three main components: (a) dataset generation, (b) multi-mask diffusion model training, and (c) multi-mask diffusion model application.

(a) Dataset generation (Figure 1(a)): A novel representation method for RC frame structures is introduced in this section to align with the input and output requirements of the diffusion model. In this study, the RC frame structure is represented layer by layer as images, with each image containing multiple channels representing information on component dimensions, layout, or design conditions. Based on the proposed structural representation method, this study extensively explores dataset construction, the incorporation of design conditions, and techniques for data augmentation, as outlined in Section 3.

(b) Multi-mask diffusion model training (Figure 1(b)): Gu et al. (2024) introduced a mask-based shear wall generation model, Struct-diffusion, which improved the density of shear wall structure data representation in images. This study extends the concept by proposing a diffusion model incorporating multi-channel masks. This approach mitigates the challenge of inadequate prediction performance for components with limited pixel proportions. The specific training process of the diffusion model will be presented in Section 4, with the benefits

of the multi-channel masks discussed in Section 5.

(c) Multi-mask diffusion model application (Figure 1(c)): Based on the training method outlined above, a dimension design model for RC frame components is obtained, enabling a more precise generation of RC frame component dimensions. The process for predicting the dimensions of frame components through the model is depicted in Figure 1(c). The detailed inference process of the diffusion model will be presented in Section 4.

3. DATASET GENERATION

The architectural design information of frame structures can be conveniently processed into structured data (Chang and Cheng, 2020). The utilization of text-prompt methods may not provide direct guidance and can result in redundant information. Therefore, this study adopts an image-prompt method to guide the diffusion process (Gu et al., 2024). Moreover, the generation strategy of Stable Diffusion mainly focuses on producing the complete image (Rombach et al., 2022). However, in practice, the pixel count occupied by frame columns and beams is considerably smaller than the total pixel count in the image ($O(1) \ll O(N) \ll O(N^2)$, where N represents the number of pixels on the longer side of the image). Therefore, using the full-image generation approach directly would result in significant inefficiency and could introduce biases during neural network training. Additionally, unlike the shear wall design task in Struct-diffusion, there are notable variations in the number of pixels occupied by beams, columns, and slabs. Therefore, a tailored multi-channel mask method needs to be devised. Based on the above content, this section will delve into the dataset construction method.

3.1 Method of dataset construction

Each standard story is regarded as a data point for analysis in this study. For each story, the method shown in Figure 2, as recommended by Han et al. (2024), was employed to encode features in the feature space and alleviate the impact of erroneous priors on the model. This study categorizes features into two groups: story-related features, demonstrated in Figure 2(a), and building-related features, illustrated in Figure 2(b). Story-related features may vary between different standard stories, such as the dimensions of beams, columns, and slabs, the arrangement of beams, columns, and slabs, loads on slabs, story heights, story numbers, etc. Building-related features, on the other hand, remain consistent within the same building, such as the total number of stories in the building, seismic design acceleration, characteristic period of the site, frame seismic resistance grade, etc.

The specific method entails transforming the current dimensions of frame components into pixelated story plans. In this study, the image pixels are configured at 256×256 , with a scaling ratio of $1 \text{ px} = 300 \text{ mm}$ between the image and the drawings. Due to the small dimension of most components after scaling by this ratio, which is not conducive for AI training, all columns are denoted by 3×3 pixel rectangles, while beams are fixed as 1-pixel-wide rectangles. It is crucial to highlight that adjusting the $1 \text{ px} = 300 \text{ mm}$ scale ratio does not significantly affect this study, as it does not pertain to layout design. The key criterion is ensuring each component can be distinctly identified on the drawing.

After converting into pixelated story plans, dimension information is allocated to the component locations channel by channel, constructing the corresponding tensors. Figure 2(a) illustrates an example within a 3×3 pixel range. Here, column l_x signifies the column width in the x-direction (measured in mm), while column l_y denotes the column width in the y-direction (measured in mm). It is essential to clarify that in this study, the x and y directions do not align with the strong and weak axes of the structure but rather with the local coordinate system of the columns. For non-rectangular columns like circular or pentagonal columns, are approximated as rectangular columns following engineering conventions. The interpretations of beam height, beam width, and slab thickness tensors are self-evident. The column mask, beam mask, and slab mask indicate the positions of columns, beams, and slabs, where a value of 1 signifies the presence of the corresponding component at that pixel point, while 0 denotes absence. The load tensor embodies the building loads, with values sourced from the Load Code for the Design of Building Structures (MOHURD, 2012). Scalar values such as story height and story number in Figure 2(a), as well as concrete material grade, total number of stories in the building, seismic design acceleration, characteristic period of site, and frame seismic resistance grade in Figure 2(b), which are independent of component positions, are directly expanded into 256×256 tensors. These tensors collectively constitute the feature tensors delineating the RC frame structure.

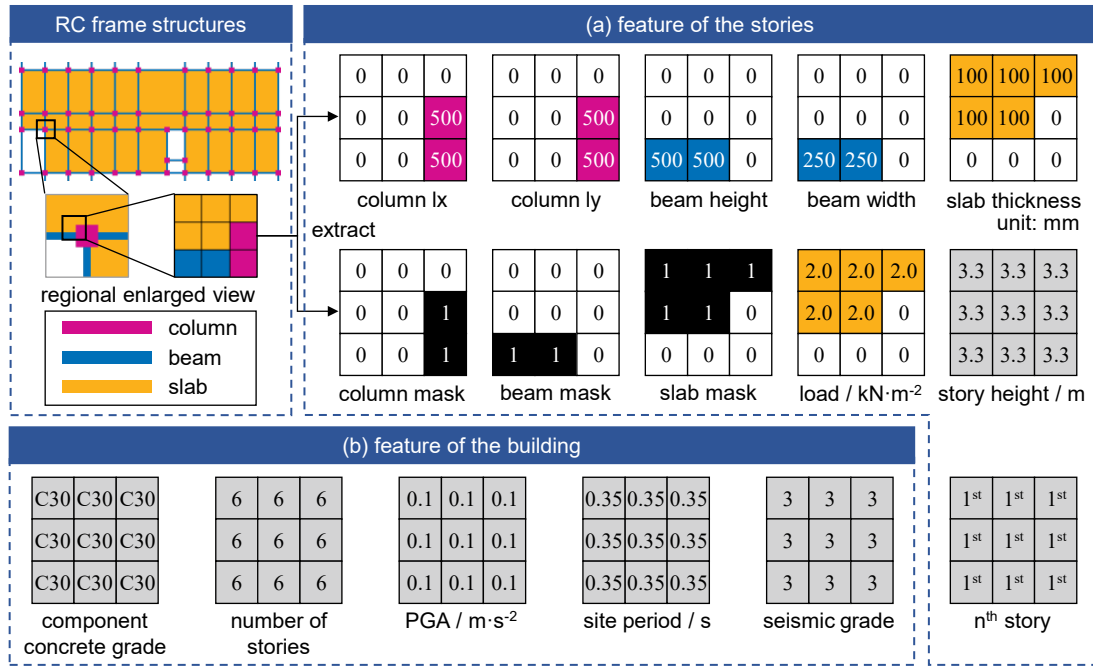


Figure 2 Representation of RC frame structures. (a) feature of the stories, and (b) feature of the building.

3.2 Data augmentation

The dataset utilized in this study is an enhanced version of the dataset introduced by Qin et al. (2024). The dataset comprises 109 buildings, with a total of 380 standard stories. The dataset is segmented into training, validation, and testing sets, with the precise allocation detailed in Table 1. To avert data leakage, the standard stories of the test buildings in dataset are ensured to be excluded from the training or validation sets.

Type	Datasets size
Train & Validation (K-fold)	362
Test	18

To address the limited number of training samples about the parameters of the diffusion model, data augmentation methods were also incorporated in this study to increase the amount of training data. The data augmentation techniques employed primarily encompass horizontal flipping, vertical flipping, and translation transformations to maintain the features of the training samples. By utilizing data augmentation techniques (Shorten and Khoshgoftaar, 2019), a training and validation dataset that is 196 times larger than the original dimension ($= 4 \times 7 \times 7$) was generated. This is beneficial for the training of the diffusion model.

4. THE ARCHITECTURE OF THE NEURAL NETWORK MODEL

4.1 The derivation of the image-prompt diffusion model with multi-mask tensor

The diffusion model primarily consists of a forward diffusion process and a reverse denoising process. As Equation 1, the forward diffusion process is a Markov process where Gaussian noise is added at each step (Ho et al., 2020):

$$q(y_t|y_{t-1}) = \mathcal{N}(y_t; \sqrt{1 - \beta_t}y_{t-1}, \beta_t I), q(y_{1:T}|y_0) = \prod_{t=1}^T q(y_t|y_{t-1}) \quad (1)$$

where β_t are hyperparameters of the noise schedule, y_t are input tensor in time t , $t=0, 1, 2, 3, \dots, T$, T is the time steps of the diffusion model, $q(x|y)$ means the conditional probability distribution of x given y , $\mathcal{N}(x; \mu; \sigma^2 I)$ means a multivariate normal distribution with a vector mean of μ and a covariance matrix of $\sigma^2 I$. In this study, $T=2000$, and the noise schedule is linear, which means that β_t varied linearly from 0.000001 to 0.01 (Dhariwal and Nichol, 2021). When $t = T$, there is no distinction between y_T and Gaussian noise. It is noteworthy that we can obtain the Equation 2:

$$q(y_t|y_0) = \mathcal{N}(y_t; \sqrt{\bar{\alpha}_t}y_0, (1 - \bar{\alpha}_t)I) \quad (2)$$

where $\alpha_t = 1 - \beta_t$, $\bar{\alpha}_t = \prod_{i=1}^t \alpha_i$. $\bar{\alpha}_t$ actually measures the noise level, meaning that a smaller $\bar{\alpha}_t$ indicates that the image contains more noise. Ho et al. (2020) show a closed form of the posterior distribution of y_{t-1} given

(y_0, y_t) as

$$q(y_{t-1}|y_0, y_t) = \mathcal{N}\left(y_{t-1}; \frac{\sqrt{\alpha_t}(1 - \bar{\alpha}_{t-1})}{1 - \bar{\alpha}_t} y_t + \frac{\sqrt{\alpha_{t-1}}}{1 - \bar{\alpha}_t} \beta_t y_0, \frac{1 - \bar{\alpha}_{t-1}}{1 - \bar{\alpha}_t} \beta_t I\right). \quad (3)$$

Based on the work of Gu et al. (2024) and Ho et al. (2020), this study additionally incorporates a multi-mask tensor. Given a noisy tensor $\tilde{y}_t = \tilde{y}_{bh,t} \oplus \tilde{y}_{bw,t} \oplus \tilde{y}_{cx,t} \oplus \tilde{y}_{cy,t} \oplus \tilde{y}_{st,t}$:

$$\tilde{y}_{k,t} = (\sqrt{\alpha_t} y_{k,0} + \sqrt{1 - \alpha_t} \varepsilon_{k,t}) \odot y_{k_mask} + y_0 \odot (1 - y_{k_mask}), \varepsilon_{k,t} \sim \mathcal{N}(0, I), k \in \{bh, bw, cx, cy, st\}; \quad (4)$$

$y_{bh_mask} = y_{bw_mask} = y_{beam_mask}$; $y_{cx_mask} = y_{cy_mask} = y_{column_mask}$; $y_{st_mask} = y_{slab_mask}$ where $\tilde{y}_{bh,t}$ means beam height channel of noisy tensor \tilde{y}_t . bw means beam width, cx means column lx, cy means column ly, and st means slab thickness, \oplus means concatenate operator. We denote $y_{mask} = y_{beam_mask} \oplus y_{beam_mask} \oplus y_{column_mask} \oplus y_{column_mask} \oplus y_{slab_mask}$. The goal is to recover the target tensor $y_0 \odot y_{mask}$, where \odot is the hadamard product operator. A neural network $M_\theta(y_{cond}, \tilde{y}_t, \bar{\alpha}_t)$ can be constructed, which takes condition tensor y_{cond} , input tensor with noise \tilde{y}_t and noise level $\bar{\alpha}_t$ as input and fits the noise vector ε by optimizing the objective

$$\mathbb{E}_{(y_0, y_{cond}, y_{mask})} \mathbb{E}_{\bar{\alpha}, \varepsilon} [\|(M_\theta(y_{cond}, \tilde{y}_t, \bar{\alpha}_t) - \varepsilon_t) \odot y_{mask}\|_p^p]. \quad (5)$$

Based on this, the process of training the denoising diffusion model can be obtained, as outlined in Table 2. The inference process is the same as Struct-diffusion.

Table 2 The process of denoising diffusion model training and inference

Algorithm 1 Training	
1:	repeat
2:	$y_{cond}, y_0, y_{mask} \sim q(y_{cond}, y_0, y_{mask})$
3:	$t \sim \text{Uniform}(\{1, \dots, T\})$
4:	$\varepsilon \sim \mathcal{N}(0, I)$
5:	Take gradient descent step on
	$\nabla_\theta \ (M_\theta(y_{cond}, (\sqrt{\alpha_t} y_0 + \sqrt{1 - \alpha_t} \varepsilon) \odot y_{mask} + y_0 \odot (1 - y_{mask}), \bar{\alpha}_t) - \varepsilon_t) \odot y_{mask}\ ^2$
6:	until converged
Algorithm 2 Inference	
1:	$z \sim \mathcal{N}(0, I)$
2:	$\hat{y}_T \leftarrow z \odot y_{mask} + y_0 \odot (1 - y_{mask})$
3:	for $t = T, \dots, 1$ do
4:	$z \sim \mathcal{N}(0, I)$ if $t > 1$, else $z = 0$
5:	$\hat{y}_{t-1} \leftarrow \left(\frac{1}{\sqrt{\alpha_t}} \left(\hat{y}_t - \frac{1 - \alpha_t}{\sqrt{1 - \alpha_t}} M_\theta(y_{cond}, \hat{y}_t, \bar{\alpha}_t) \right) + \sqrt{\beta_t} z \right) \odot y_{mask} + y_0 \odot (1 - y_{mask})$
6:	end for
7:	return \hat{y}_0

It is worth noting that when $y_{beam_mask} = y_{column_mask} = y_{slab_mask} = y_{mask}$, this method degenerates to Struct-diffusion. In this study, we consider $y'_{beam_mask} = y'_{column_mask} = y'_{slab_mask} = y_{beam_mask} \mid y_{column_mask} \mid y_{slab_mask}$ as a single mask scenario, where $y_{beam_mask}, y_{column_mask}, y_{slab_mask}$ represent the layout of beams, columns, and slabs obtained in Section 3.1.

4.2 Training details

This study employed a U-Net network with an attention mechanism and temporal encoding (Dhariwal and Nichol, 2021) as the denoising model. To investigate the impact of the various factors mentioned above on the model results, this study conducted two sets of training. The specific training IDs and corresponding model parameters can be found in Table 3. In the table, the ID column represents the name of the trained model. The ID names are designated based on the presence of multi-channel masks. For example, for model with multi-channel masks, the model ID is labeled as M. Here, “M” stands for Multi-mask, and “S” stands for Single-mask. The term Epoch denotes the number of training cycles, whereas Epoch 21600 signifies a total of 21600 training cycles. The Seed indicates the chosen random seed, which controls processes such as neural network weight initialization and Gaussian noise sampling. If the random seeds are the same, it indicates identical outcomes for these processes. It is crucial to mention that each set of training has three random seeds, indicating that each set underwent three training runs to reduce the impact of random errors on the research conclusions. The Training hours column represents the total training time. Further explanations on Mask type is provided in Section 4.1.

Table 3. Training details

ID	Epoch	Seed	Training hours	Mask type
S	6400	42, 88, 8888	58.30	Single
M	11200	42, 88, 8888	110.00	Multi

The training strategy for the aforementioned experiments was consistently maintained, wherein the models were trained with a learning rate of 5×10^{-5} following optimization of other hyperparameters. Mean square error calculation was employed for both training and validation processes. Validation error assessment occurred every 200 Epochs, and training concluded after 30 consecutive validations without enhancement. The model with the minimal validation error was chosen for prediction and subsequent metric computations. Furthermore, to ensure model reproducibility, the random seed specified in Table 3 governed the model's random initialization, irrespective of the dataset splitting methodology. To ensure equitable comparisons, the dataset-splitting method remained uniform. Table 3 also presents the training duration of the models (on a single GPU, measured in hours). The computing platform configuration included OS: Ubuntu 22.04 LTS; CPU: Intel Xeon E5-2682 v4 @ 64x 3GHz; RAM: 32GB; GPU: NVIDIA GeForce RTX 3090 24GB.

5. DISCUSSION

5.1 Evaluation metrics and postprocessing

Following the methodology proposed by Qin et al. (2024), this study employed the Root Mean Square Error (RMSE) metric for evaluation. The metric calculates the root mean square error between the results obtained from the diffusion model design and the engineering design (Ground Truth, GT). The reason for choosing this metric is that it maintains the same units as the original data, allowing for a direct assessment of the variability in the corresponding predicted results. In this study, we primarily compared the RMSE values for beam height, beam width, column length in the x-direction, column length in the y-direction, and slab thickness. It is noteworthy that in practical design, the modulus is often applied. For instance, if the designed beam width is 195mm, it might be rounded up to 200mm. Typically, the modulus for beams and columns is 50mm, while for slabs, it is 20mm.

For the predictions produced by the diffusion model, the average value of the respective channel for each component was extracted to acquire the dimensional details of that component. Subsequently, by consolidating all the dimensional information of the building, the RMSE was computed as the evaluation metric in this study. Hence, the RMSE is component-centric rather than pixel-centric, aligning with real-world applications.

5.2 Test results

Using the RMSE metric provided in Section 5.1, the two sets of models shown in Table 3 were tested, and Table 4 presents the average performance of these models on the prediction dataset. Each column in the table represents the model's ID, as well as the RMSE values for beam width, beam height, column length in the x-direction, column length in the y-direction, and slab thickness, all measured in millimeters.

Table 4. RMSE error on the Test dataset (unit: mm)

ID	Beam height	Beam width	Column lx	Column ly	Slab thickness
S	152.0696	44.4171	122.5505	128.0622	13.5619
M	114.3256*	32.4199*	62.6379*	76.3565*	13.3414*

Comparing the results of IDs S with M, it can be observed that the multi-channel masks outperform the single-channel masks. In the case of column and beam channels, the absence of multi-channel masks leads to a high number of zero values, significantly surpassing the valid values. This scenario can mislead the neural network during backpropagation, as it may predict the positions directly as zeros, thereby causing larger prediction errors. Additionally, it is noticeable that the improvement in prediction errors for slabs is relatively minor compared to beams and columns, which aligns with the aforementioned discussion.

5.3 Summary

This section analyzed the impact of various factors on the results generated by the diffusion model. Among them, the model utilizing multi-channel masks achieved the best performance. Furthermore, the top-performing model had all dimension predictions within two modulus, except for beam height prediction which was within three modulus. It can be concluded that this method is suitable for predicting the dimension of RC frame components.

6. CONCLUSIONS

This study introduces a novel diffusion model-based approach for predicting component dimensions in RC frame structures, effectively addressing the limitations of traditional heuristic methods and GAN-based approaches in capturing the nonlinear and coupled features of structural design. By integrating multi-channel masking, the proposed Frame-dimension-diffusion model offers a more robust and precise prediction of beam, column, and slab dimensions. The principal contributions of this research are as follows:

(a) A novel framework based on diffusion models has been developed specifically to predict the component dimensions of RC frames. This approach leverages advanced generative modeling techniques inherent to diffusion processes, enabling more precise estimations of component dimensions.

(b) The framework incorporates multi-channel masks to improve feature representation for individual components. By applying this technique, the model achieves higher density and accuracy in capturing the nuanced features of RC frame components.

(c) A novel dataset construction methodology is proposed, capturing key characteristics of standard stories and building seismic conditions, thereby enhancing the training process and efficacy of the diffusion model.

The findings indicate that the Frame-dimension-diffusion model is capable of generating component dimensions that comply with engineering standards, exhibiting reduced prediction error. Future research should focus on improving model efficiency through advanced architectural techniques, such as model distillation, and incorporating spatial layout information to extend the diffusion model's applicability to more complex structural design scenarios.

ACKNOWLEDGMENTS

This work was supported by the Ministry of Housing and Urban-Rural Development of the People's Republic of China, Science and Technology Program (2022-K-073).

REFERENCES

- Chang, K.-H., Cheng, C.-Y., 2020. Learning to Simulate and Design for Structural Engineering, in: Proceedings of the 37th International Conference on Machine Learning. Presented at the International Conference on Machine Learning, PMLR, pp. 1426–1436.
- Dhariwal, P., Nichol, A., 2021. Diffusion Models Beat GANs on Image Synthesis. <https://doi.org/10.48550/arXiv.2105.05233>
- Gu, Y., Huang, Y.L., Liao, W.J., Lu, X.Z., 2024. Intelligent design of shear wall layout based on diffusion models. *Computer-Aided Civil and Infrastructure Engineering*. <https://doi.org/10.1111/mice.13236>
- Han, J., Lu, X.Z., Gu, Y., Liao, W.J., Cai, Q., Xue, H.J., 2024. Optimized data representation and understanding method for the intelligent design of shear wall structures. *Engineering Structures* 315, 118500. <https://doi.org/10.1016/j.engstruct.2024.118500>
- Ho, J., Jain, A., Abbeel, P., 2020. Denoising Diffusion Probabilistic Models. <https://doi.org/10.48550/arXiv.2006.11239>
- Kaveh, A., Hamedani, K.B., Hosseini, S.M., Bakhshpoori, T., 2020. Optimal design of planar steel frame structures utilizing meta-heuristic optimization algorithms. *Structures* 25, 335–346. <https://doi.org/10.1016/j.istruc.2020.03.032>
- Li, G., Lu, H.Y., Liu, X., 2010. A hybrid simulated annealing and optimality criteria method for optimum design of RC buildings. *Structural Engineering and Mechanics* 35, 19–35. <https://doi.org/10.12989/sem.2010.35.1.019>
- Liao, W.J., Lu, X.Z., Fei, Y.F., Gu, Y., Huang, Y.L., 2024. Generative AI design for building structures. *Automation in Construction* 157, 105187. <https://doi.org/10.1016/j.autcon.2023.105187>
- Mariniello, G., Pastore, T., Bilotta, A., Asprone, D., Cosenza, E., 2022. Seismic pre-dimensioning of irregular concrete frame structures: Mathematical formulation and implementation of a learn-heuristic algorithm. *Journal of Building Engineering* 46, 103733. <https://doi.org/10.1016/j.jobbe.2021.103733>
- MOHURD, 2012. Load Code for the Design of Building Structures (GB 50009-2012).
- Peng, B., Flager, F., Barg, S., Fischer, M., 2021. Cost-based optimization of steel frame member sizing and connection type using dimension increasing search. *Optimization and Engineering* 23, 1525–1558. <https://doi.org/10.1007/s11081-021-09665-5>
- Qin, S.Z., Liao, W.J., Huang, Y.L., Zhang, S.L., Gu, Y., Han, J., Lu, X.Z., 2024. Intelligent Design for Component Size Generation of Reinforced Concrete Frame Structures Using Heterogeneous Graph Neural Networks. *Automation in Construction*.
- Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B., 2022. High-Resolution Image Synthesis with Latent Diffusion Models. <https://doi.org/10.48550/arXiv.2112.10752>
- Shorten, C., Khoshgoftaar, T.M., 2019. A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data* 6, 60. <https://doi.org/10.1186/s40537-019-0197-0>
- Trapani, F.D., Sberna, A.P., Marano, G.C., 2022. A genetic algorithm-based framework for seismic retrofitting cost and expected annual loss optimization of non-conforming reinforced concrete frame structures. *Computers & Structures* 271, 106855. <https://doi.org/10.1016/j.compstruc.2022.106855>
- Xu, L., Gong, Y.L., 2001. Preliminary Design of Long-Span King-Post Truss Systems with a Genetic Algorithm. *Computer-Aided Civil and Infrastructure Engineering* 16, 94–105. <https://doi.org/10.1111/0885-9507.00216>

- Xu, L.H., Yan, X.T., Li, Z.X., 2018. Development of BP-based seismic behavior optimization of RC and steel frame structures. *Engineering Structures* 164, 214–229. <https://doi.org/10.1016/j.engstruct.2018.03.012>
- Zhang, C., Tao, M.X., Wang, C., Yang, C., Fan, J.S., 2024. Differentiable automatic structural optimization using graph deep learning. *Advanced Engineering Informatics* 60, 102363. <https://doi.org/10.1016/j.aei.2024.102363>

CONSTRUCTABILITY-BASED MULTI-OBJECTIVE OPTIMIZATION FOR REINFORCING BAR DESIGN IN RECTANGULAR CONCRETE BEAMS

Luis F. Verduzco¹, Jack C.P. Cheng²

1) PhD student, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology (HKUST), Hong Kong SAR. Email: lfvm@connect.ust.hk

2) PhD, Assoc. Head Prof., Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology (HKUST), Hong Kong SAR. Email: cejcheng@ust.hk

Abstract: Constructability-based optimization design of reinforcing bar (rebar) in concrete structures (RC) has been attracting attention in recent years when aiming towards industrialized sustainable construction. This paradigm enables it to more effectively link the practicality of reinforced concrete designs and their associated material usage and construction cost. The problem itself is multi-objective (MO), and the development of effective optimization algorithmic frameworks to approach its solution is essential. For this purpose, Artificial Intelligence (AI) based optimization with enhanced Meta-heuristic algorithms (MA) has demonstrated to be the key to reduce the computational demand. Particularly for beams, the deployment of Artificial Neural Networks (ANN) and Graph Neural Networks (GNN) has proven to be of the most effective AI-based optimization approaches. Nonetheless, its application for these elements has been limited, so far, to single-objective (SO) optimization and not for MO optimization, which entails further considerations to effectively reach optimal Pareto Fronts (PF) in a time-efficient manner. Additionally, the lack of constructability metrics, at this point, for rebar design in RC structures, in the literature, is still evident. Even though some efforts have been made in the last years, for some types of elements, there is still a gap when it comes to elaborate and flexible constructability models that may be used in general, for any project at hand.

This work presents the development of a novel MO optimization framework with GNN-Enhanced Metaheuristics (MA), for rebar design in multi-span beams. For this purpose, the development of a constructability score (CS) model is proposed based on rebar cuts and labor assembling complexity. The Non-Sorting Genetic Algorithm II (NSGA-II) is used for enhancement. The performance of each algorithm is analyzed and compared between Non-Enhanced, ANN and GNN, in terms of convergence and time efficiency.

Keywords: Constructability, Multi-Objective Optimization, Graph Neural Networks, Rebar design, Reinforced Concrete Beams

1. INTRODUCTION

Optimization of rebar for RC structures has proven to be a complex problem, of such magnitude, that many researchers have referred to it to be NP-hard (Afzal M., 2020). specially for beams, columns and slabs. The use of pure meta-heuristic optimization algorithms has demonstrated, thus, to be time expensive, and without necessarily reaching acceptable optimum solutions. For RC beam elements, in comparison to RC column elements, the rebar crosses multiple design cross-section along the element's length, so that free clash must be assured at all locations, which makes the optimization problem of higher complexity. In this regard, most of the optimization procedures developed by researchers have been focused on the section level, separately, failing in considering the parametric relationship of the various rebar layouts between such sections and between multiple structural elements. In turn, the design results usually lack the necessary and essential *constructability*, leading to non-optimized material waste and non code-compliance.(Li M. et al., 2023).

Constructability has become in recent years an essential metric for industrialized construction. It promotes the optimum use of knowledge and experience in conceptual planning and field operations, in a coordinated systematic workflow, to facilitate the construction of infrastructure designs. When considering constructability for structural design purposes, the problem is multi-objective, given the trade-off with construction costs or rebar design volumes (Lao W.-L. et al., 2023). In this context, when referring to RC structures, the development of constructability metrics, specifically for rebar designs, is of paramount importance for a more proper estimation of construction costs. Nonetheless, most of the existing research and constructability metrics, up to now, have been focused at the building level and not at the element level, ignoring the constructability of the reinforcement itself.

To solve this optimization design problem related to rebar design, meta-heuristics alone have demonstrated to be computationally expensive (Afzal M., 2020). In recent years, however, and most specifically, in the present decade, the use of Machine Learning (ML) and Deep Learning (DL) models have proven considerable efficiency to aid meta-heuristic optimization algorithms for rebar design of RC structures, both in terms of convergence and time execution. This approach, along with BIM technology, has demonstrated incredible potential with huge reductions of computational time and convergence efficiency to global optimum solutions. In (Li M. et al., 2023), for instance, Graph Neural Networks (GNN) were used to aid Exploratory Genetic Algorithms (EGA), as an automated pipeline for free-clash rebar design optimization of rebar layouts, both for structural

elements and for different kinds of joints of RC buildings. It was demonstrated that under this approach, the computation time could be reduced by 75% to 90% with great convergence results. The key to this success is that, through GNN, the representation of rebar designs as graphs would enable it to consider the parametric relationship between different rebar groups in a single structural element, or among multiple ones, enhancing the efficiency of clash-free rebar design optimization.

Up to now, however, the application of AI-based optimization design has been mostly focused on single-objective (SO) optimization and not MO optimization. The applications of ANN and GNN are no exception. In fact, according to (Afshari, 2019) most of the formulations for RC beams have been SO, focused mostly on construction cost. Those few works focused on multi-objective optimization are mostly concerned with construction cost and flexural limit states, with only a couple of works focused on deflection limit states. In more recent works, a focus on constructability-based optimization has been emphasized. In (Li M. et al., 2021), for instance, a multi-objective optimization approach between material cost and labour cost was adopted. In (Lao W.-L. et al., 2023), constructability-scores of reinforcement were formulated for precast beams and columns to account for standardization in a BIM-enabled framework in a multi-objective optimization formulation with the NSGA-II-GD. Nonetheless, there seems to be an existing gap in regards to constructability-based optimization design with AI-based optimization processes. Additionally, the lack of solid formulations of constructability scores models for reinforcement designs is also evident.

This work aims at bridging the gap in the literature focused on RC beam optimization between construction cost, limit state objectives and constructability. For this purpose, a constructability score model focused entirely on the rebar layout along the whole length of beam span is proposed, by considering the number of different diameter sizes, number and distribution of the rebar itself and rebar cuts to avoid clashes. The approach to be adopted is multi-objective (constructability - construction cost / rebar volumes), to complement the current state-of-art related to AI-based optimization of rebar design in RC structures. More specifically, a GNN-Enhanced MA is proposed for single-span beams and compared to other ML methods, such as ANN, in terms of time efficiency and convergence.

2. LITERATURE REVIEW

Since the commercialization of computers, the investigation of novel optimization design methods and frameworks for engineering design started to take place. In this regard, meta-heuristics, soon, took over the conventional mathematical optimization methods given their versatility to adapt to non-linear problems of multiple variables. Perhaps the first algorithms to be used for optimization design of RC structures is the Genetic Algorithm, in the 90's. Examples of the earliest works with the GA for concrete beams are (Coello, 1997), (Koumoussis & Arsenis, 2002) for beams and deep beams, considering construction costs and weight as the main objective functions.

A few years later, with the development of new meta-heuristics, hybrid meta-heuristic optimization algorithms began to be developed for the optimum design of reinforcement in RC structures, such as in (Bekdas & Nigdeli, 2012), where the Big-Bang Crunch and the Harmony Search (HS) algorithm were used for the optimum design of T-shaped beams. These advances enabled, as well, the further development of optimization processes for systems of continuous beams, as in (Govindaraj & Ramasami, 2005) where the GA was used, based on rebar templates and prototypes. In later years, BIM technology began to be used, along with these optimization processes, to work as the visualization platform for better appreciation of the optimum results and automation of detail drawings, and eventually to serve as the programming platform itself, as in (Li, M. et al., 2021).

In most of these studies, however, due to the high complexity of the problem and search space of rebar solutions, many simplifications and assumptions have been formulated for the optimization processes to have relatively acceptable efficiency in time execution (Shaqfa & Orban, 2019). Thus, most of the solutions found usually turn out to be far from the global optimum ones. In (Govindaraj & Ramasami, 2005), for instance, the authors used reinforcement tables for each beam cross-section, separately, without considering the rebar as an optimization variable, limiting the feasible search space to what the database would provide. Similar approaches have been adopted in other works, as in (Kwak & Kim, 2008), where no global optimum solutions are ever found, but only improved ones from a given initial proposed solution. The main issue lies on identifying the key features related to the relationship between the rebar designs of the different critical cross-sections along a beam span or several spans. That is where ANN/GNN and ML/DL, in general, can have the greatest advantages.

3. CONSTRUCTABILITY ANALYSIS FOR REBAR IN CONCRETE BEAMS

Constructability refers to the optimum use of construction knowledge and experience in the conceptual planning, engineering, procurement and field operations phases to achieve the overall project objectives. When referring to rebar design in RC structures, the number of rebars and its distribution as well as the rebar diameter sizes are the two main factors to consider. For this purpose, the specification of rebar design patterns or prototypes should be created. These patterns may vary, according to the design mechanism of each structural element. For beams, for instance, it is common to have different rebar layers in the tension zone, one on top of another. In this

work, the rebar prototype shown in Figure 1. For such rebar design prototype, the following enlisted distribution constraints are considered.

1. A maximum of three layers of rebars in tension is considered for each design cross-section
2. Each layer can be only composed of two rebar diameter sizes
3. Any rebar on the upper layers has to be placed right above the bottom layer
4. A minimum of two bars in tension and two in compression should always be provided, for confining reinforcement

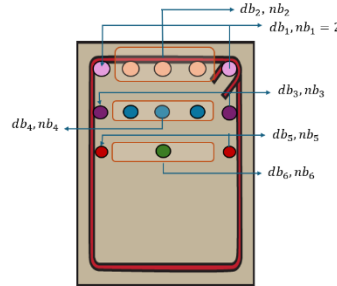


Figure 1. Rebar design prototype for a beam cross-section.

However, specifically for beams, additional factors, in relation to other structural elements, should be also considered. For instance, the number of cuts to perform would be of great influence on the constructability of the element. In general, cuts are usually sought to minimize, to reduce waste. Additionally, free-clash constraints must be satisfied at all costs, both at the element level as well as when considering multi-span beams with joints between them.

3.1 Constructability Score model

In summary, for beams, three main variables should be considered when measuring the constructability of rebar designs: uniformity of number of rebars and their distribution UNB , uniformity of rebar diameter sizes UND and number of rebar cuts NC , as expressed next in (1):

$$CS = UNB + UND + NC \quad (1)$$

The variable UNB is the uniformity of number of rebars distributed over the main three design cross-sections (left, middle and right). This variable depends on the number of rebar layers $nlay$ at each cross-section i , as following. The maximum number of layers at each cross-section is considered as three $[1 \leq nlay_i \leq 3]$:

$$UNB_i = \begin{cases} \frac{1}{(nlay_i)^{W_{UNB}}} \sum_{j=1}^{nlay_i-1} UNBL_{i,j}, & nlay > 1 \\ 1, & nlay = 1 \end{cases} \quad (2)$$

In the previous equation, the variable $UNBL_j$ refers to the uniformity of number of rebars $nb_{j,i}$ between each layer j of the cross-section i , with respect to the bottom layer $nlay_{1,i}$, as follows:

$$UNBL_{i,j} = \begin{cases} \frac{nb_{i,j+1}}{nb_{i,1}}, & 2 \leq nb_{i,j} \leq nb_{i,1} - 1 \\ 1, & otherwise \end{cases} \quad (3)$$

The weight factor $W_{\{UNB\}}$ penalizes the number of layers, so that the UNB score may decrease as the number of layers increase. Thus, it is recommended for this weight factor to be $[1 \leq W_{\{UNB\}} \leq 2]$. In general, for each cross-section $[0 < UNB_i \leq 1]$.

The variable UND refers to the uniformity of rebar diameter sizes of each cross-section, and is defined as follows:

$$UND_i = \frac{1}{ND_{i,j}^{W_{\{ND\}}}} \quad (4)$$

where $ND_{i,j}$ is the number of different rebar diameter sizes $[1 \leq ND_{i,j} \leq 3]$. Because the ND score decreases as the number of rebar diameter sizes increase, it is recommended that the corresponding weight factor W_{ND} is between $[0.1 \leq W_{ND} \leq 1]$.

Finally, the variable UC refers to the uniformity in which rebar cuts at each cross-section take place. It is defined as following:

$$UC = \begin{cases} \frac{1}{nlay_i^{W_{NC}}} \sum_{j=1}^{nlay_i} \sqrt{\frac{nb_{i,j} - nb_{c,i,j}}{nb_{i,j}}}, & 1 \leq nb_{i,j} \leq nb_{i,1} - 1 \\ 1, nlaycut_i = 0 \end{cases} \quad (5)$$

As it could be observed, the corresponding weight factor of this last variable W_{NC} penalizes the number of layers in which cuts are going to take place $nlaycut_i$, instead of the number of cuts per layer or per section. The reason is that usually, the complexity of executing cuts on rebars would depend more likely on the number of layers that require cuts, rather than the number of cuts themselves. This weight factor is recommended to have values between $[0.1 \leq W_{NC} \leq 1]$.

4. OPTIMIZATION FRAMEWORK

4.1 Design constraints

According to the ACI 318-19 (ACI318, 2019) beams of special moment frames should comply with the following design constraints, both in regard to dimensions and reinforcement quantity:

1. Maximum and minimum rebar cross-section area:

$$A_{s,min} \geq \begin{cases} \left(\frac{3\sqrt{f'_c}}{f_y} \right) b_w d \\ \frac{200}{f_y} b_w \end{cases} \quad (6)$$

$$A_{s,max} \leq 0.0025bh \quad (7)$$

2. Minimum rebar separation

For parallel nonprestressed reinforcement in a horizontal layer, clear spacing shall comply with:

$$sep_{min,x} = \max \begin{cases} 25mm \\ d_b \\ \frac{4}{3}d_{agg} \end{cases} \quad (8)$$

For parallel reinforcement placed in two or more horizontal layers, reinforcement in the upper layers shall be placed directly above reinforcement in the bottom layer with a clear spacing between layers of at least $sep_{min,y} \geq 25mm$.

3. Side rebars must be provided, in case the height dimension is greater than $900mm$. For such case, they should be distributed uniformly for a distance $h/2$ from the tension face. Spacing of this skin reinforcement should not be in accordance to (9) where $f_s = 2/3f_y$.

$$sep_{min,y} = \max \begin{cases} 15 \left(\frac{40000}{f_s} \right) - 2.5C_c \\ 12 \left(\frac{40000}{f_s} \right) \\ 250mm \end{cases} \quad (9)$$

4.2 AI-based optimization design process

The relative new tendency of AI-based optimization design in engineering considers that the training of the ML/DL model is done before executing any optimization process. In order to take the most advantage of the AI model, the engineering process itself must be decomposed in two or more stages, so that the AI model may be used, in as many as possible ones, for leverage.

The whole optimization design process is summarized as shown below in Figure 2, both for the Non-AI-Enhanced optimization algorithm with pure MAs (left panel), and for the AI-based optimization design process with MAs (right panel). In general, the process for both paradigms consists of a three-stage hybrid optimization process with the PSO and GA. In total, twelve main optimization variables take place. Such variables are the rebar diameter sizes and number of rebars for the left and middle section of the beam, over their respective three layers in tension. After rebar cut design takes place, additional analogous optimization variables are considered for the

right cross-section. In this case, the DL/ML model is used in two sub-stages of the general optimization design process, namely in (1) design of reinforcement in left and middle cross-sections and (2) design of reinforcement in right cross-section, after cut design.

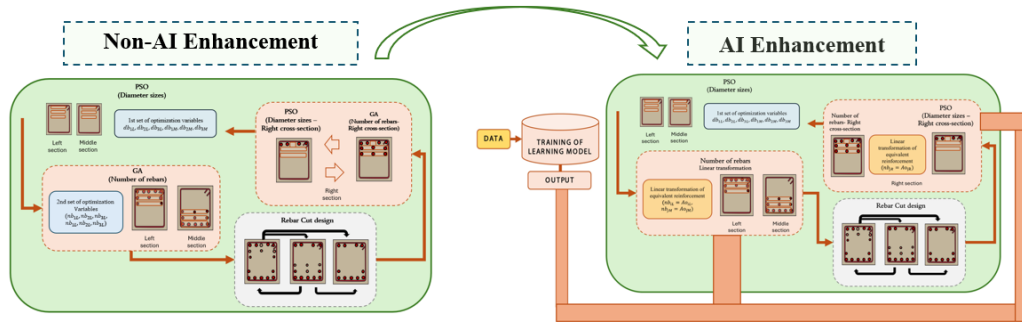


Figure 2. Summarized optimization design process of rebar in rectangular concrete single-span beams.

4.3 Constructability-based MO optimization

When formulating a constructability-based MO optimization design of structures, the constructability is sought to be maximized, while minimizing at the same time, the construction costs or material usage, as shown next in Figure 3. When the problem is stated in terms of construction cost, it can be expressed mathematically as in (10), where x is the decision vector is composed by the $n = 12$ variables just mentioned previously (namely the discrete variables corresponding the rebar diameters for the left and middle cross-section). S is the feasible objective space with $m = 2$ objective functions, namely the total construction cost C_s of reinforcement $f_1(x)$ and the constructability CFA $f_2(x)$, under the $i = 1, \dots, q = 5$ constraints just mentioned (max and min rebar separation, structural resistance efficiency and max/min rebar cross-section area).

$$\begin{aligned} \min [F(x) &= (f_1(x), 1 - f_2(x))] \\ S.T.: x &= [x_1, x_2, \dots, x_n] \in S \\ \text{Under: } g_i(x) &\leq y_i, i = 1, 2, \dots, q \end{aligned} \quad (10)$$

For this work, a hybrid three-staged MO optimization algorithm with the NSGA-II was formulated, as shown next in Figure 3:

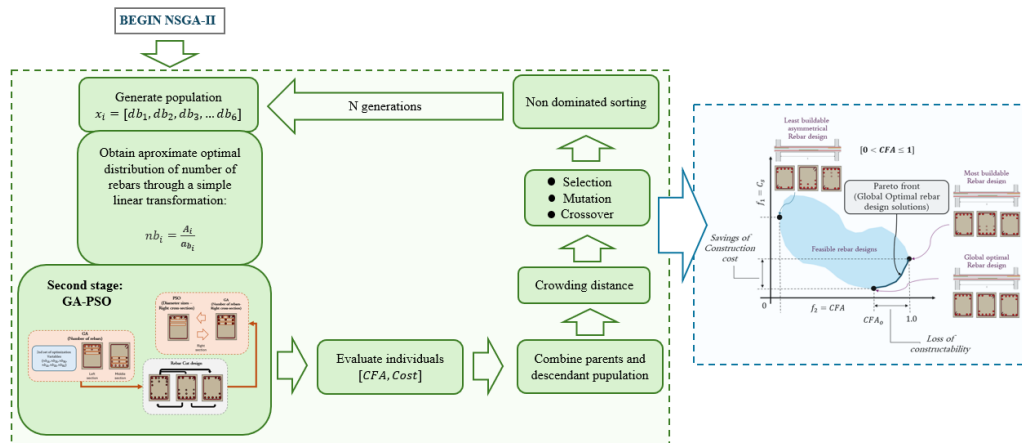


Figure 3. Flow diagram depicting the formulated non-enhanced three-staged hybrid NSGA-II for MO optimization design of rebar in concrete beams.

4.4 Generation of training data

Differing from what happens in AI-Enhanced SO optimization, where only one optimum solution is sought for convergence, in AI-Enhanced MO optimization it has been suggested in the literature to work with Multi-Levels of Optimum Targets (M-LOT). That is, to generate training data sets through non-enhanced MO Optimization, with pure meta-heuristics, for instance, and then, extract multiple optimum solutions from the PFs to generate multiple output entries for the training data. For this specific problem, this formulation of AI training

is depicted next in Figure 4, where, for instance, three-LOT are considered. The input data, on the other hand, consists of six entries, namely: cross-section dimensions b, h concrete's compressive strength f'_c , span's length L and pure bending load actions M_{uL}, M_{uM}, M_{uR} .

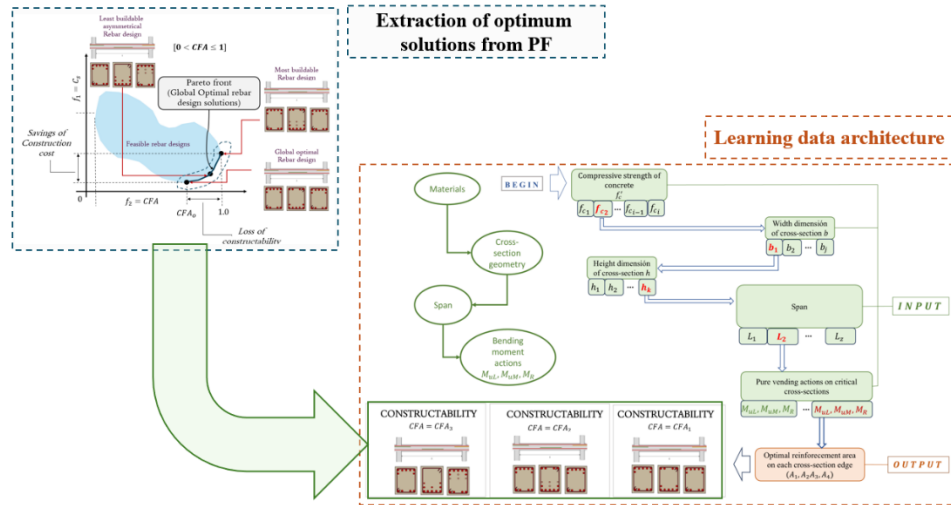


Figure 4. Generation of training data sets for AI-based optimization of rebar in beams, through MLOT.

4.5 AI-Enhanced optimization design framework

Based on the previous formulations, the Non-enhanced MO optimization framework from Figure 3 would be transformed into the one shown next in Figure 5 under an AI-based paradigm. Here, given that M-LOT are considered for the training of the ML/DL model, multiple models are created for each constructability level of rebar design. Therefore, in this MO AI-based optimization process, when generating an individual, a random selection process takes place to choose a constructability level of rebar design, to predict the required amount of rebar cross-sectional area for each cross-section. This process gives each generation more diversity of individuals, and therefore, better approximations to the optimum PF.

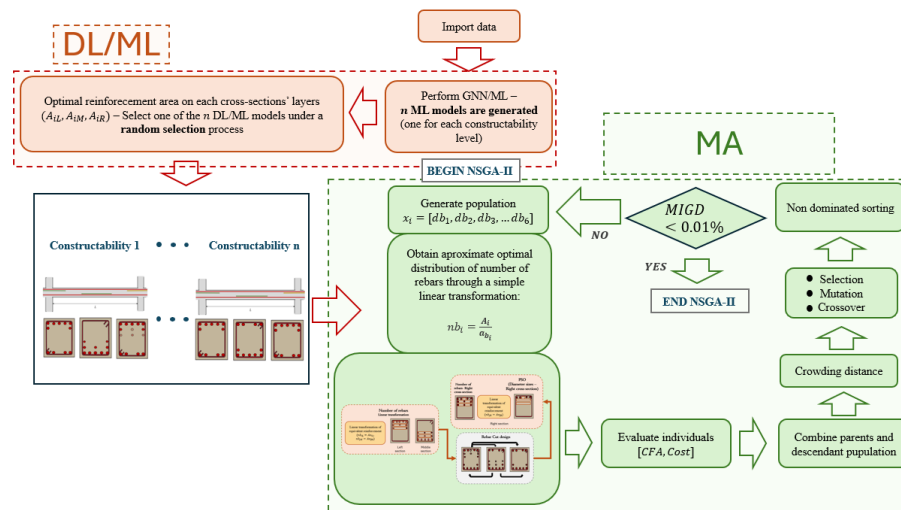


Figure 5. Proposed AI-Enhanced constructability-based MO optimization framework for the design of rebar in concrete beams through MLOT.

4.6 Graph Neural Network representation for rebar design in beams

The versatility of GNN have made them widely popular in applied AI for multiple disciplines, given that many phenomena can be represented as graphs. Particularly, for rebar design in concrete beams, the rebar prototyping of each cross-section is represented by a node, and the relationship between each cross-section design as edges, as shown next in Figure 6. In previous studies, related to application of GNN for this problem of rebar design in beams, usually three critical cross-sections are considered for each span, as it was also the case in this work. Each node consists of six features. Features related to b, h, L, f'_c are considered to be constant for each node

in the graph representation, where the other features vary for each node, namely the critical design bending moments at left, middle and right span cross-sections M_{u-left} , M_{u-mid} , $M_{u-right}$.

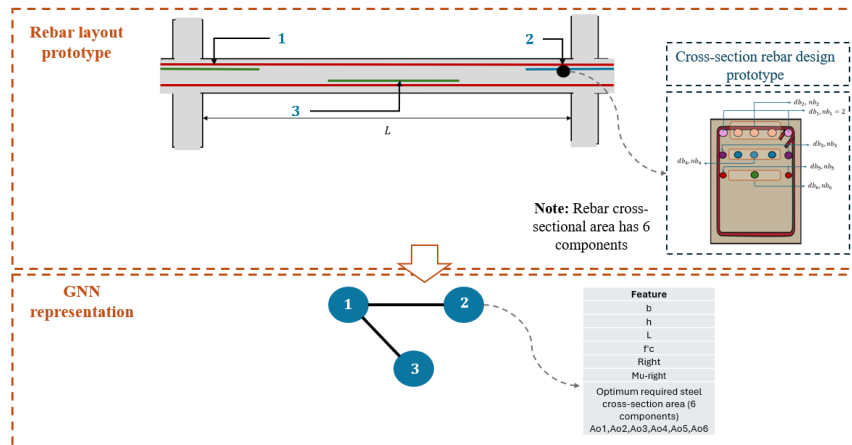


Figure 6. Architecture of the GNN for rebar representation in single-span concrete beams.

5. RESULTS

5.1 Performance of optimization framework

Based To assess and compare the performance of the framework with each ML enhancement (ANN vs GNN), a dataset with a total of 1600 samples was generated. A relation 7/3 was established for training/testing. The regression coefficients (R) are shown as follows in Figure 7 and the summary of performance comparison is presented in Table 1.

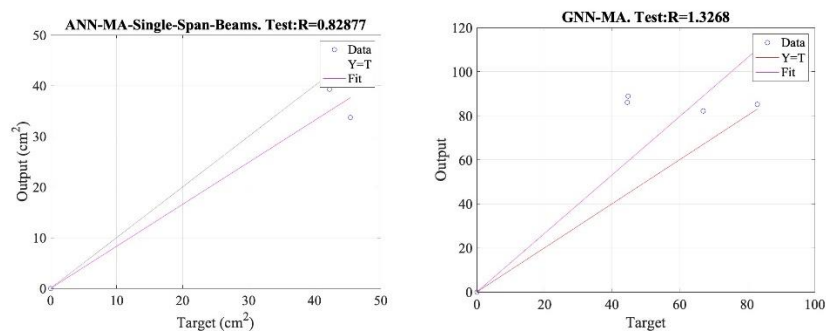


Figure 7. Regression coefficients for test samples. (Left panel) ANN-Enhanced MA, (Right panel) GNN-Enhanced MA.

Table 1. Performance comparison

Performance metric/ML-Enhancement	ANN-MA	GNN-MA
R coef.	0.8287	0.7537
R^2 score	0.9972	0.9986
MRE	0.2686	0.2577
Accuracy	0.7313	0.7455

5.2 Showcase of the framework: convergence and time efficiency

To showcase the proposed optimization formulation, the following beam model of Figure 6 is considered.

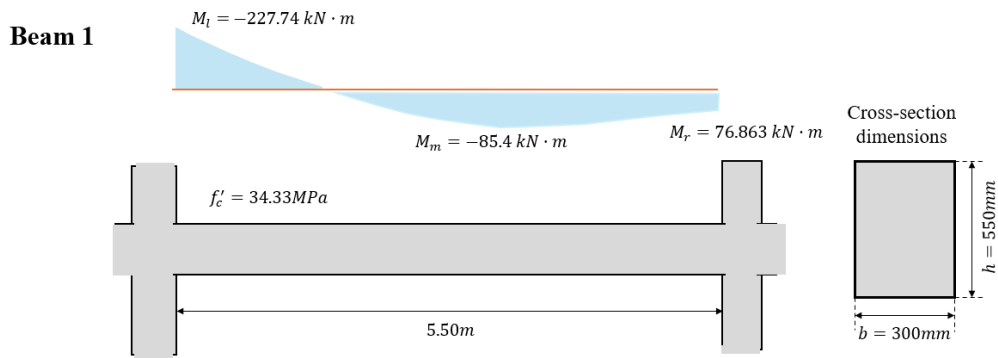


Figure 6. Experimental beam model for showcase of the optimization design proposed framework.

The final PF from each algorithm is shown below, with their respective execution time, for comparison between Non-ML Enhanced MA and ML-Enhanced-MA, and summarized in Table 2, in terms of time and Mean Inverted Generational Distance (MIGD).

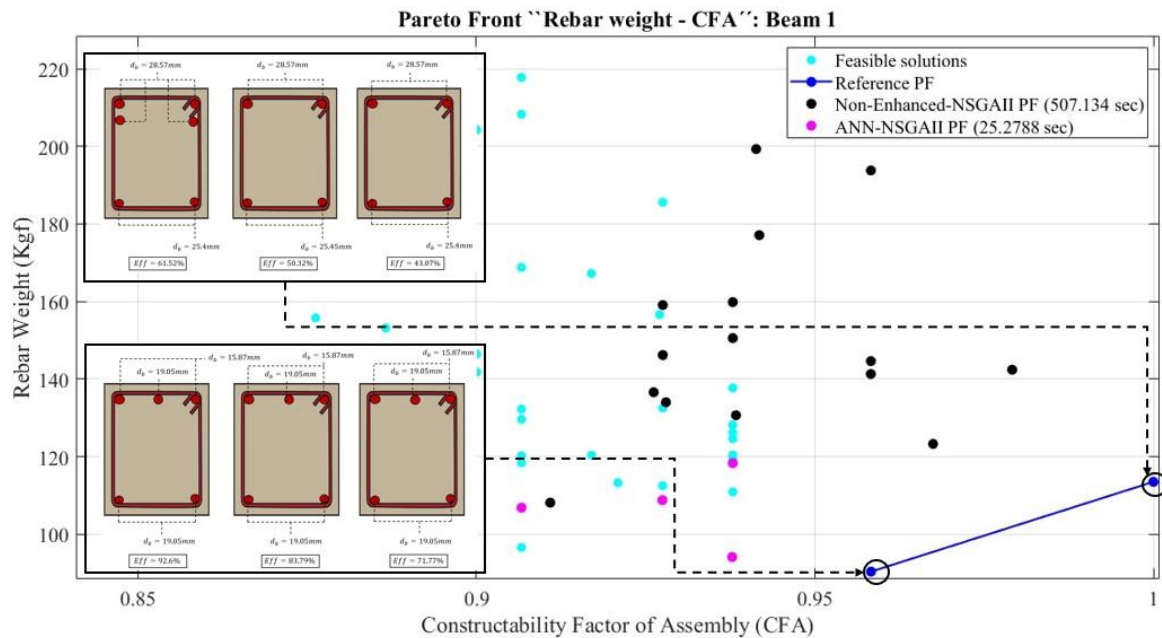


Figure 7. Final PF with each ML-Enhancement.

Table 2. Convergence comparison

ML-Enhancement	Time (sec)	MIGD
Non-Enhanced NSGA-II	507.13	167.56
ANN-NSGAII	25.28	150.44

The evolution of the PFs in time is shown below in Figure 8. The advantages of the AI-Enhancement are evident. The PF evolution for the ANN-NSGAII is more organized and uniform, although less dense than the non-enhanced one.

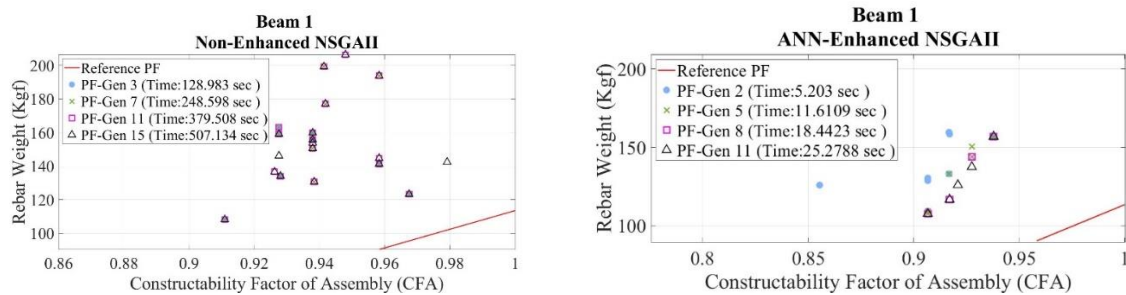


Figure 8. Evolution of PF convergence in time. (Left panel) for the Non-Enhanced NSGA-II, (Right panel) for the ANN-Enhanced NSGA-II.

4. DISCUSSION

The GNN-MA turned out to have slightly better efficiency than the ANN-MA, for most of the performance metrics considered, with respect to regression coefficients.

The dataset used for training and testing was of limited size, and therefore, when using larger datasets, the results might present important differences. Thus, it is important to further investigate the performance of the framework with larger data.

For the GNN, one ANN was used at the beginning to estimate a rebar cross-section area quantity in either of the six components of the rebar prototype for one beam cross-section (see Figure 1). This quantity was used later in the GNN, for re-scaling. That is, the GNN was only used for the estimation of relationship between the different rebar cross-section area quantities of each of the three main beam cross-sections, and its efficiency would, thus, depend on the efficiency of an ANN. To enhance even more the capabilities of the GNN, it is recommended to establish a more proper re-scaling function, based on mechanics of materials.

The superiority of a GNN, in comparison to other ML methods, for this design problem, is expected to be even greater when working with multi-span beams. In this work, only single-span beams were considered, both for training and for showcase. It is recommended to extend the framework for multi-span beams and analyze its performance.

Only one-LOT was considered in this work, for showcase, due to limited availability of time. It is recommended to experiment with MLOT, as mentioned in previous sections, and make comparison of performance between the different levels. For this purpose, a formulation could be made to estimate the required levels of optimum targets, according to a given CS model, depending on the chosen weight factors.

5. CONCLUSIONS

A novel ML-based MO optimization algorithmic framework for reinforcement bar design of rectangular concrete beams was proposed in this work. The advantages obtained in terms of time execution were major, with savings of approximately 95% of the time needed for a Non-ML Enhanced MO optimization algorithm.

The use of GNN in comparison to conventional ANN demonstrated advantages in most of the performance metrics considered, with the exception of regression coefficients. In this sense, the GNN would slightly overestimate the rebar cross-section quantities.

A novel Constructability Score (CS) model for rebar designs in rectangular single-span concrete beams was proposed in this work. With the CS model, savings of material usage of approximately 23.7% were observed for the beam model used for showcase, with only a reduction of 4% of constructability.

ACKNOWLEDGMENTS

I would like to thank the HKUST for the financial support that was provided to execute this research through a full scholarship. Additional thanks to my PhD supervisor Jack C.P. Cheng for his support.

REFERENCES

- Afshari, H., Hare, W., Tesfamariam, S. (2019). Constrained multi-objective optimization algorithms: Review and comparison with application in reinforced concrete structures. *Applied Soft Computing Journal* 83.
- Afzal, M., Liu, Y., Cheng, J.C.P., Gan, V. (2020). Reinforced concrete structural design optimization: A critical review. *Journal of Cleaner Production* 260, 73–79.
- American Concrete Institute ACI-318, (2019). Building Requirements for Structural Concrete and Commentary.
- Coello, C.A., Christiansen, A.D., Hernández, F.S. (1997). Ga simple genetic algorithm for the designs of reinforced concrete beams. *Engineering with Computers* 13, 195–106.
- G. Bekdas, Nigdeli, S. (2012). Cost optimization of t-shaped reinforced concrete beams under flexural effect according to aci-318. In: 3rd European Conference of Civil Engineering (2-4).

- Koumoussis, V.K., Arsenis, S. (2002). Genetic algorithms in optimal detailed design of reinforced concrete members. *Comput-Aided Civ. Inf.* 13(1), 43–52.
- Kwak, H.-G., Kim, J. (2008). Optimum design of reinforced concrete plane frames based on predetermined section database. *Computed-Aided Design* 40(3).
- Lao, W.L., Li, M., Wong, B., Gan, V., Cheng, J.C.P. (2023). Bim-based constructability-aware precast building optimization using optimality criteria and combined non-dominated sorting genetic ii and great deluge algorithm (nsga-ii-gd). *Automation in Construction* 155, 169–188.
- Li, M., Liu, Y., Wong, B., Gan, V., Cheng, J.C.P. (2023). Automated structural design optimization of steel reinforcement using graph neural network and exploratory genetic algorithms. *Automation in Construction* 146, 1439–1446.
- Li M., Wong, B., Lao, Y., Chan. C.M., Gan, V., Cheng, J.C.P. (2021). Dfma-oriented design optimization for steel reinforcement using bim and hybrid metaheuristic algorithms. *Journal of Building Engineerin* 44, 999–1019.
- Shaqfa, M., Orban, Z. (2019). Modified parameter-setting-free- harmony search (psfhs) algorithm for optimizing the design of reinforced concrete beams. *Struct Multidiscip Optim* 60(3), 999–1019.
- V. Govindaraj, Ramasami, J. (2005). Optimum detailed design of reinforced concrete continuous beams using genetic algorithms. *Comput. Struct.* 84(1-2), 34–48.



Part 4

**XR (VR/AR/MR),
Visualization, and Simulation**

AUTOMATIC DETECTION OF BUILDING EQUIPMENT USING LASER INTENSITY OF MLS POINT CLOUD

Arisa Kobayashi¹, Tomohiro Mizoguchi²

1) Master Course Student, Computer Science Course, College of Engineering, Nihon University, Koriyama, Fukushima, Japan. Email: cear23014@g.nihon-u.ac.jp

2) Ph.D., Prof., Department of Informatics and Data Science, Faculty of Engineering, Sanyo-Onoda City University, Sanyo-Onoda, Yamaguchi, Japan. Email: mizoguchi@rs.socu.ac.jp

Abstract: Building Information Modeling (BIM) is being increasingly used in the maintenance of buildings. Since 3D models do not exist for old buildings, it is common to create BIM from 3D scanned point clouds. To date, it has become possible to construct simple BIM consisting of major components such as floors, walls, ceilings, and columns almost automatically. However, automatic construction of detailed BIM including building equipment (lighting, air conditioning, fire alarm, etc.) necessary for building maintenance has not yet been achieved. These are often attached to ceilings and walls, and are difficult to recognize because of their small surface area and thickness. In this paper, we propose a method to detect building equipment using laser reflection intensity to automatically construct detailed BIM from point clouds of buildings by mobile laser scanners (MLS). In this method, first, the effects of distance and incident angle included in the reflection intensity are eliminated based on polynomial approximation, and the reflection intensity value of ceilings and walls made of the same material is approached to a constant value. Next, since the corrected intensity follows a normal distribution, a set of points that deviate from the normal distribution is extracted as building equipment candidates by thresholding. Finally, the point cloud is converted into an image representation, and each equipment is extracted using morphological and labeling process. Through various experiments targeting the ceilings and a wall of buildings, the proposed method achieved a high detection rate.

Keywords: Laser scanning, Point cloud, 3D modeling, Building equipment.

1. INTRODUCTION

Building Information Modeling (BIM), which combines 3D shape and attribute information, is being used in the maintenance of buildings. Since 3D models do not exist for old buildings, it is common to create BIM from 3D scanned point cloud. To date, it has become possible to construct simple BIM consisting of major components such as floors, walls, ceilings, and columns almost automatically (Gourguechon et al., 2022). However, it has not yet been possible to automatically construct detailed BIM including building equipment (lighting, air conditioning, fire alarms, etc.) necessary for the maintenance of buildings, as shown in Figure 1. These are often attached to the ceilings or walls, and have a small surface area and thickness, making it difficult to recognize them by evaluating only their shape (Hossain et al., 2021). Therefore, a method is needed to detect building equipment from information other than the shape (coordinate values) of the point cloud and incorporate it into simple BIM to construct detailed BIM.

In this paper, we propose a method to detect building equipment using laser reflection intensity contained in the scanned point cloud obtained by mobile laser scanners (MLS) to automatically reconstruct detailed BIM. Compared to conventional terrestrial laser scanners (TLS), MLS has the advantage of being able to scan a wide area efficiently and has rapidly become popular in recent years (Stefano et al., 2021). Laser reflection intensity is a value that indicates the intensity of reflected light from the target surface. Since it varies depending on the color and material of the target, it is considered to be extremely effective for detecting building equipment in the point cloud. However, since the reflection intensity value obtained from the scanner includes the effects of the distance and incident angle, it is necessary to appropriately exclude them and correct it to a value that reflects only the difference in color and material. Generally, ceilings and walls are wide and flat, and made of the same material, and building equipment of different colors and materials are attached there. Therefore, the basic idea of this study is that if the reflection intensity value of ceilings and walls can be appropriately corrected and approached close to a constant value, point sets with values different from them can be detected as building equipment.

Object recognition in point clouds generally consists of an extraction step, in which partial point clouds corresponding to each object are individually extracted, and a classification step, in which the class of each object (lighting, fire alarm, etc.) is estimated. Of these two steps, this study only deals with the first step, the extraction step. If extraction is possible, we believe that by using the corresponding partial images for each object in the classification step, highly accurate classification will be possible using AI with appropriate image processing.

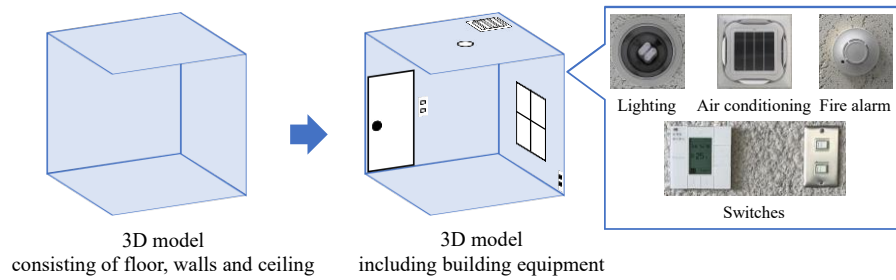


Figure 1. Construction of detailed BIM from point cloud

2. RELATED WORKS

2.1 BIM Reconstruction from Point Cloud

There have been several studies on detecting building equipment from point clouds. Akiyama et al. (Akiyama et al., 2023) proposed a method to detect building equipment on the ceiling surface from TLS scanning point clouds. First, as preprocessing, 2D ceiling regions are extracted from the point cloud data and structured. Then, it detects partial point clouds corresponding to each building equipment such as lighting and air conditioning, and then fits circles and rectangles to their boundary points. However, it is difficult to detect objects with small thickness using only the shape information of the point cloud. In addition, the detection rate decreases when the scanning distance exceeds 5 meters, so the number of scans increases as the target area becomes wider. Note that classification processing was not performed in this study. Pan et al. (Pan et al., 2022) proposed a method to detect building equipment using both images and point clouds. First, two 3D point clouds are obtained by SfM-MVS and laser scanning, and then these are synthesized. On the other hand, it used a convolutional neural network (CNN) on video data to detect and classify building equipment inside buildings and added them to the point cloud to construct detailed BIM. However, since a 3D model is constructed by combining two types of point cloud data, there are problems such as cumbersome processing and a large amount of time. In addition, while object detection using CNN is easy to detect objects with small shape variations such as fire extinguishers and fire alarms, a large amount of training data must be prepared for objects with large shape variations such as ceiling lights. Anjanappa et al. (Anjanappa et al., 2023) also detected and classified building equipment inside buildings using CNN on 3D point clouds. However, similar to the above, it is necessary to prepare sufficient training point cloud data to detect a wide variety of building equipment.

2.2 Laser Intensity Correction

The reflectance intensity values obtained from a laser scanner are generally affected by the reflectance of the object as well as the distance, incident angle, sensor characteristics, and atmospheric conditions, so it is not recommended to use them directly without correction (Bai et al., 2023).

When using one scanner to scan in a short time under the same environment, there is no need to consider the effects of the sensor characteristics and atmospheric conditions mentioned above, and the value changes according to the radar equation due to the reflectance of the target, the incident angle, and the distance. Furthermore, these three factors are independent of each other and can be expressed by Equation (1) (Tan & Cheng, 2020).

$$I(\rho, \theta, d) = f_1(\rho) \cdot f_2(\theta) \cdot f_3(d) \quad (1)$$

Where $I(\rho, \theta, d)$ is the initial reflection intensity, f_1 , f_2 , and f_3 are the functions of reflectance of the target ρ , incident angle θ , and distance d respectively.

Theoretically, the reflection intensity is directly proportional to the cosine of the incident angle and inversely proportional to the square of the distance. However, considering the complexity of the effects of the distance and incident angle, a method of eliminating the effects of the distance and incident angle using a polynomial, as shown in Equation (2), has been found to be effective (Tan & Cheng, 2020).

$$I_{cor}(\rho) = I(\rho, \theta, d) \cdot \frac{N_2(\alpha_i \theta^i)}{N_2(\alpha_i \theta^i)} \cdot \frac{N_3(\beta_i d^i)}{N_3(\beta_i d^i)} \quad (2)$$

Where α_i and β_i are polynomial parameters, and N_2 and N_3 are the degrees of the polynomial. To estimate the polynomial parameters, it is necessary to derive the relationship between the reflected intensity and the distance or incident angle, and define a correction equation.

One method for calculating the correction equation is to use a target and find the relationship between

reflection intensity, distance, and incident angle, respectively. In the method by (Tan & Cheng, 2015), the TLS point cloud was mainly used, and the target was scanned multiple times from different distances and angles in a laboratory, and a polynomial for correction was calculated based on the obtained reflection intensity values. First, the target was scanned from a fixed distance between 0deg and 90deg at small intervals (for example, 5deg pitch) to obtain samples of angle and reflection intensity. Similarly, it is scanned while changing the distance at a fixed pitch while keeping the angle constant, and it obtains samples of distance and reflection intensity. Least-squares fitting of polynomial to this set of sample points, we can obtain correction equations that express the relationship between the incident angle-reflection and the distance-reflection intensity. However, the correction equation for the distance-reflection intensity is more complicated than that for angle and requires scans at fine intervals up to distances of several tens of meters, which results in problems such as a huge amount of time required for the experiment. Therefore, a method has been proposed to calculate the correction equation for the distance and reflection intensity from the actual scanning point cloud (Tan & Cheng, 2020). First, angle correction is performed as shown in Equation (3) from the correction equation that expresses the relationship between the incident angle and reflection intensity calculated using the target as described above.

$$I_{\theta}(\rho, d) = I(\rho, \theta, d) \cdot \frac{\sum_{i=0}^{N_2} (\alpha_i \theta^i)}{\sum_{i=0}^{N_2} (\alpha_i \theta^i)} \quad (3)$$

The obtained incident angle corrected reflection intensity $I_{\theta}(\rho, d)$ is a reflection intensity that is affected only by the distance, so the distance-reflection intensity correction equation is calculated using least squares regression on the reflection intensity after incident angle correction. The final corrected reflection intensity $I_{cor}(\rho)$ can be calculated using Equation (2) based on the calculated polynomial parameters.

On the other hand, in addition to the method using the polynomial mentioned above, there is also a method of correction using a look-up table (LUT) (Jeong & Kim, 2018). This method focuses on the fact that the ratio of the reflection intensity of two scanning surfaces at the same distance and incident angle must be the same as the ratio of the reflectance. A LUT is composed of 2D cells with the vertical axis as the distance and the horizontal axis as the incident angle and is created by calculating the average value and standard deviation of the reflection intensity for the distance and incident angle corresponding to each cell. To obtain the data used for creation, a point cloud scanned on a wide plane made of the same material, such as a large wall several meters square, is used as a reference surface. Using this LUT, the reflectance of the reflection intensity of the scanning surface can be estimated as shown in Equation (4).

$$I_{cor}(\rho) \propto I(\rho_{ref}) \cdot \frac{I(\rho, \theta, d)}{I(\rho_s, \theta, d)} \quad (4)$$

Where the reflection intensity $I(\rho_s, \theta, d)$ of the reference surface is the average value of the reflection intensity corresponding to each cell of the LUT. Also, the true reflectance of the reference surface $I(\rho_{ref})$ is unknown, so it is set arbitrarily.

Different from the above-mentioned method, a method has been proposed to calculate the correction equation using only the obtained scanning point cloud without preparing a correction equation or LUT in advance. For example, when measuring a road surface using a vehicle-mounted LiDAR, if the height between the scanner and the road surface is constant, the incidence angle can be described as a function of the distance. Therefore, as shown in Equation (5), by correcting only the effect of the distance, the incidence angle can also be corrected at the same time (Wan et al., 2019).

$$I_{cor}(\rho) = I(\rho, \theta, d) \cdot \frac{\sum_{i=0}^{N_3} (\beta_i d^i)}{\sum_{i=0}^{N_3} (\beta_i d^i)} \quad (5)$$

The authors have demonstrated a certain degree of effectiveness of the proposed reflection intensity correction through experiments on extracting road markings from MLS point clouds. Although this method requires certain scanning conditions, correction is possible using only the point cloud, and it can be widely deployed in a variety of applications.

3. METHOD

3.1 Overview of The Proposed Method

In this method, first, the area to be modeled is manually extracted, and preprocessing is performed to sample the point cloud so that the point density is spatially uniform. After that, the effects of distance and incident angle contained in the reflection intensity are eliminated based on polynomial approximation, and the reflection

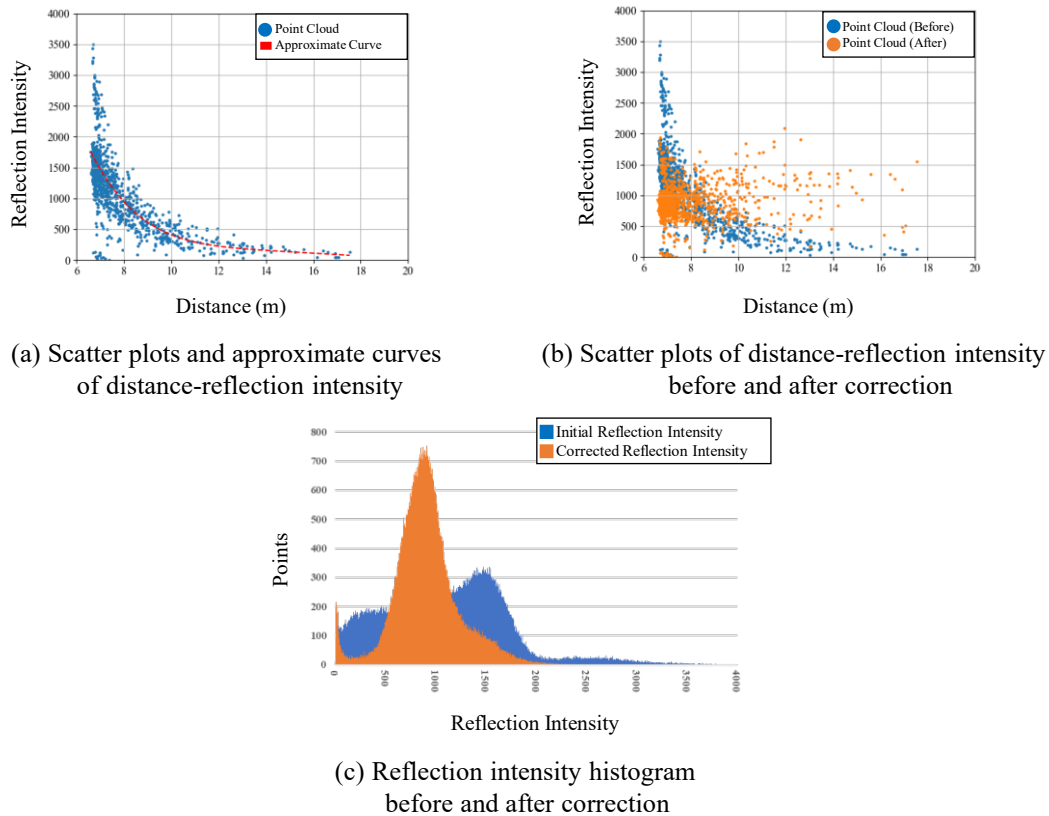


Figure 2. Reflection intensity correction

intensity value of walls and ceilings made of the same material is corrected to approach a constant value. Next, since the corrected reflection intensity follows a normal distribution, a set of points that deviate from the normal distribution are extracted as equipment candidates by thresholding. Finally, the point cloud is converted into a binary image, and each equipment is extracted using morphological and labeling process. In this report, we focus on the area scanned while keeping the distance between the scanner and the scanning surface constant. The details are described below.

3.2 Pre-processing

The point cloud obtained from the scanner is high-density data that includes the surrounding area of the target. Furthermore, as the scanning target becomes wider, the amount of data becomes enormous, ranging from tens of millions to hundreds of millions of points. Therefore, to make the point cloud easier to handle, we use the free point cloud processing software Cloud Compare to cut out the target area and thin out the point cloud at 1 cm intervals.

3.3 Reflection Intensity Correction

In this method, the distance and incident angle are simultaneously corrected using the method of Wan et al. (Wan et al., 2019) introduced in Section 2.2. A scatter plot is created from the scanning point cloud, with the horizontal axis as the distance and the vertical axis as the reflection intensity, and an approximate polynomial is obtained for this set of points by the least-squares method. Referring to the method of Wan et al., the degree of the polynomial was set to $N_3 = 3$. Furthermore, ds in Equation (5) was set to the average value of the distance. Figure 2(a) shows the result of superimposing the scatter plot of distance-reflection intensity and the calculated correction equation. Figure 2(b) shows the sample points before and after correction. It can be confirmed that the reflection intensity value, which decreased with the distance, is approaching a constant value regardless of the distance. This can also be visually confirmed from the point cloud before and after correction shown in Figure 3(a) and (b). From the histogram of reflection intensity before and after correction shown in Figure 2(c), the reflection intensity value after correction approximately follows a normal distribution. This is thought to be because the reflection intensity values of points corresponding to areas with a large surface area and made of the same material were appropriately corrected.

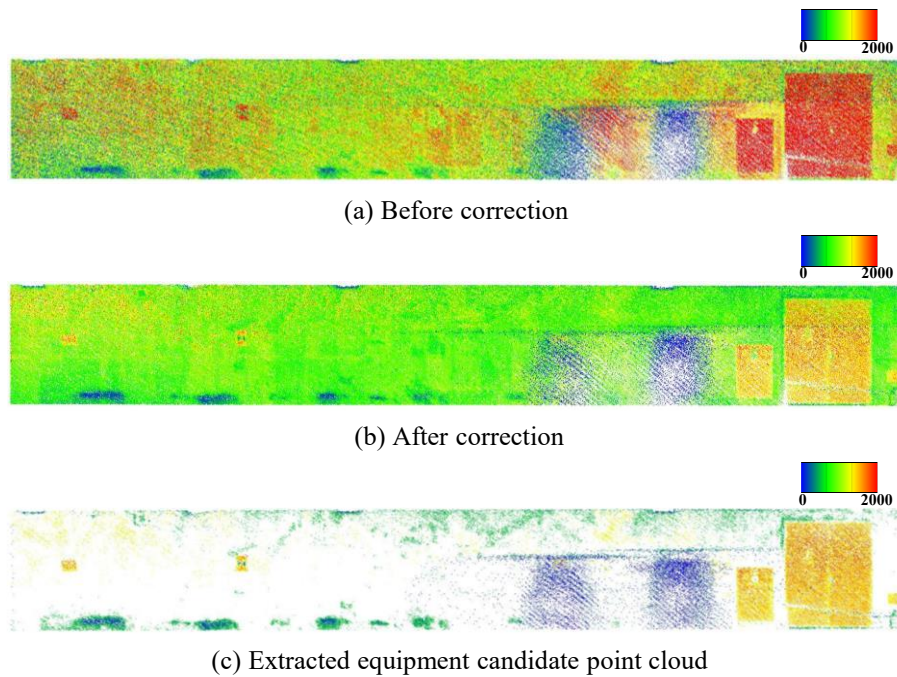


Figure 3. Results of reflection intensity correction on point cloud

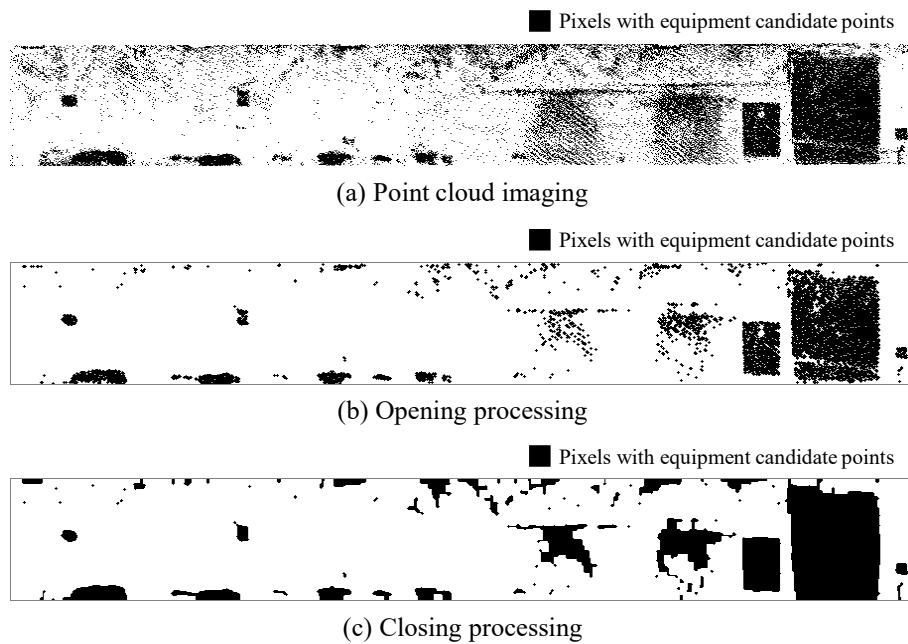


Figure 4. Morphological process

3.4 Extraction of Equipment Candidate Points

From the normalized and corrected reflection intensity, it is considered that points corresponding to walls and ceilings are within a certain range from the average value, and points corresponding to building equipment are outside the range. Therefore, using the average value μ and standard deviation σ of the reflection intensity values, the points corresponding to equipment candidates are extracted by thresholding. The threshold is set as $th = \mu \pm \alpha \sigma$, where α is a user-set parameter. Figure 3(c) shows an example of the extracted equipment candidate point cloud.

3.5 Imaging of Point Cloud

From the equipment candidate point cloud obtained in the process in the previous section, the point cloud is grouped by equipment. First, the candidate point cloud is projected onto the planes corresponding to walls or

Table 1. Specification of scanners

Instrument	OS0 (Ouster)	VLP-16 (Velodyne)
Weight	430g	830g
Maximum distance	75m (80% reflectivity) 35m (10% reflectivity)	100m
Accuracy	$\pm 2.5\text{cm}$ (Lambert Target) $\pm 5.0\text{cm}$ (Retroreflective Target)	$\pm 3\text{cm}$
Field of view	$360^\circ \times 90^\circ$	$360^\circ \times 30^\circ$
Data acquisition rate	5.2M points/second	0.3M points/second

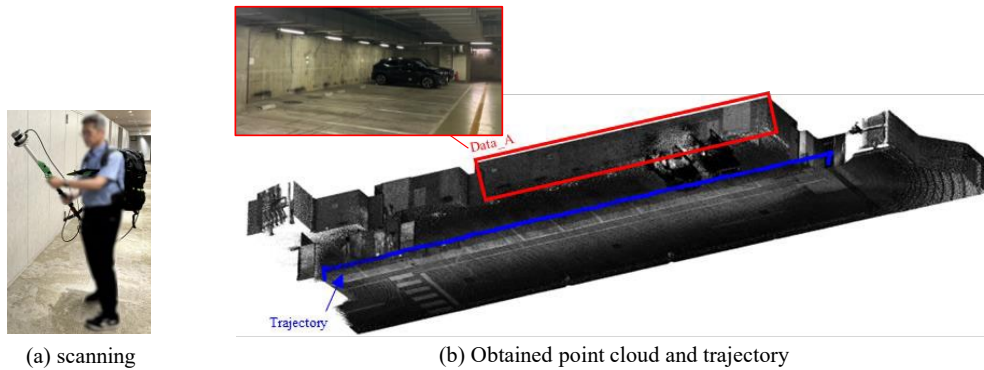


Figure 5. Scanning with Ouster LiDAR OS0

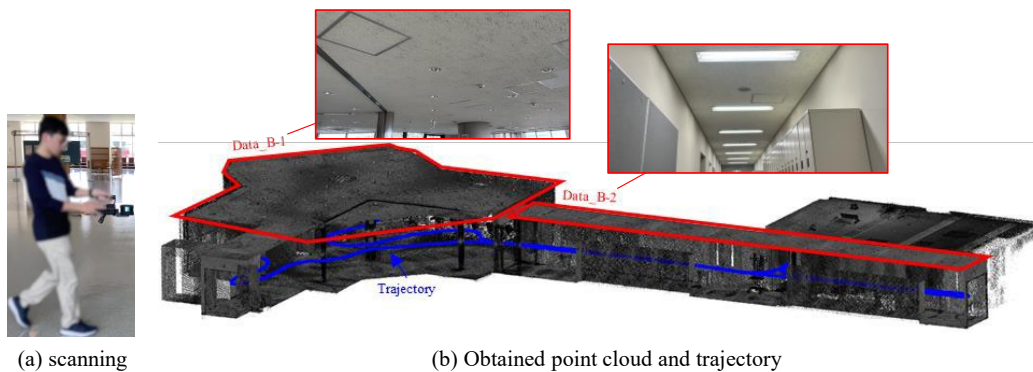


Figure 6. Scanning with Velodyne LiDAR VLP-16

ceilings. And then the projection plane is gridded to create a binary image. The grid size (image resolution) was set to 2 cm. An example is shown in Figure 4(a). In the binary image, pixels that contain one or more points are shown in black, and the others are in white.

3.6 Extraction of Equipment Candidates

From the image obtained through the processing in the previous section, it can be seen that pixels other than the original equipment have been incorrectly extracted, and that there are areas where pixels identified as single equipment have been disconnected. Therefore, morphological process is performed on the created binary image. Morphological process is based on dilation process which expands the binarized area by one pixel, and erosion process which shrinks inversely. In the opening process, it first iterates shrinking n times and then iterates the same number of expanding. This enables to remove isolated pixels. However, if the number of iterations is too large, even one piece of equipment may be divided into multiple pieces, or small equipment may be removed and cannot be extracted. On the other hand, in the closing process, it first iterates expanding and then shrinking n times. It has the effect of closing small holes in binary images and connecting unconnected components. However, if the number of repetitions is too large, adjacent multiple equipment will be connected and extracted as one. We set the number of iterations n to one to four depending on the dataset. Figure 4(b) and (c) shows images after opening and closing process respectively. It can be confirmed that the opening process removes isolated pixels and the closing

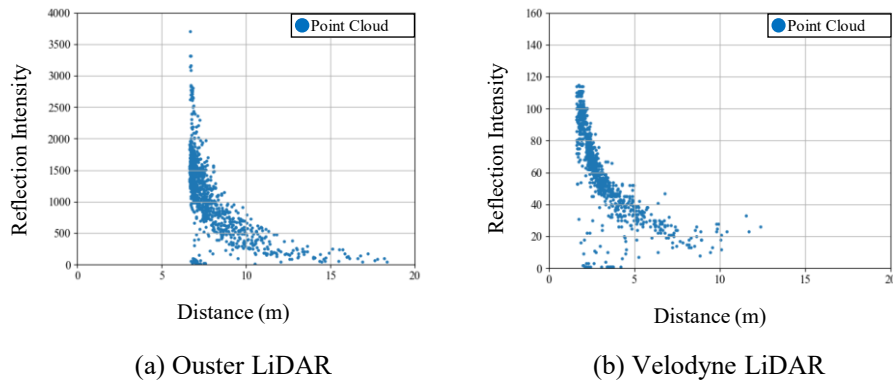


Figure 7. Reflection intensity distribution of Ouster and Velodyne LiDAR

Table 2. Results of reflection intensity correction

Instrument	Target area		Average	Standard deviation
Ouster LiDAR	Wall of underground parking (Data_A)	Before	1101.0	605.5
		After	904.9	309.3
Velodyne LiDAR	Ceiling of elevator hall (Data_B-1)	Before	66.4	27.2
		After	61.9	16.6
	Ceiling of the hallway (Data_B-2)	Before	78.7	22.5
		After	76.0	16.6

process connects disconnected pixels.

Finally, the labeling process is used to assign a label to each connected component in the binary image and extract the equipment in the image. The same label is also assigned to the points inside the cells.

4. RESULTS

4.1 Test Site and Scanners

The proposed method was applied to point cloud datasets obtained from two different scanners, Ouster OS0 and Velodyne VLP-16. Table 1 shows the specifications of each scanner.

OS0 has a viewing angle of 360° horizontally and 90° vertically and is prone to blind spots in the vertical direction. Therefore, as shown in Figure 5(a), the scanner mounted on a monopod was held at an angle so that the laser hit the ceiling surface vertically and scanned the underground parking lot while walking at a normal speed. As shown in Figure 5(b), the scanning point cloud of the wall surface of the underground parking lot (Data_A) used in the experiment was scanned by walking parallel to one direction while keeping a position about 6 m away from the target wall surface. The scan time was about 3 minutes, and the total number of points was about 1M points which was then reduced to 0.5M points after the sampling process. The smallest building equipment identified from this dataset was a rectangular fire extinguisher sign approximately $25\text{ cm} \times 20\text{ cm}$.

On the other hand, VLP-16 has a viewing angle of 360° horizontally and 30° vertically, and this also tends to have blind spots in the vertical direction. Therefore, as shown in Figure 6(a), a mobile laser scanner Hovermap from Emesent equipped with the VLP-16 was held parallel to the floor and scanned the interior of the university building while walking at a normal speed. By rotating the VLP-16 attached to the tip of the Hovermap, it is possible to evenly irradiate a laser over a $360^\circ \times 360^\circ$ area. Figure 6(b) shows the scanning point cloud of the ceiling surface of the elevator hall (Data_B-1) and the ceiling surface of the corridor (Data_B-2) used in the experiment. The scanning time was about 5 minutes. The total number of points in Data B-1 was about 4M and then sampled to 1.3M. The total number of points in Data B-2 was about 2M and sampled to 0.6M. The smallest building equipment identified from this dataset was a circular fire alarm with a diameter of approximately 10 cm.

4.2 Results of Intensity Correction

Figure 7 shows the results of comparing the reflection intensity distribution of point clouds scanned using OS0 and VLP-16. A study by Viswanath et al. (Viswanath et al., 2024) shows that there are significant differences in the reflection intensity values obtained from these two scanners. The reflection intensity values obtained with

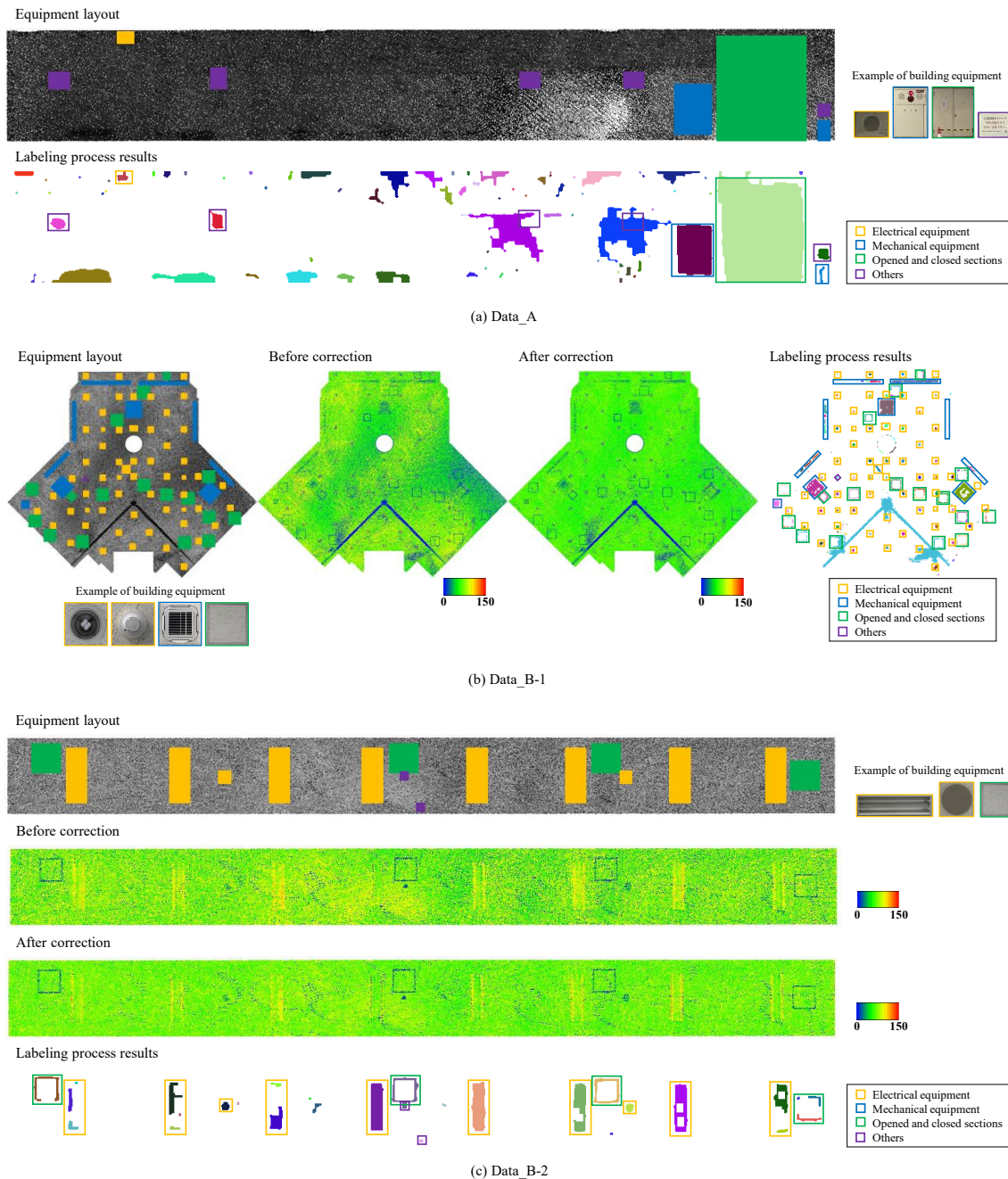


Figure 8. Visual evaluation of building equipment

Ouster LiDAR are raw data that have not been pre-calibrated, while those obtained with Velodyne LiDAR are data that have been calibrated for distance and laser power, as described in the manual. The data obtained from OS0 showed a trend like that described in the paper, as shown in Figure 7(a). However, the data obtained from VLP-16 showed a tendency for the reflection intensity to decrease as the distance increases as shown in Figure 7(b). Therefore, as introduced in Section 3.3, the reflection intensity was corrected using the method of Wan et al. (Wan et al., 2019) for both data sets.

Table 2 shows the results of reflection intensity correction. Data_A was corrected using the 6th-order equation since we found it tightly approximated the sampled points from experiments than 3rd-order. Comparing the initial and corrected reflection intensity, the standard deviation is smaller after correction for both Ouster and Velodyne LiDAR, confirming the correction effect. Therefore, it was found that simultaneous correction of distance and incident angle using Equation (5) can be applied when scanning while keeping the distance between the scanning device and the surface constant. As in Figure 2(c), it was also confirmed visually that the corrected reflection intensity roughly follows a normal distribution.

Table 3. Result of building equipment detection

	Data A		Data B-1		Data B-2	
Number of equipment	9	-	87	-	16	-
Number of extracted regions	77	855.6%	172	197.7%	28	175.0%
Correct detection	9	100.0%	124	142.5%	23	143.8%
Miss detection	0	0.0%	0	0.0%	0	0.0%
Over detection	68	88.3%	48	27.9%	5	17.9%

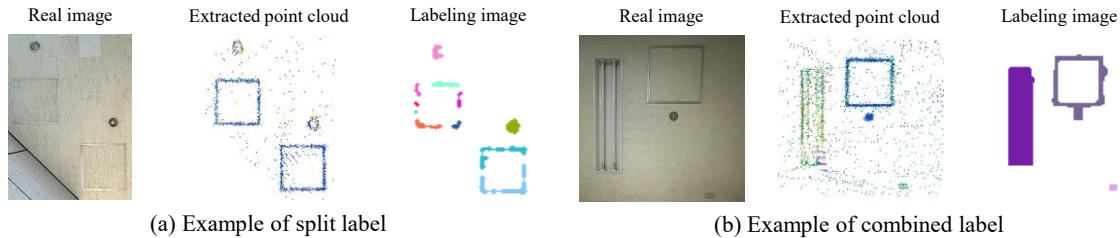


Figure 9. Results of labeling process

4.3 Results of Equipment Detection

Figure 8 shows the results of visual detection of building equipment. The equipment layout was visually compared with the actual site images. And if more than one labeled pixel exists in the corresponding position, the building equipment was considered to have been detected correctly. Table 3 shows the detection results for building equipment. We set $\alpha=1.0$ for Data_A and $\alpha=2.0$ for Data_B-1 and Data_B-2.

Overall, many regions were extracted compared to the actual number of equipment. Correct detections matched the actual number of equipment in Data_A, but were detected more than the actual number of equipment in Data_B-1 and Data_B-2. Conversely, there were no miss detections. Over detections differed greatly depending on the data.

As for Data_A, our method could detect nine all equipment. However, two of them could not detect only equipment correctly and extracted as wide areas including its surrounding wall surface. The reason is that cars were parked in front of the signs while scanning. And it made it difficult to scan correctly because they were obstructed, and the reflection intensity value of the corresponding wall area was abnormally low. The large number of extracted regions is due to the high over detection rate of 88.3%. This is because the material of the wall used in this study is concrete, which is easily stained. Therefore, it is believed that the intensity value changes depending on the degree of stains and some of these are detected as equipment candidates. Our method assumes that the target area is composed of the same material, but this concrete surface does not follow it due to the stains. Therefore, it is difficult to extract only the equipment by extracting only the reflection intensity.

As for Data_B-1 and Data_B-2, it has low over detection rates of 27.9% and 17.9% respectively. Since these indoor data well follow the assumption, good results were obtained. The large number of extracted regions is due to the high correct detection rates of 197.7% and 175.0%, respectively. This is due to the image processing aspects of the proposed method rather than the intensity correction. Even if morphological process is applied, the pixels are not sufficiently connected, and the labels are detected as divided for equipment. For example, this was typically seen in cases of ceiling inspection hatches which are composed of long and thin metal frames, as shown in Figure 9(a). Conversely, the multiple pieces of equipment were sometimes labeled as a single unit when they were located close to each other, as shown in Figure 9(b). Another example is shown in Figure 8(b). In this figure, our method did not work well for the pillars of the fire shutters made of metal and detected a wide area including the surrounding equipment as a single unit. Therefore, it is necessary to revise the concatenation conditions of the pixels extracted as equipment candidates for more correct detection.

5. CONCLUSIONS

In this study, we proposed a method of detecting building equipment based on reflection intensity correction with the aim of constructing detailed BIM from MLS scanned point clouds. In the proposed method, we first demonstrated that the effects of distance and incident angle contained in the laser reflection intensity can be eliminated from the scanned point cloud of ceilings and walls obtained by MLS using polynomial approximation. And we verified that the reflection intensity of flat ceilings and walls made of the same material can be corrected to approach a constant value. We also demonstrated that the corrected reflection intensity can be used to detect building equipment without omission through thresholding and morphological process. An advantage of the proposed method is that it uses MLS, so it can scan a wide area more efficiently than conventional

methods using TLS or SfM. In addition, compared to conventional methods using images and deep learning, it has the advantage that no preparation of training data is required and extraction process can be performed without being affected by lighting or shadows.

A limitation of the proposed method is the case where there is variation in the reflection intensity of the background as shown in the example of Figure 8. When there are differences in the color or material in the background, such as stains on the wall or deteriorated paint, and the reflection intensity of the background varies, it may erroneously extract things other than the equipment. A possible solution to this problem is to use the image and reflection intensity in combination for comprehensive analysis, rather than using them alone. Future issues include removing regions where equipment is over-detected and dealing with regions where equipment is divided into multiple parts. We will also compare the method with other correction methods and consider its application to buildings with non-flat shapes. We also need to consider how to classify the equipment classes obtained.

ACKNOWLEDGMENTS

The authors would like to express their gratitude to KYUDENKO CORPORATION for their support in carrying out this study.

REFERENCES

- Akiyama, R., Date, H., Kanai, S., and Yasutake, K. (2023). FOOTPRINT DETECTION OF CEILING EQUIPMENT FROM TLS POINT CLOUDS, *Proceedings of the 5th International Conference on Civil and Building Engineering Informatics*, 465-470.
- Anjanappa, G., Nikoohemat, S., Oude Elberink, S., Voûte, R. L., and Lehtola, V.V. (2023). NEEDLE IN A HAYSTACK: FEASIBILITY OF IDENTIFYING SMALL SAFETY ASSETS ROM POINT CLOUDS USING DEEP LEARNING, *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, X-1/W1-2023, 461-468.
- Bai, J., Niu, Z., Gao, S., Bi, K., Wang, J., Huang, Y., and Sun, G. (2023). An exploration, analysis, and correction of the distance effect on terrestrial hyperspectral LiDAR data, *ISPRS Journal of Photogrammetry and Remote Sensing*, 198, 60-83.
- Di Stefano, F., Chiappini, S., Gorreja, A., Balestra, M., and Pierdicca, R. (2021). Mobile 3D scan LiDAR: a literature review, *Geomatics, Natural Hazards and Risk*, 12 (1), 2387-2429.
- Hossain., M., Ma., T., Watson., T., Simmers., B., Khan., J. A., Jacobs., E. and Wang., L. (2021). Building Indoor Point Cloud Datasets with Object Annotation for Public Safety, *Proceedings of the 10th International Conference on Smart Cities and Green ICT Systems (SMARTGREENS 2021)*, 45-56.
- Jeong, J., and Kim, A. (2018). LiDAR Intensity Calibration for Road Marking Extraction, *2018 15th International Conference on Ubiquitous Robots (UR)*.
- Pan, Y., Braun, A., Brilakis, I., and Borrmann, A. (2022). Enriching geometric digital twins of buildings with small objects by fusing laser scanning and AI-based image recognition, *Automation in Construction*, 140, 104375.
- Tan, K., and Cheng, X. (2015). Intensity data correction based on incidence angle and distance for terrestrial laser scanner, *Journal of Applied Remote Sensing*, 9, 094094.
- Tan, K., and Cheng, X. (2020). Distance Effect Correction on TLS Intensity Data Using Naturally Homogeneous Targets, *IEEE GEOSCIENCE AND REMOTE SENSING LETTERS*, 17 (3), 499-503.
- Viswanath, K., Jiang, P., PB, S., and Saripalli, S. (2024). Off-Road LiDAR Intensity Based Semantic Segmentation, arXiv:2401.01439 [cs.CV].
- Wan, R., Huang, Y., Xie, R., and Ma, P. (2019). Combined Lane Mapping Using a Mobile Mapping System, *Remote Sens.* 2019, 11(3), 305.

BIM-AR Driven Collaborative Management of Complex MEP Equipment Installation

Youde Zheng¹, Liyun Yao², Silin Li³, Yugui Zhou⁴, Yi Tan^{1*}

1*) Ph.D., Assoc. Prof., Sino-Australia Joint Research Center in BIM and Smart Construction, Shenzhen University, Shenzhen, China. Email: tanyi@szu.edu.cn

1) Master's Candidate, Sino-Australia Joint Research Center in BIM and Smart Construction, Shenzhen University, Shenzhen, China. Email: zhengyoude2023@email.szu.edu.cn

2) Master's Candidate, Sino-Australia Joint Research Center in BIM and Smart Construction, Shenzhen University, Shenzhen, China. Email: 2410214014@mails.szu.edu.cn

3) Ph.D. Candidate, Sino-Australia Joint Research Center in BIM and Smart Construction, Shenzhen University, Shenzhen, China. Email: 529063420@qq.com

4) Engineer, China Construction Fifth Division Southern Construction Subsidiary CORP.LTD, China. Email: 82066491@qq.com

Abstract: With the increasing complexity of construction projects, the efficient installation and management of Mechanical, Electrical, and Plumbing (MEP) equipment has become particularly important. Currently, the management of complex MEP equipment faces challenges of multi-disciplinary collaboration, errors, and collisions during installation, as well as inefficient progress management and problem feedback. Therefore, this paper proposes a BIM and Augmented Reality (AR) driven collaborative management method for the installation of complex MEP equipment, including intelligent scene localization, accurate matching of model information, and collaboration of multi-terminal devices to reduce errors and improve efficiency. In this method, the physical built structural scene is firstly collected and accurately aligned with the MEP BIM model, and then the MEP component data is processed and matched using Dynamo and Unity. The human-machine interaction functions of installation guidance and review, information display and filtering, progress simulation, and data management are then realized for the complex MEP equipment. Additionally, the collaborative application and data management of multi-terminal devices utilizing AR headsets and handheld devices significantly improves the efficiency of communication. Through the experiment in a public building project under construction in Shenzhen, the feasibility and application effect of the method are verified, which provides strong support for the intelligent construction and collaborative management of complex MEP equipment, and has a wide range of application prospects.

Keywords: Augmented Reality; BIM; Collaborative Management; Human-machine interaction; MEP;

1. INTRODUCTION

The installation and management of mechanical, electrical, and plumbing (MEP) equipment face challenges as the scale and complexity of construction projects increase (Xie et al., 2022). Complex MEP equipment in this study refers to complex, narrow and curved MEP piping systems in large public buildings involving highly interwoven and space-constrained layouts, and traditional construction management methods have limitations in information transfer, progress control, and coordination (Li et al., 2022), which often lead to equipment installation conflicts, accuracy deviations, rework and schedule delays (Y. Wang et al., 2022; Zhao et al., 2021), thus affecting project quality and efficiency. Therefore, how to effectively manage complex MEP equipment and improve the level of construction intelligence is significant.

Building Information Modeling (BIM) technology provides a digital three-dimensional management platform for construction projects, which can realize information sharing in the design, construction, and operation phases (Vignali et al., 2021). However, the application of traditional BIM models mostly stays at the design level, and the real-time interaction with the construction site is insufficient, especially in the installation guidance and schedule control of complex MEP equipment. Augmented Reality (AR) technology provides a new solution for the application of BIM in construction by virtue of the fusion of reality and reality (Chu et al., 2018). Through AR devices, construction personnel can view the equipment installation location and parameter information in the BIM model in real time, which can dynamically guide the construction and reduce errors and communication costs. (Al-Sabbag et al., 2022; Catbas et al., 2022)

However, the existing combined BIM-AR applications mostly focus on the display of simple equipment, and lack systematic solutions for complex MEP equipment. In practical applications, there is still a lack of exploration on how to realize the accurate alignment of BIM model and field environment, the sharing of model geometry and semantic information, and the collaborative management of multi-terminal equipment.

To this end, this paper proposes a BIM-AR-driven collaborative management method of complex MEP equipment installation. First, a panoramic camera is used to capture the building scene to achieve high-precision alignment between the BIM model and the project scene. Subsequently, Dynamo and Unity are used to process and match the semantic information of components to ensure the sharing of BIM data among multiple devices. In the end, the head-mounted AR device and handheld device are used to realize installation guidance and review of

hidden components, information display and screening, progress simulation, and data management among multiple devices, which improves the efficiency of construction management and collaboration level.

This paper is organized as follows: Section 2 reviews the research on the application of BIM in MEP and the application of AR in architecture; Section 3 introduces the methodology proposed in this paper and the implementation details; Section 4 shows the experiments and results in a public building project in Shenzhen; Section 5 summarizes the main conclusions and looks forward to the future direction.

2. RELATED WORK

2.1 The Application of BIM in MEP

In recent years, with the rapid development of BIM technology, its application in MEP equipment has received wide attention. Pärn et al. (2018) investigated the application of BIM technology in multidisciplinary collaboration, but also pointed out that design conflicts still cannot be completely avoided in complex MEP equipment. J. Wang et al. (2016) developed a BIM framework that combines design, construction, and prefabrication, and it significantly improves the efficiency during the construction phase. In the O&M phase, Kalasapudi et al. (2014) demonstrated how laser scanning point clouds can be combined with BIM models for spatial analysis and facility management of MEP equipment, which provides effective support for long-term maintenance and operation, and Cheng et al. (2020) improved the efficiency of facility management of MEP equipment through the combination of BIM models and Internet of Things (IoT) technology.

Although BIM technology has shown many advantages in MEP facilities, there are still problems of design conflicts in multidisciplinary collaboration, lagging real-time feedback in the construction phase, and insufficient interaction between dynamic updates and actual scenarios in operation and maintenance.

2.2 The Application of AR in Construction

The application of AR in the construction field has gradually gained widespread attention, especially in construction management, quality control and safety monitoring, etc. (Nguyen et al., 2021; Sadhu et al., 2023). Morikawa & Ando (2019) developed an augmented reality-based pipeline management system, which recognizes and operates AR tags through a tablet computer. Park et al. (2013) proposed an AR-based defect detection system by combining BIM, AR and building defect management. Mirshokraei et al. (2019) developed a 4D BIM model using BIM and AR technology to enhance the quality management process in construction using AR. In the field of safety control, El Ammari & Hammad (2019) proposed a collaborative BIM-based augmented reality framework for supporting on-site safety inspections and remote collaboration. Diao & Shih (2019) developed an AR-based building maintenance system that provides safety training and real-time construction path guidance to ensure safe operation of workers.

Despite the progress of AR in the construction field, there is a lack of research on the accurate integration of AR with BIM geometric and semantic information, real-time dynamic interaction in complex environments, and collaborative management of multi-terminal devices. Meanwhile, the application of BIM-AR is limited by the computational capability and positioning accuracy, and the development and implementation costs are high, which makes it difficult to popularize it especially in large-scale projects.

Therefore, this study proposes a BIM-AR driven collaborative management method for complex MEP equipment installation. The method optimizes the management of MEP equipment by real-time data interaction and collaborative application of multi-terminal equipment through the accurate integration of AR and BIM model, and is applied in a large public building project, which fills the gaps in the existing research and has important practical application value.

3. METHODOLOGY

The framework diagram of the proposed method in this study is shown in Figure 1. It mainly includes the following four parts: (a) acquisition of building scene; (b) matching geometry and semantic information of model; (c) human-machine interaction application of complex MEP equipment; and (d) collaborative management of multi-terminal devices. In this section, the methodology of each part is described.

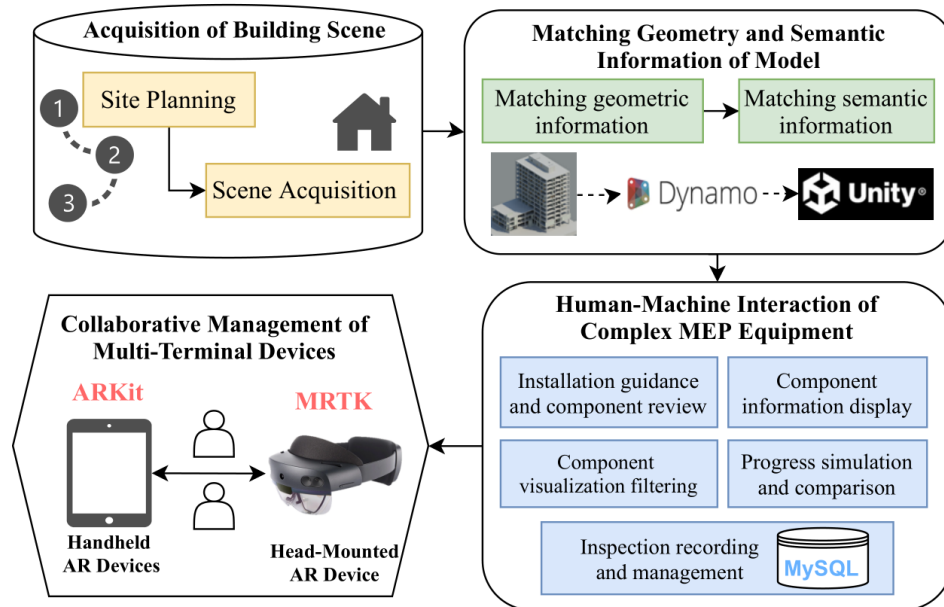


Figure 1. Framework diagram of the proposed method

3.1 Acquisition of Building Scene

The acquisition of the building scene is a prerequisite to realize the AR effect. In this study, a Matterport Pro 2 panoramic camera is used to acquire the 3D scene inside the building and generate high-resolution 3D point cloud data and texture maps.

In order to ensure comprehensive coverage and high-precision acquisition of scene data, site planning for scenes is crucial. The spacing between sites is adjusted according to the spatial complexity and is usually set between 1.5 and 2.5 meters. At the same time, key locations such as windows and holes in the building are marked immediately after each acquisition operation to improve the matching effect.

After planning the sites, the Matterport Pro 2 panoramic camera performs a 360-degree omni-directional scanning at each acquisition site, combining with the depth sensor to capture spatial information of the scene and generating accurate point cloud data to represent the geometry of the building. At the same time, the high-resolution images captured by the camera are used to generate texture maps. The acquired sites are schematically shown in Fig. 2(a).

After the acquisition is completed, the system combines the texture mapping of each site with the point cloud data through a specialized processing algorithm to generate a 3D scene model with rich details and high-quality visual effects, which provides a reliable reference for the geometrical alignment of the BIM model, as shown in Figure 2(b).

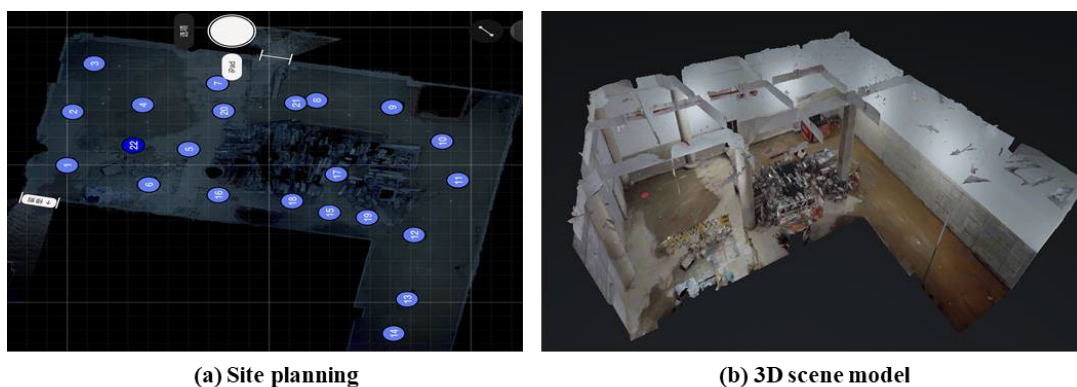


Figure 2. Acquisition of building scenes

3.2 Matching Geometric and Semantic Information of Model

(1) Matching Geometric Information of Model

In the process of geometric alignment of MEP models, this study adopts a markerless-based matching method to accurately align the existing MEP models linking architectural and structural models with the captured

texture maps of the scene, and the results are shown in Figure 3.

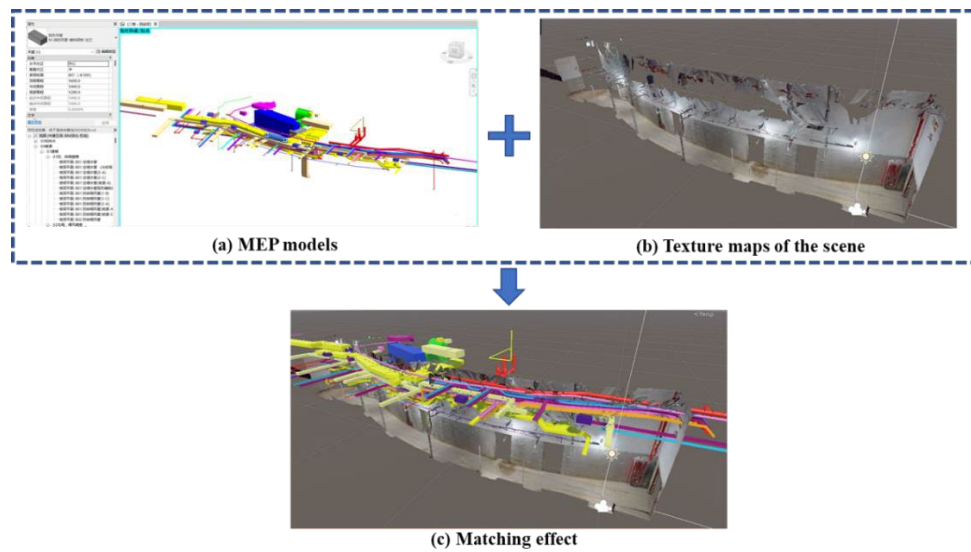


Figure 3. Markerless-based matching method for geometric information

The key to the matching process is to select a number of obvious architectural reference points, such as columns, wall corners and other geometric feature points. Based on the coordinate relationships of these reference points, the matching algorithm calculates the transformation matrix and automatically adjusts the position, angle and scale of the MEP model to ensure that it is completely consistent with the geometry and spatial location of the actual built environment. The advantage of this automatic markerless-based matching method is its flexibility and accuracy, which is more efficient than the traditional manual adjustment and can adapt to the complex geometry of different building environments.

In addition, this study is based on Vuforia's Area Target localization technology, which generates feature maps by capturing the salient geometry of the target area. When the AR device scans the physical environment in real time, the system automatically compares the environmental features captured by the camera with the pre-generated feature maps, thus realizing the precise alignment between the virtual model and the actual scene.

In this way, the MEP model can be accurately superimposed in the augmented reality environment, which provides a reliable location basis for construction guidance, on-site inspections, and so on.

(2) Matching Semantic Information of Model

The matching of semantic information enables the virtual model not only to have physical geometry but also to display detailed component attributes in an augmented reality environment. These attributes include key information such as the ID, category, construction period, and floor of the MEP component.

The semantic information of the components is exported from the BIM model by writing a program in Dynamo, as shown in Figure 4. The exported data forms a semantic information table corresponding to the geometric model. This information table provides rich data support for subsequent matching and interaction operations.

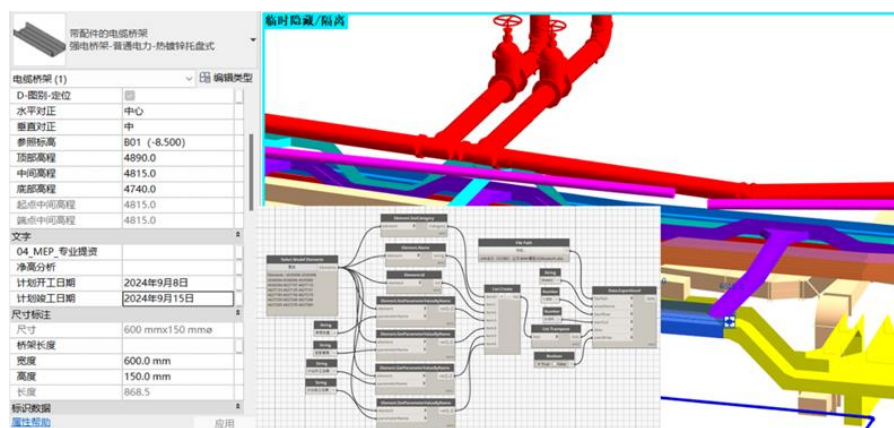


Figure 4. Component information export by Dynamo

Next, the exported semantic information needs to be matched with the model artifacts in the Unity environment. After the BIM model is imported into Unity, each component's name contains a unique ID that corresponds to the ID exported by Dynamo. With this unique identifier, the semantic information can be accurately associated with the geometry model in Unity, realizing the fusion of geometric and semantic data. An example of associating semantic information in Unity is shown in Figure 5.

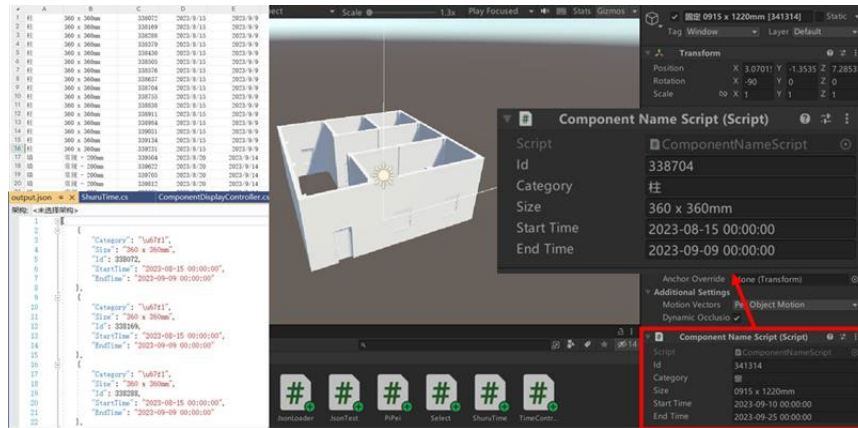


Figure 5. Matching semantic information in Unity

3.3 Human-machine Interaction Application of Complex MEP Equipment

Based on the acquisition and matching of MEP component data, this paper proposes a set of AR-based human-machine interaction application solutions. By combining the components in the MEP model with the actual scene, this solution realizes efficient component installation guidance, information display, progress management and inspection records in the AR environment, and its effect is shown in Figure 6.

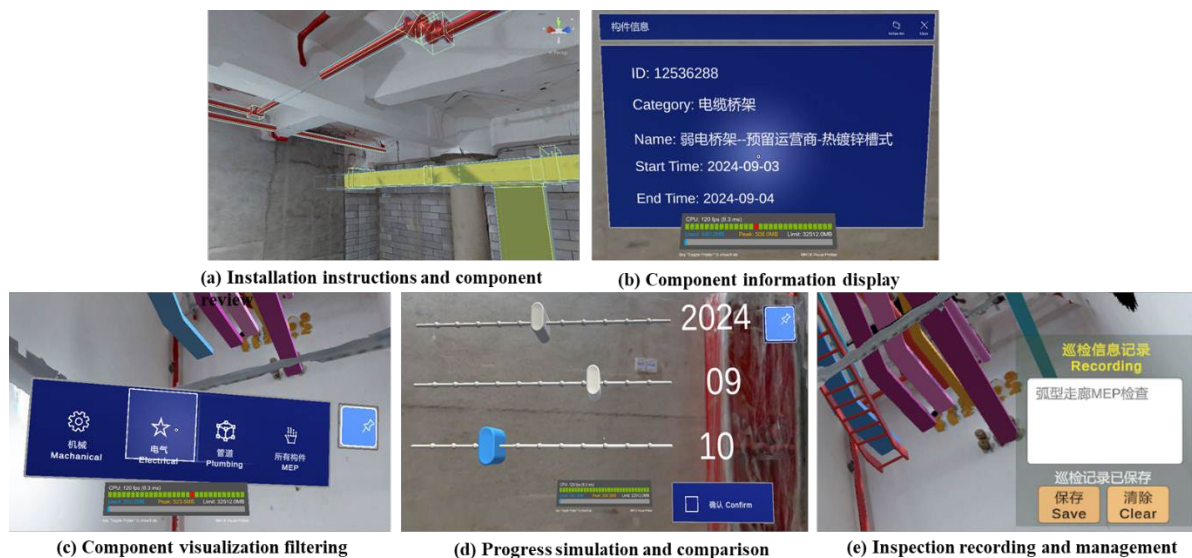


Figure 6. Human-machine interaction application of complex MEP equipment

(1) Installation Guidance and Component Review

During the installation of complex MEP equipment, the accurate AR positioning effect setting can visualize the installation position of each component and provide clear guidance to the construction personnel, such as the location of holes and the determination of the installation sequence. In addition, AR technology also supports the review of hidden components, the effect of which is shown in Figure 6(a). The feature helps constructors quickly understand the arrangement of complex equipment, reduces installation errors, and improves installation efficiency.

(2) Component Information Display

In the AR scene, the detailed attribute information of the components can be displayed in real time by

interacting with the model, the effect of which is shown in Figure 6(b). Through the development of collision bodies, pointers and rays, construction personnel can select any MEP component by simple gesture operation and display its ID, category, duration and other attributes, so that construction personnel can quickly determine the specific installation requirements and status of the component.

(3) Component Visualization Filtering

In order to help constructors quickly locate and view specific types of components, the application performs visual filtering based on the category, floor, room and other attributes of the MEP components. Users can filter out specific categories of MEP components through simple menu operations, which can reduce visual interference and quickly locate the required components, the effect of which is shown in Figure 6(c).

(4) Progress Simulation and Comparison

The progress simulation function allows the user to set the construction date in the AR scene. As shown in Figure 6(d), when the user sets a certain date, the AR scene will display all the components that should be completed before that date. Users can visually check whether the components are completed on schedule, which helps managers adjust the construction plan in time and provides a powerful support for the construction progress management of MEP equipment.

(5) Inspection Recording and Management

On the construction site, the recording and management of inspection information is an important part of ensuring the quality of construction. The AR system in this study manages all the inspection records through MySQL database and supports the synchronized operation of multiple terminal devices. As shown in Figure 6(e), the construction personnel can record the inspection results after completing the component inspection in the AR scene, and the system will automatically upload this information to the MySQL database for subsequent query and management, ensuring that the problems in the construction process will be fed back and solved in a timely manner, which improves the efficiency of construction management.

3.4 Collaborative Management of Multi-Terminal Devices

This study introduces AR headset and hand-held devices into on-site construction management, realizing the collaborative management of multi-terminal devices. Through the database, the inspection information is unified and synchronized in real time among multiple devices, and the managers and construction personnel can use different devices to view and update the equipment information in real time, which improves the communication and collaboration efficiency at the project site.

(1) Deployment of AR Headset

HoloLens 2 provides an immersive 3D visualization experience. Combined with Microsoft's Mixed Reality Toolkit (MRTK), an application was developed that allows users to view component properties, check construction progress, and enter inspection information into the system for management through gesture presses, hand rays, or voice commands. Through multiple rounds of deployment testing, its UI design was optimized to enable stable operation in complex field environments and seamless collaboration with other devices.

(2) Deployment of AR Handheld Devices

Handheld devices such as iPad are widely used in project sites due to their portability and ease of operation. iPad provides real-time access to project data and recording of inspection results, and the developed UI interface focuses on touch-screen interaction and supports zooming in and out of the 3D model, rotating the viewing angle and component filtering. Users can filter MEP components according to floor, category and other conditions, and inspection data is automatically uploaded and synchronized to the database. Deployment tests of AR kit show that it performs stably in scenes with poor signal or complicated construction, especially suitable for frequently moving occasions.

4. EXPERIMENTS

In order to verify the feasibility and application effect of the proposed method in this study, experiments were conducted in the basement MEP equipment scenario of a public building project under construction in Shenzhen. The experiments mainly focus on the matching accuracy of geometric and semantic information of MEP models, the application effect of human-machine interaction of complex electromechanical equipment, and the collaborative application of multiple terminals to validate its practical value for MEP equipment installation and management.

4.1 Experimental Setting

Two complex areas in the negative basement floor were selected for the experimental scenario: a curved corridor and a large room, as shown in Figure 7. The corridor area contains complex piping and ventilation equipment, while the room involves the installation and commissioning of multiple MEP equipment.

In order to ensure accurate data acquisition and precise model alignment, the experiments were conducted using a Matterport Pro 2 panoramic camera for on-site 3D scanning, Revit 2020 for processing the MEP model, Dynamo for developing a program for exporting the MEP component data, and Unity 2020 and 2021 for the development of UWP and IOS versions of the AR apps, respectively. HoloLens 2 and iPad Pro were used for AR human-machine interaction in the field.

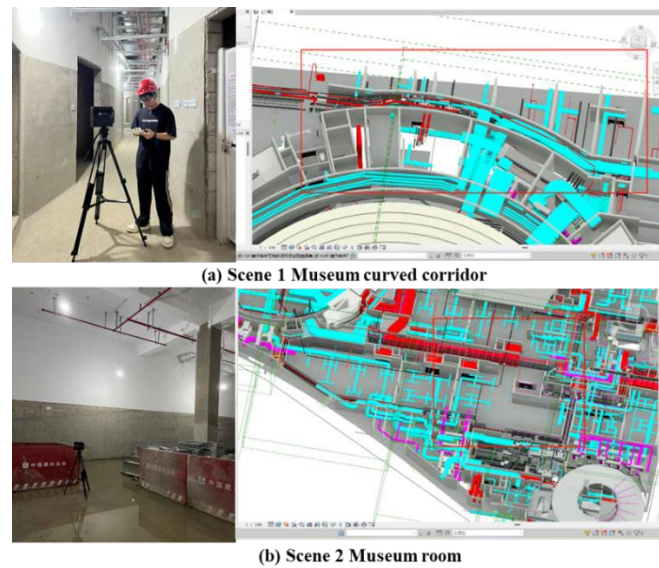


Figure 7. Experimental scenes and acquisition

4.2 Experimental Implementation and Results

(1) Matching of Model Information

The experiment captured 3D texture maps of two scenes by Matterport Pro 2 and accurately aligned the scenes with the BIM models in Unity. The columns and corners in the building were selected as reference points during the matching process to ensure the precise alignment of the MEP model with the actual environment. At the same time, the semantic data of the MEP components exported by Dynamo are bound in Unity to realize the integration of geometric and semantic information of the virtual model. As shown in Figure 8, the geometric positioning error of the model reaches centimeter level, and the semantic data is matched accurately, which can display the detailed attribute information of the components in the AR scene in real time.



Figure 8. Matching effect of model information

(2) Human-Machine Interaction Applications on AR Headset

On the HoloLens 2 device, the experimenter was able to visualize the 3D model of the MEP equipment through the AR environment. By interacting with the virtual model through gestures and voice commands, the

constructor was able to obtain component information, check the installation location, and review the status of hidden components in real time, the effect of which is shown in Figure 9. The experimental results show that the HoloLens device can accurately display the installation position of the MEP equipment, the superposition of the AR model and the actual space is good, the program can run at a frame rate of up to 60 FPS, with an average frame rate of 46 FPS, and the operation is smooth, which significantly improves the visualization effect of the construction of complex MEP equipment.

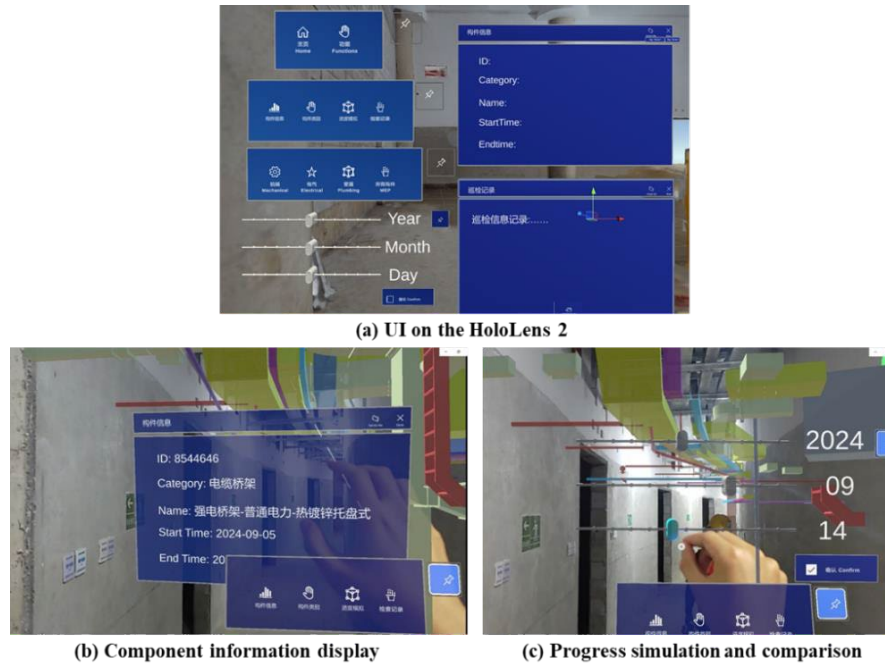


Figure 9. Human-machine interaction applications on HoloLens 2

(3) Human Machine Interaction Applications on AR Handheld Devices

On the iPad Pro, the experimenter was able to view and query the information of MEP components through touch screen operation. As shown in Figure 10, the intuitive interactive interface allows users to quickly click on the components to view their attributes and use the filtering function to categorize and display the components according to conditions such as floor and category. The experimental results show that iPad Pro has fast response time for data loading and operation, stable enhanced localization effect, and high portability, which is suitable for quick inspection and information query in mobile scenes, and has more flexibility compared with HoloLens 2.

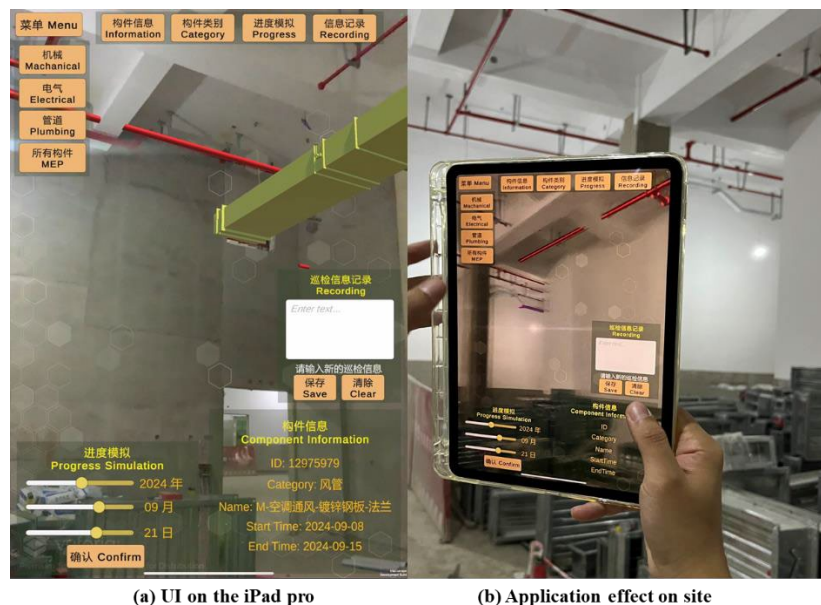


Figure 10. Human-machine interaction applications on iPad Pro

(4) Collaborative Management of Multi-Terminal Devices

In the multi-terminal collaborative application experiment, the collaborative application of devices such as HoloLens 2 and iPad Pro can better communicate about the situation of complex MEP equipment, and at the same time synchronize the inspection data through the MySQL database, and the experimental effect is shown in Figure 11. The experiment shows that the multi-terminal collaborative application significantly improves the information sharing and communication efficiency at the project site, and the construction personnel and management personnel are able to provide real-time feedback on the problems through different devices and ensure the unified management of inspection records.



Figure 11. Collaborative Management of Multi-Terminal Devices

4.3 Experimental Summary

The method proposed in this study is effectively verified through experiments in the MEP equipment scenario in the basement of a public building project under construction in Shenzhen. The experimental results show that the method can achieve centimeter-level accuracy in the localization of model geometric information, and at the same time, the semantic information of components can be accurately matched. The enhanced visualization effect of multi-terminal AR devices is obvious, in which the HoloLens 2 program can run at a frame rate of up to 60 FPS in the field, with an average frame rate of 46 FPS, and the collaborative application improves the efficiency of communication and feedback in the field. Overall, this method can smoothly realize the interaction of complex MEP equipment, significantly improve the efficiency of equipment installation and collaborative management, and has high practical application value.

5. CONCLUSIONS

This study proposes a BIM-AR-driven collaborative management method for complex MEP equipment, which significantly improves the efficiency of MEP equipment installation management through the acquisition of architectural scenes, the high-precision matching of model geometry and semantic information, and the collaborative application of multi-terminal devices. The introduction of AR technology enables construction personnel to interact with the virtual model in a real environment, realizing installation guidance, component information display, progress simulation and other human-machine interaction functions, which greatly improves the intuitiveness and accuracy of on-site operation. Through the experimental verification in a public building project under construction in Shenzhen, this method shows good feasibility and application prospects in real projects.

However, this study still has some shortcomings in practical application. First, although AR significantly improves the interactive experience, the stability of the equipment and the real-time data transmission still need to be further optimized in the complex construction environment. Second, the synchronization of BIM data across multiple devices needs to be further optimized to ensure seamless communication and real-time updates. Future work will focus on improving the scalability of the system in large-scale projects, optimizing the data transmission and multi-terminal collaboration mechanism, and exploring advanced features such as automated management and machine learning-based detection to further enhance the intelligence of MEP equipment management.

ACKNOWLEDGMENTS

This study was supported by the National Natural Science Foundation of China [Grant Nos. 52308319].

REFERENCES

- Al-Sabbag, Z. A., Yeum, C. M., & Narasimhan, S. J. A. E. I. (2022). Interactive defect quantification through extended reality, *Advanced Engineering Informatics*, 51, 101473
- Catbas, F. N., Luleci, F., Zakaria, M., Bagci, U., LaViola Jr, J. J., Cruz-Neira, C., & Reiners, D. J. S. (2022). Extended reality (XR) for condition assessment of civil engineering structures: A literature review, *Sensors*, 2223, 9560
- Cheng, J. C., Chen, W., Chen, K., & Wang, Q. J. A. i. C. (2020). Data-driven predictive maintenance planning framework for MEP components based on BIM and IoT using machine learning algorithms, *Automation in Construction*, 112, 103087
- Chu, M., Matthews, J., & Love, P. E. J. A. i. C. (2018). Integrating mobile building information modelling and augmented reality systems: an experimental study, *Automation in Construction*, 85, 305-316
- Diao, P.-H., & Shih, N.-J. J. A. s. (2019). BIM-based AR maintenance system (BARMS) as an intelligent instruction platform for complex plumbing facilities, *Applied Sciences*, 98, 1592
- El Ammari, K., & Hammad, A. J. A. i. C. (2019). Remote interactive collaboration in facilities management using BIM-based mixed reality, *Automation in Construction*, 107, 102940
- Kalasapudi, V. S., Turkan, Y., & Tang, P. (Year). Toward automated spatial change analysis of MEP components using 3D point clouds and as-designed BIM models, *2014 2nd International Conference on 3D Vision*, 2, 145-152
- Li, X., Yang, D., Yuan, J., Donkers, A., & Liu, X. J. A. i. c. (2022). BIM-enabled semantic web for automated safety checks in subway construction, *Automation in Construction*, 141, 104454
- Mirshokraei, M., De Gaetani, C. I., & Migliaccio, F. J. A. S. (2019). A web-based BIM-AR quality management system for structural elements, *Applied Sciences*, 919, 3984
- Morikawa, K., & Ando, T. J. F. S. T. J. (2019). Reduction of piping management person-hours through use of AR technology at shipbuilding sites, *Fujitsu Sci. Tech*, 5501, 20-26
- Nguyen, D.-C., Jin, R., Jeon, C.-H., & Shim, C.-S. J. C. I. (2021). BIM-based mixed-reality application for bridge inspection and maintenance, *Construction Innovation*, 223, 487-503
- Park, C.-S., Lee, D.-Y., Kwon, O.-S., & Wang, X. J. A. i. c. (2013). A framework for proactive construction defect management using BIM, augmented reality and ontology-based data collection template, *Automation in Construction*, 33, 61-71
- Pärn, E., Edwards, D., & Sing, M. C. J. A. i. C. (2018). Origins and probabilities of MEP and structural design clashes within a federated BIM model, *Automation in Construction*, 85, 209-219
- Sadhu, A., Peplinski, J. E., Mohammadkhorasani, A., & Moreu, F. J. J. o. S. E. (2023). A review of data management and visualization techniques for structural health monitoring using BIM and virtual or augmented reality, *Journal of Structural Engineering*, 1491, 03122006
- Vignali, V., Acerra, E. M., Lantieri, C., Di Vincenzo, F., Piacentini, G., & Pancaldi, S. J. A. i. C. (2021). Building information Modelling (BIM) application for an existing road infrastructure, *Automation in Construction*, 128, 103752
- Wang, J., Wang, X., Shou, W., Chong, H.-Y., & Guo, J. J. A. i. c. (2016). Building information modeling-based integration of MEP layout designs and constructability, *Automation in Construction*, 61, 134-146
- Wang, Y., Zhang, L., Yu, H., & Tiong, R. L. J. A. E. I. (2022). Detecting logical relationships in mechanical, electrical, and plumbing (MEP) systems with BIM using graph matching, *Advanced Engineering Informatics*, 54, 101770
- Xie, X., Zhou, J., Fu, X., Zhang, R., Zhu, H., & Bao, Q. J. B. (2022). Automated rule checking for MEP systems based on BIM and KBMS, *Buildings*, 127, 934
- Zhao, Y., Deng, X., & Lai, H. J. A. i. C. (2021). Reconstructing BIM from 2D structural drawings for existing buildings, *Automation in Construction*, 128, 103750

Enhanced Curtain Wall Construction Progress Monitoring via Tower Crane Perspective and Integrated 3D Reconstruction

Tao Zhong¹, Yujie Lu², Shuo Wang³, Xuanjun Li⁴

1) Ph.D. Candidate, College of Civil Engineering, Tongji University, Shanghai, China. Email: 2410045@tongji.edu.cn

2) Ph.D., Prof., Department of Architecture, Tongji University, Shanghai, China. Email: lu6@tongji.edu.cn

3) Ph.D. Candidate, College of Civil Engineering, Tongji University, Shanghai, China. Email: wangshuo2207@tongji.edu.cn

4) Postgraduate, College of Civil Engineering, Tongji University, Shanghai, China. Email: 2430928@tongji.edu.cn

Abstract: Urban development has driven the widespread proliferation of high-rise buildings, making curtain wall construction a critical aspect of progress management. However, the extensive surface areas of curtain walls and the severe perspective distortions present significant challenges to efficient construction progress monitoring. To address these challenges, a novel method is proposed for tracking curtain wall installation progress in high-rise buildings through integrated 3D reconstruction. Cameras are strategically mounted at the ends of adjacent tower crane booms, with multiple ground control points (GCPs) deployed on-site as hardware anchors. The process begins with capturing multi-view images using the crane-mounted cameras. The COLMAP 3D reconstruction pipeline is enhanced by incorporating GCPs to ensure accurate 3D reconstruction within a real-world coordinate system. The building structure is subsequently extracted from the site model, and rectified facade images are generated using projection techniques. The curtain wall installation progress is assessed at both the floor and overall building levels using the YOLOv8 image segmentation model. The proposed method was validated through a case study on a super high-rise construction site. This approach achieved decimeter-level accuracy in 3D reconstruction and a precision rate of 95.9% in curtain wall progress identification, meeting project requirements. These findings establish a robust framework for managing large-scale outdoor construction progress, particularly for high-rise curtain walls. Additionally, the site modeling methodology enables more refined and timely monitoring practices, offering significant potential for the development of digital twin models.

Keywords: Curtain wall, Progress identification, Digital twin, Three-dimensional modeling, Computer vision

1. INTRODUCTION

As urbanization accelerates and urban land becomes increasingly scarce, high-rise buildings have become a crucial architectural solution, widely adopted for offices, commercial spaces, and other purposes (Quesada-Olmo et al., 2018). Effective schedule management in high-rise construction plays a pivotal role in controlling project costs (Baduge et al., 2022). Meanwhile, curtain wall projects, known for their intensive workloads, high installation risks, and low levels of automation, are critical elements that directly impact key project milestones (Kim et al., 2021). Therefore, monitoring and managing the progress of curtain wall installation is essential throughout the construction process of high-rise buildings.

The objectives of monitoring facade installation progress include assessing the progress of each facade and identifying uninstalled areas (Eom & Kang, 2022). Manual inspections face difficulties in accurately quantifying progress and are labor-intensive. Recent studies have explored the use of algorithms such as computer vision (Lu et al., 2023) and 3D reconstruction (Lei et al., 2019) to automate the identification of construction elements for progress calculation. Computer vision techniques analyze two-dimensional visual features of construction elements to determine progress based on the proportion of object pixels within the image. In contrast,

3D reconstruction methods create 3D point clouds from images or laser scans, followed by semantic annotation through BIM-based comparison to calculate construction volume. However, these approaches pose significant challenges. Visual recognition methods that rely solely on images, captured manually or via drones, are often limited by shape distortions and the lack of accurate size information, making it difficult to obtain precise progress data. Furthermore, due to the reflective and transparent properties of glass, the resulting point clouds of curtain walls are typically too sparse to form a complete 3D representation, leading to the loss of key visual features (Lin & Golparvar-Fard, 2020).

Additionally, existing methods typically rely on either manual inspections or drone flights to monitor construction progress and quality. However, manual inspections are inefficient for large construction sites, while the use of drones is constrained by the need for professional operators, favorable weather conditions, and compliance with licensing requirements. As a result, these approaches fail to support efficient daily site monitoring. In contrast, crane-mounted cameras provide an effective alternative by autonomously capturing construction data during routine operations without requiring manual intervention or permits (Masood et al., 2020). Positioned on the crane's boom, these cameras capture multi-view imagery of the construction site, which is subsequently utilized for 3D reconstruction of site point clouds. As autonomous hardware systems, they operate continuously without human oversight, enabling real-time monitoring of curtain wall progress and overall site activity.

This paper proposes a method for monitoring curtain wall installation progress in high-rise buildings based on tower crane perspectives combined with 3D reconstruction (see Figure 1). Our approach uses tower crane cameras as the primary hardware, leveraging 3D reconstruction for accurate dimensional analysis and computer vision techniques for visual progress tracking.

The structure of this paper is organized as follows: The first section presents the importance of monitoring curtain wall progress in high-rise buildings and the limitations of current research. The second section outlines the research methods, including the integration of ground control points (GCPs) for accurate 3D reconstruction, extraction of rectified facade images from the site model, and curtain wall identification for progress analysis. In the third section, we present the experimental setup along with the results of site modeling and progress quantification. The fourth section discusses the benefits and practicality of the proposed method. Finally, the contributions and limitations of this work are summarized in the fifth section.

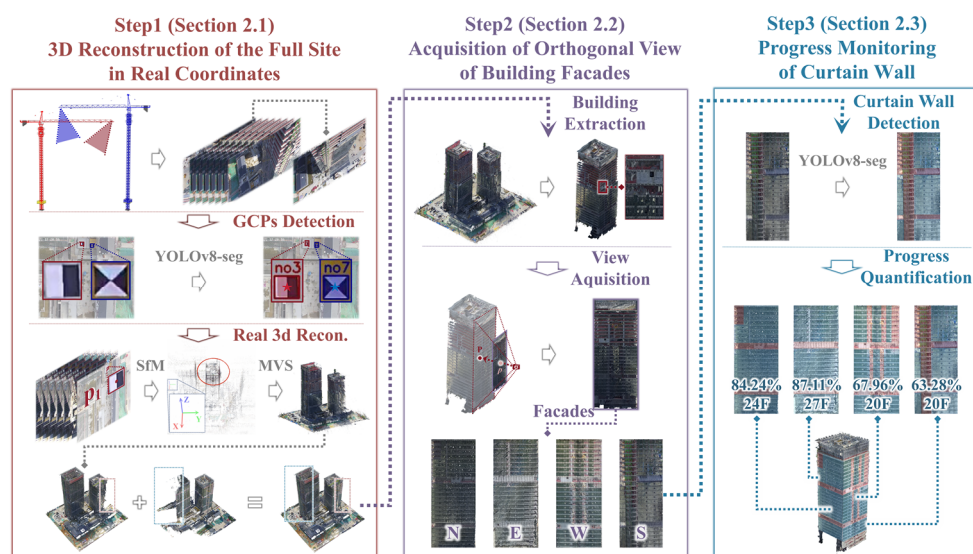


Figure 1. Curtain wall progress identification framework using multi-view 3D reconstruction

2. METHOD

2.1 3D Reconstruction of Construction Site with GCPs Integration

Cameras are mounted at the front ends of tower crane booms, enabling the capture of multi-view images from multiple tower cranes to generate a comprehensive site point cloud model. This model serves as a foundational dataset for extracting visual information related to the curtain wall. Ground Control Points (GCPs) are integrated into the process by identifying their 2D coordinates and using the COLMAP 3D reconstruction algorithm to accurately determine their corresponding 3D coordinates.

(1) Detection of GCP Pixel Centers

GCPs with known real-world coordinates are utilized to establish a coordinate transformation between the image coordinate system and the real-world coordinate system. The process begins by obtaining the pixel coordinates of the GCPs in each multi-view image. These GCPs are fixed markers strategically positioned at various heights across the construction site, each featuring a clearly defined geometric center. To determine the distribution of GCPs within the images, the YOLOv8-seg model is employed. The centroid of the corresponding image mask is then calculated to identify the precise GCP center (see Figure 2).

The YOLOv8-seg model, a segmentation-specific variant of the YOLOv8 model, is optimized for image segmentation tasks. Its architecture integrates a multi-detection head and an anchor-free design, enabling it to effectively detect target objects of varying sizes. The model consists of three main components: the backbone, neck, and head. The backbone extracts multi-scale features, the neck aggregates and bridges features across scales, and the head predicts object categories and masks using its multiple detection heads. Finally, by leveraging the symmetrical geometric shape of the GCPs, the geometric centroid of each object mask is computed to determine the precise pixel coordinates of the GCPs.

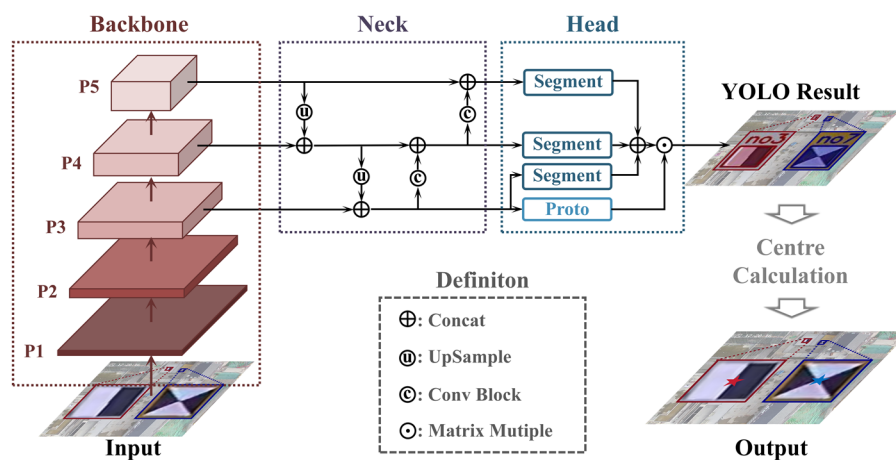


Figure 2. Process for identifying GCP pixel centers using the YOLOv8-seg model

(2) Framework for COLMAP-based 3D Reconstruction with GCPs Integration

COLMAP (Schonberger & Frahm, 2016) is a well-established open-source framework for 3D reconstruction from multi-view images and has demonstrated strong performance across various outdoor scenarios. However, since the modeling process relies on relative pose estimation between images, it lacks absolute coordinate information, including real-world scale and orientation. To address this limitation, we propose a COLMAP-based 3D reconstruction framework with GCP integration (see Figure 3) to incorporate real-world coordinates and recover precise 3D information. The original COLMAP framework consists of two key components: sparse reconstruction (SfM) and dense reconstruction (MVS). Sparse reconstruction estimates camera poses through feature matching, triangulation, and bundle adjustment, while dense reconstruction

generates the point cloud model via depth estimation and point cloud fusion. We integrate GCP information during the sparse reconstruction phase to recover accurate poses by triangulating corresponding GCPs to obtain normalized 3D coordinates P_n (as shown in Equation (1)). The transformation T between the normalized coordinates P_n and real-world coordinates P_r is estimated (see Equation (2)) and mapped onto the camera pose E_i (see Equation (3)) to recover the real-world coordinates of both the cameras and the point cloud model. With each local point cloud model reconstructed from individual tower crane cameras aligned within the real-world coordinate system, a complete site model is formed through the direct aggregation of these local models, as shown in Table 1.

$$P_n = \tau(p_a, p_b, E_a, E_b, I_a, I_b) \quad (1)$$

where, a and b are two images capturing the same GCP point P . The pixel coordinate of this point in image i is denoted as p_i , and E_i and I_i represent the extrinsic and intrinsic parameters of image i respectively. We use the DLT method (Robust, 2020) for triangulation τ .

$$T = P_r^* \cdot (P_n^*)^{-1} \quad (2)$$

where, P^* denotes the homogeneous coordinates of the 3D point P .

$$E_r = T \cdot E_i \quad (3)$$

where, E_r represents the camera pose in the real-world coordinate system.

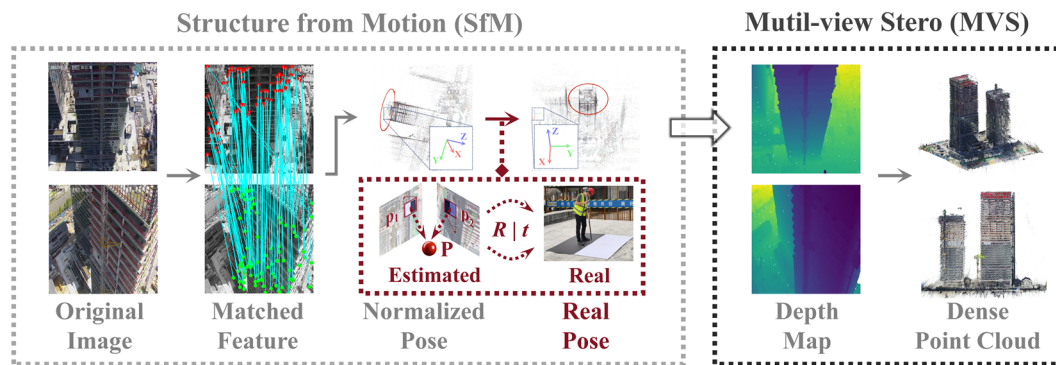


Figure 3. COLMAP 3D reconstruction framework with GCP integration, where red sections indicate the introduced module for recovering real-world poses

2.2 Acquisition of Rectified Facade Images

Building appearance images are captured using the site model to obtain geometric information about the curtain walls. These images must meet three essential criteria: unobstructed views, minimal distortion, and an accurate scale aligned with real-world dimensions. Adhering to these criteria ensures reliable identification of completed and incomplete sections of the curtain wall. To achieve this, the building structure is first extracted from the site model, and a corresponding appearance model is generated. Rectified facade images of the building are then created using projection principles, ensuring alignment with real-world geometry.

(1) Acquisition of the Building Visual Model

The building structure is extracted from the site model using point cloud projection, and a mesh model is generated using the Poisson surface reconstruction algorithm (Kazhdan et al., 2006) (see Figure 4). The mesh model, composed of numerous triangular patches, creates a continuous surface with realistic visual attributes, serving as the foundation for feature analysis. The process begins with horizontally projecting the site point cloud to calculate the horizontal point density distribution (see Figure 4(b)). Next, density gradients along the X and Y

axes are computed, where significant changes highlight object edges (see Figure 4(c)). Point clustering is then applied to distinguish individual objects, identifying the structure with the largest horizontal area as the building (see Figure 4(d)). Subsequently, the convex hull of the edge points is calculated, followed by vertical cropping to isolate the building structure's point cloud. Finally, the Poisson algorithm is applied to reconstruct the mesh model from the processed point cloud.

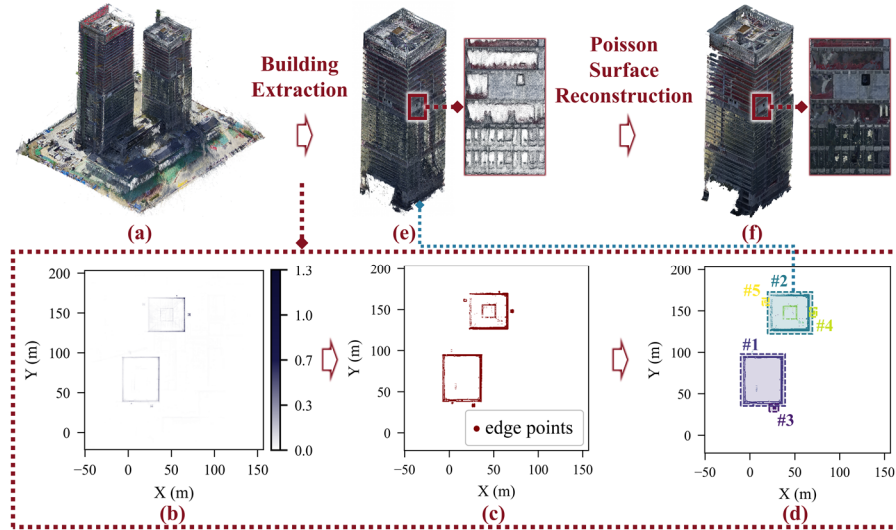


Figure 4. Workflow for generating the visual model of the building structure, where (a) represents the complete site point cloud, (e) represents the building structure point cloud, and (f) represents the mesh model of the building structure

(2) Acquisition of Building Facade Images

Virtual cameras are configured to capture rectified facade images of the visual model (see Figure 5). As described in Section 2.2 (1), the height H and width B of each building facade are determined from the geometric dimensions of the point cloud. Each pixel value in the image is calculated as follows: the optical center of the camera and each pixel p form a spatial ray \vec{r}_w (see Equation (4)), which intersects the mesh model of the building at point P . The RGB value at the intersection point P is assigned as the pixel value of p . For the external camera parameters R and t , the viewing direction is aligned with the normal of the facade. The optical center is positioned at half the height $H/2$, and the distance D is set to the width of the facade. For the internal camera parameters I , the aspect ratio of the image is matched to the height-to-width ratio of the building facade, and the focal length f is set to the pixel width I_u of the image. According to the pinhole camera model, the camera is positioned to capture the entire facade (see Equation (5)). This process produces rectified facade images for each building structure, which serve as the data foundation for subsequent curtain wall progress quantification.

$$\vec{r}_w = R \cdot \vec{r} + t, \quad \vec{r} = \frac{\begin{bmatrix} (i-c)/f & (j-c)/f & 1 \end{bmatrix}^T}{\sqrt{(i-c)^2/f^2 + (j-c)^2/f^2 + 1}} \quad (4)$$

where, i and j denote the pixel's 2D coordinates; R and t represent the camera's external parameters; f and c denote the focal length and optical center, respectively.

$$U = 2 \cdot \frac{I_u \cdot D}{2f} = 2 \cdot \frac{I_u \cdot B}{2I_u} = B \quad (5)$$

where, U is the actual width in the camera's field of view, I_u denotes the pixel width of the image, D represents the

distance between the optical center and the object, and f denotes the focal length of the camera.

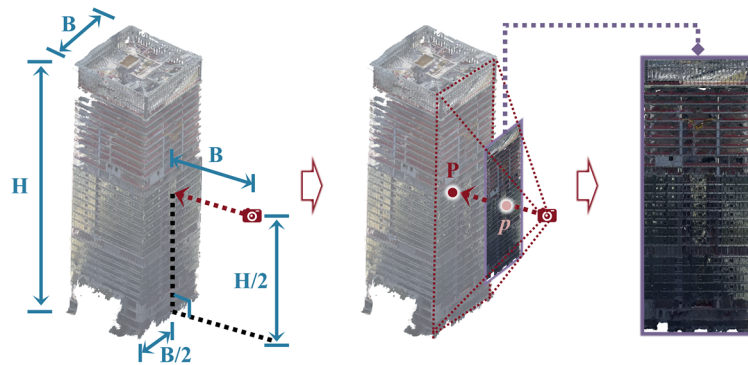


Figure 5. Workflow of rectified facade image generation

2.3 Construction Progress Analysis of Curtain Wall

(1) Curtain Wall Distribution Identification

The YOLOv8-seg model is employed to identify the distribution of installed curtain walls on the building facade, enabling precise analysis of construction progress. By leveraging the distinct visual differences between curtain walls and structural elements, the image segmentation algorithm effectively delineates their boundaries. The recognition results are further refined using geometric priors of the facade. First, since the actual boundaries between curtain walls and structural elements are either horizontal or vertical, we straighten the detected edges to align with these orientations (see Figure 6(c)). Second, since only curtain walls and structural elements are present in the facade image and they do not intersect, the mean value of their adjacent boundaries is calculated and used as the actual dividing line (see Figure 6(d)).

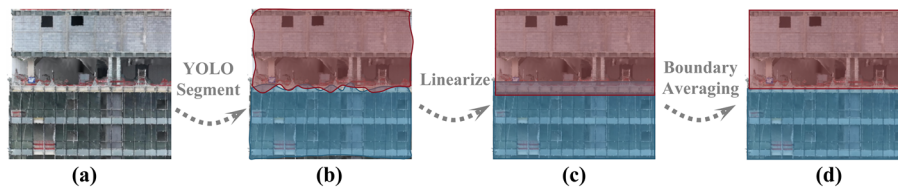


Figure 6. Post-processing diagram of curtain wall recognition results

(2) Curtain Wall Construction Progress Calculation

The overall progress, floor-specific progress, and the identification of the standard floor under construction are calculated as key indicators of curtain wall installation progress. Since the image-to-real-world scale is established during the generation of rectified facade images, pixel positions in the image can be accurately mapped to their real-world coordinates (see Equation (6)). Consequently, the geometric data in the image reflects real-world dimensions with precision. The overall curtain wall progress is defined as the ratio of the total curtain wall area to the facade area. Using the elevations of standard floors, boundary lines for each floor are determined within the image. The progress for each floor is calculated as the ratio of the curtain wall area on that floor to the total area of the floor. Finally, the boundary where significant progress differences are observed from top to bottom is identified as the standard floor currently under construction.

$$[X, Y]^T = [i, j]^T \cdot I_u / U \quad (6)$$

where, X and Y are the real-world facade coordinates, and i and j are the pixel coordinates.

3. RESULTS

3.1 Experiment Site Overview

Experiments were conducted at a super high-rise construction site in China, with data collected weekly over four months. During this time, primary construction activities included curtain wall installation, secondary structure installation, and interior decoration. The site consisted of two super high-rise commercial buildings, standing 180 meters and 135 meters tall, respectively. Two attached tower cranes facilitated material transport and installation: Tower #1 was positioned on the north side of Building #A, while Tower #2 was located on the west side of Building #B (see Figure 7). Pan-tilt-zoom cameras were mounted at the front ends of both crane booms to capture a panoramic view of the entire site. Additionally, seven Ground Control Points (GCPs) were placed around the site to integrate real-world coordinates into the 3D data. A total of 600 images were collected for GCP identification, while 800 images were used for curtain wall recognition.

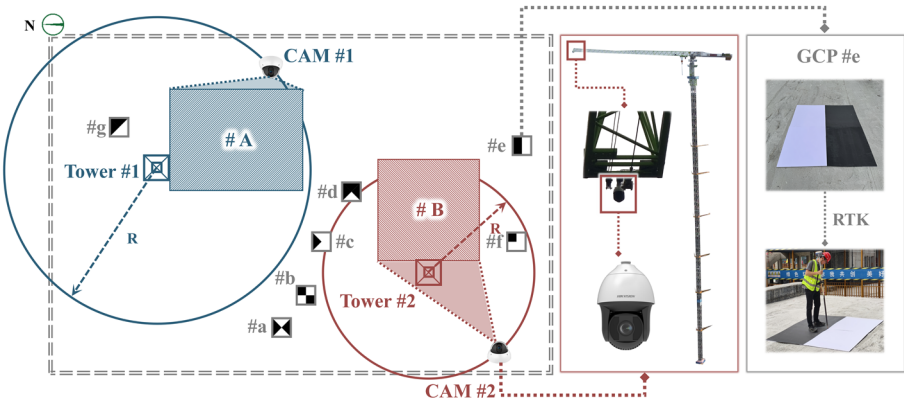


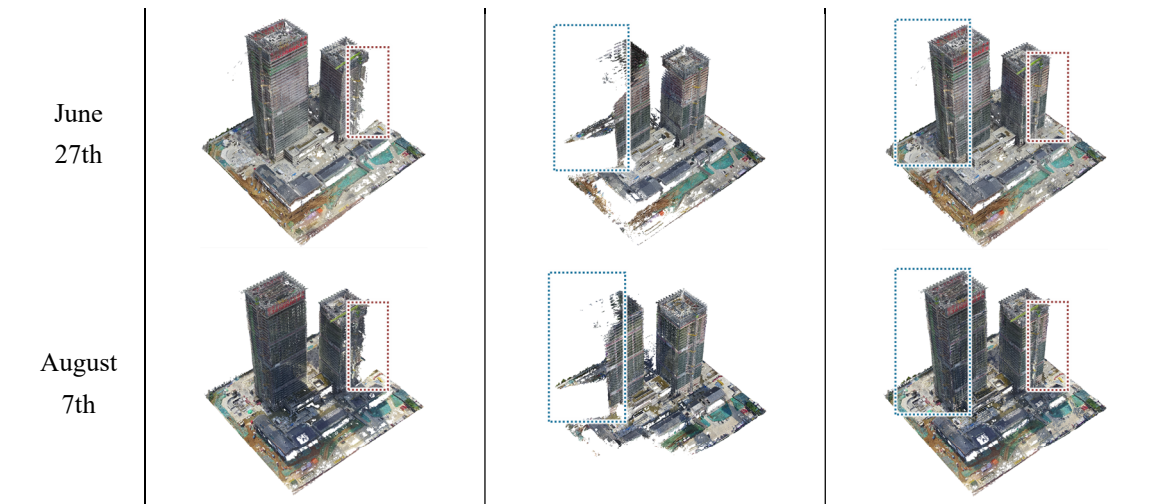
Figure 7. Experimental setup and hardware configuration

3.2 The Results of Integrated 3D Reconstruction

A comprehensive 3D reconstruction of the construction site was achieved through integrated modeling using dual tower crane cameras. When compared to the point cloud model reconstructed from drone flights, which served as the ground truth, the proposed model demonstrated an average error of 0.18 meters. This level of accuracy effectively reflects the actual conditions of the construction site, meeting the precision requirements for progress monitoring.

Table 1. Visual results of 3d joint modeling of the construction site

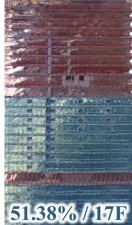
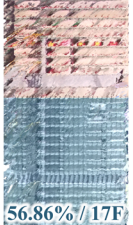

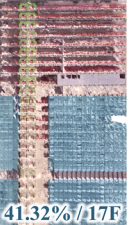
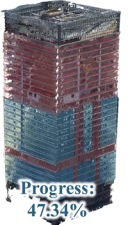
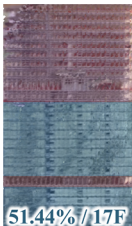

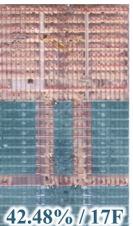

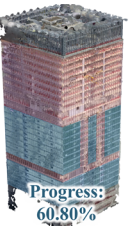
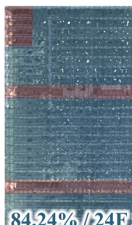
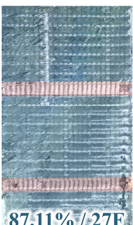
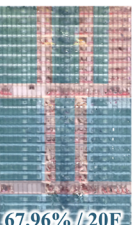
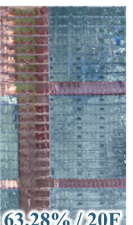

Date	From tower#1	From tower#2	Full site
April 18th			



3.3 The Results of Curtain Wall Progress Quantification

Three time points were selected to illustrate the quantified results of curtain wall progress recognition (see Table 2). As of August 7th, the overall curtain wall installation progress had reached 75.65%. The remaining uninstalled curtain wall areas were primarily located at equipment floors, construction hoists, and tower crane attachment points, which will be addressed after the removal of these temporary structures. It is important to note that minor deviations in progress recognition, typically ranging from 1% to 2%, may occur due to partial surface feature blurring in the mesh model or slight inaccuracies in image recognition. For instance, the east facade showed a 2% recognition error between June 27th and August 7th, despite no actual progress occurring during this period.

Table 2. Curtain wall construction progress visualization, where blue sections represent curtain walls, and red sections represent structural components

Date	North	East	South	West	All
April 18th					
June 27th					
August 7th					

4. DISCUSSION

4.1 Applicability of YOLOv8 in This Scenario

The YOLOv8-seg algorithm was employed for image recognition tasks, including the detection of GCPs and curtain walls. To assess its performance in detecting objects of varying sizes, YOLOv8 was benchmarked against two state-of-the-art algorithms, SOLOv2 and Mask2Former, using the mAP50 metric (see Table 3). The mAP50 metric evaluates detection accuracy based on a minimum 50% overlap between predicted and ground truth bounding regions. The results indicate that YOLOv8 achieves exceptional accuracy in detecting GCPs due to their distinct visual features, which enhances the precision of 3D modeling. For curtain walls, YOLOv8 attained a recognition accuracy of 95.9%, satisfying the precision requirements for construction progress monitoring. Furthermore, YOLOv8 outperformed SOLOv2 and Mask2Former, demonstrating superior accuracy and requiring fewer parameters, thus offering a more practical solution for large-scale construction monitoring.

Table 3. Comparison of segmentation accuracy of different algorithms

Algorithm	Params	GCPs	Curtain wall	Structural part
SOLOv2	46.26M	0.958	0.910	0.836
Mask2Former	44.06M	0.930	0.890	0.815
YOLOv8	27.24M	0.981	0.959	0.875

4.2 Outperformance of the Generated Rectified Images

Our contribution is the development of an efficient method for generating complete rectified facade images using tower crane cameras. Unlike traditional methods such as manual inspection, UAV photography, or standard crane camera captures, our approach produces rectified images that provide comprehensive facade coverage, remove distortions and obstructions, and accurately map pixels to real-world coordinates (as shown in Figure 8). This enables accurate progress recognition and precise localization of uninstalled curtain wall sections. Moreover, compared to methods that directly correct 2D images (Wei et al., 2023), our approach preserves a clear correspondence with real-world dimensions (see Equation (6)), offering more accurate scaling and a distinct advantage in floor-by-floor progress identification.

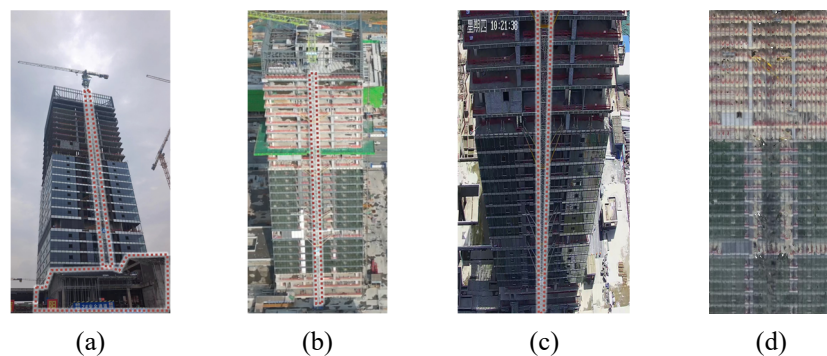


Figure 8. Comparison of the generated rectified image with common captures: (a) captured with a mobile phone, (b) captured by a UAV from a distance, (c) captured by a tower crane camera, and (d) our generated rectified image, the red dotted box marks the obstructed areas of the facade

5. CONCLUSION

Curtain wall progress recognition is a crucial aspect of curtain wall construction management. To address the challenges of inaccurate progress recognition and difficulty in localization for high-rise buildings, we propose a curtain wall progress recognition method based on joint 3D reconstruction from the aerial perspective of tower cranes. First, the real 3D scene is reconstructed using multi-view images captured by tower crane cameras. Next,

rectified facade images are extracted from the 3D model. Finally, progress quantification and analysis are performed on the rectified images, achieving a modeling accuracy of 0.18 meters and a curtain wall recognition accuracy of 95.9%. With the introduction of real 3D information, our method enables precise calculation of curtain wall progress at any location and accurate localization of uninstalled curtain wall sections. This accurate progress information allows managers to track project status, assess individual task efficiency, and optimize material deliveries and scheduling, facilitating precise on-site management.

The contribution of this paper lies in proposing a spatial appearance analysis method for large-scale objects on construction sites, addressing distortion, lack of scale, and incompleteness inherent in traditional 2D images. This method provides a reference for analyzing facades, foundation pits, and other large-scale outdoor construction information. It also demonstrates the practical application of digital twins on construction sites. However, our method requires regular building facades to obtain distortion-free surface images. In the future, surface unfolding techniques can be explored to fully analyze the visual features of irregular high-rise buildings.

ACKNOWLEDGMENTS

The study was supported by the National Key Research & Development Program of China (2022YFC3801700), Fundamental Research Funds for the Central Universities (2024-1-ZD-02).

REFERENCES

- Baduge, S. K., Thilakarathna, S., Perera, J. S., Arashpour, M., Sharafi, P., Teodosio, B., Shringi, A., & Mendis, P. (2022). Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications. *Automation in Construction*, 141, 104440.
- Eom, J., & Kang, Y. (2022). Curtain Wall Construction: Issues and Different Perspectives among Project Stakeholders. *Journal of Management in Engineering*, 38(5), 04022054.
- Kazhdan, M., Bolitho, M., & Hoppe, H. (2006). Poisson surface reconstruction. Proceedings of the fourth Eurographics symposium on Geometry processing,
- Kim, T., Kim, Y.-W., & Cho, H. (2021). A simulation-based dynamic scheduling model for curtain wall production considering construction planning reliability. *Journal of cleaner production*, 286, 124922.
- Lei, L., Zhou, Y., Luo, H., & Love, P. E. (2019). A CNN-based 3D patch registration approach for integrating sequential models in support of progress monitoring. *Advanced Engineering Informatics*, 41, 100923.
- Lin, J. J., & Golparvar-Fard, M. (2020). Visual and virtual progress monitoring in Construction 4.0. In *Construction 4.0* (pp. 240-263). Routledge.
- Lu, Y., Wei, W., Li, P., Zhong, T., Nong, Y., & Shi, X. (2023). A deep learning method for building façade parsing utilizing improved SOLOv2 instance segmentation. *Energy and Buildings*, 295, 113275.
- Masood, M. K., Aikala, A., Seppänen, O., & Singh, V. (2020). Multi-building extraction and alignment for as-built point clouds: a case study with crane cameras. *Frontiers in Built Environment*, 6, 581295.
- Quesada-Olmo, N., Jimenez-Martinez, M. J., & Farjas-Abadia, M. (2018). Real-time high-rise building monitoring system using global navigation satellite system technology. *Measurement*, 123, 115-124.
- Robust, A. (2020). Revisiting Projective Structure from Motion: A Robust and Efficient Incremental Solution. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, 42(2).
- Schonberger, J. L., & Frahm, J.-M. (2016). Structure-from-motion revisited. Proceedings of the IEEE conference on computer vision and pattern recognition,
- Wei, W., Lu, Y., Lin, Y., Bai, R., Zhang, Y., Wang, H., & Li, P. (2023). Augmenting progress monitoring in soil-foundation construction utilizing SOLOv2-based instance segmentation and visual BIM representation. *Automation in Construction*, 155, 105048.

3D INDOOR RECONSTRUCTION BASED ON SPATIAL LAYOUT ESTIMATION USING PANORAMIC INSPECTION VIDEO FROM BUILDING SITES

Shuo Wang¹, Yujie Lu², Tao Zhong³, Lijian Zhong⁴, and Yufan Chen⁵

- 1) Ph.D. Candidate, College of Civil Engineering, Tongji University, Shanghai, China. Email: wangshuo2207@tongji.edu.cn
- 2) Ph.D., Prof., College of Civil Engineering, Tongji University, and Key Laboratory of Performance Evolution and Control for Engineering Structures of Ministry of Education, Tongji University, Shanghai, China. Email: lu6@tongji.edu.cn
- 3) Ph.D. Candidate, College of Civil Engineering, Tongji University, Shanghai, China. Email: zhingzt@163.com
- 4) Master Student, College of Civil Engineering, Tongji University, Shanghai, China. Email: lijianz@tongji.edu.cn
- 5) Master Student, College of Civil Engineering, Tongji University, Shanghai, China. Email: 2330894@tongji.edu.cn

Abstract: As the demand for digital delivery of construction projects in the Architectural Engineering and Construction (AEC) industry continues to increase, the importance of worksite inspection and supervision is emphasized. Digital twin modeling of construction processes can reflect real-time site conditions, aiding refined management and project delivery. This paper explores the task of 3D layout reconstruction of interior construction sites through inspection by employing a portable 360-degree panoramic camera. The method uses visual simultaneous localization and Mapping (vSLAM) technology to precisely estimate camera poses during inspections, generating a motion trajectory and selecting key panoramic frames through an optimal capture point searching algorithm. Before reconstruction, the system integrates Inertial Measurement Unit (IMU) data to determine positional relationships between panoramic camera viewpoints, aligning multiple panoramic images into a unified coordinate system for accurate spatial reconstruction. Three-dimensional indoor layouts are reconstructed from panoramic images using a deep learning-based algorithm to automatically detect vertices through panoramic geometry calculations from a single panorama. An experiment with the existing floor plan is conducted to demonstrate the validity of the proposed method. This research introduces a novel approach that enhances the real-time capabilities and automation of spatial layout modeling for construction sites, laying the groundwork for intelligent inspections and holding significant engineering potential for rapid spatial layout recovery and object space mapping in future applications.

Keywords: Intelligent construction, Layout estimation, 3D reconstruction, Panoramic inspection, Simultaneous localization and mapping

1. INTRODUCTION

Digital delivery in civil and infrastructure projects is vital for ensuring project success, quality assurance, owner satisfaction, and long-term operations and maintenance (Guo et al. 2017). A fundamental concept of the digital delivery model is the transition from two-dimensional to three-dimensional management frameworks, enabling more comprehensive and dynamic oversight of project processes. The digital model of existing construction scenes generated through 3D reconstruction can visually demonstrate the differences between the construction site conditions and the design model, enabling project managers to assess conditions efficiently, make informed decisions, and conduct post-project verification and documentation, serving it as a critical technology in the development of an effective project delivery management system.

3D layout reconstruction extracts object profiles, including dimensions, positions, and range data, which is crucial for creating digital twins with detailed geometrical information and enabling human interaction and predictive analysis (Lu et al. 2024). Data for 3D layout reconstruction can be obtained using structured light sensors, laser scanners, X-ray scanners, or imaging devices (Vervikou and Ioannidis 2023). In recent years, digital cameras have become a preferred choice due to their quality, efficiency, and cost-effectiveness. Thus, image-based 3D layout reconstruction techniques have gained increased attention from both theoretical and practical perspectives in the field of construction engineering and management.

Current mainstream methods for 3D reconstruction of indoor layouts fall into two categories. The first method involves using multiple consecutive images with Structure from Motion (SfM) (Ullman 1979) or Simultaneous Localization and Mapping (SLAM) (Durrant-Whyte and Bailey 2006) techniques for depth estimation, resulting in a sparse point cloud model. However, in indoor environments with limited texture and features, image-based 3D reconstruction requires capturing a large number of images from various angles, leading to time-consuming data processing and insufficient model accuracy. The second method employs a single panoramic image to estimate depth or directly infer layout parameters for indoor reconstruction, which is more efficient in data acquisition and processing, providing higher accuracy specifically for layout reconstruction of interior space.

This study examines the digital delivery of 3D construction modeling through site inspection following the sequence of inspection trajectory mapping, capture point determination, and 3D layout reconstruction, as shown in Figure 1. Panoramic video serves as the data source to develop a 3D digital site layout information model. The method employs a portable 360-degree panoramic camera to capture indoor panoramic images and addresses two main challenges in the modeling process: (1) To determine the optimal panoramic capture points during

dynamic inspections, an OpenVSLAM-based path trajectory mapping and capture point retrieval algorithm is proposed; (2) To address the issue of missing the actual size of the room layout model, the HorizonNet algorithm is applied for single panoramic image-based 3D layout reconstruction of interior spaces, enabling the recovery of accurate room boundary dimensions. The following sections will focus on related work, proposed methods, experimental analysis, and conclusions regarding 3D layout reconstruction of interior construction based on panoramic site inspection.

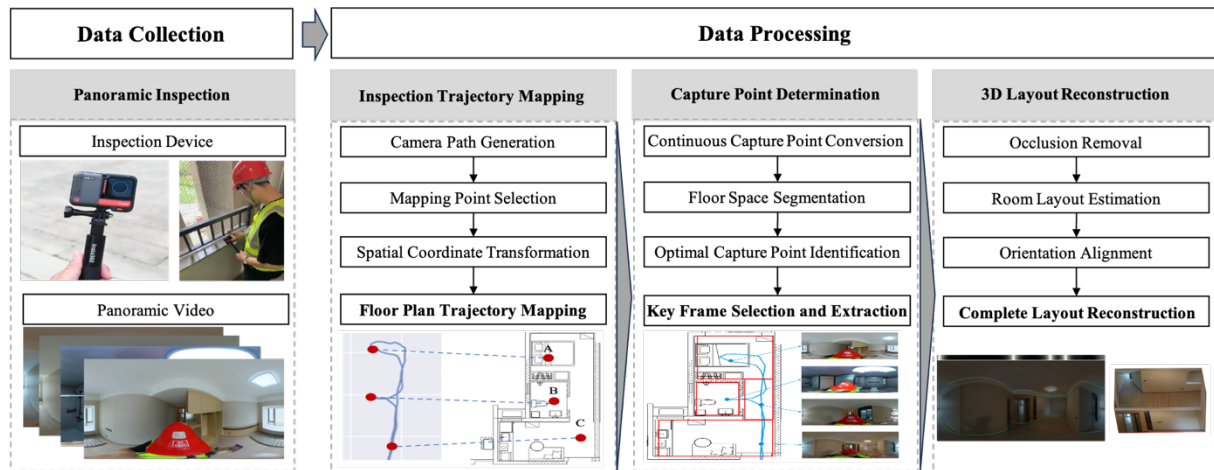


Figure 1. Framework of panoramic image-based 3D layout reconstruction for interior construction space

2. RELATED WORK

2.1 Vision-based 3D Reconstruction

The generation of vision-based 3D reconstruction can be achieved using active or passive methods, distinguished by how depth information is extracted from images. Active methods use RGB-Depth cameras directly to acquire depth data, with Structured Light (Geng 2011) and Time-of-Flight (ToF) (Cui et al. 2010) being common techniques. In contrast, passive methods rely on environmental reflections, such as natural light, to estimate geometric data and depth, which is then used to generate a dense point cloud. Due to limitations of active methods, including environmental and equipment constraints, recent research has increasingly focused on passive vision approaches. Generally, the passive methods can be categorized based on the number of cameras: monocular, binocular, and multi-view systems (Khan et al. 2020).

Monocular vision uses a single camera for 3D reconstruction and is a simple, flexible, and cost-effective method with fast processing times. Its versatility makes it widely applicable in areas such as 3D object measurement and site inspection. Monocular vision extracts features from images like brightness, depth, texture, contours, geometric shapes, and key points. The data processing includes feature extraction, matching, camera motion estimation, bundle adjustment, and depth recovery (Lu et al. 2024). The first four steps form the Structure from Motion (SfM) process, which estimates camera movement and generates sparse point clouds. This sparse reconstruction is then enhanced with Multi-View Stereo (MVS) to achieve spatial consistency, cluster points, and fill in missing data, resulting in a dense point cloud (Furukawa and Hernández 2015). However, the narrow field of view of a standard monocular camera requires capturing multiple images to obtain information about the entire scene, making it challenging to cover all angles. Additionally, since feature point matching is necessary, the images used for reconstruction must have sufficient overlap, increasing the data acquisition complexity.

2.2 Single Panoramic Image-Based 3D Layout Reconstruction

Single-image reconstruction simplifies the image-based 3D reconstruction problem by relying on a single input image rather than multiple images, providing parallax for depth estimation. This approach benefits from easy data acquisition and a simplified model. With the growing prevalence of 360-degree cameras and panoramic techniques applications in engineering fields, layout reconstruction based on panoramic images has become a popular research focus in computer vision (Cinnamon and Jahiu 2023). Single panoramic image-based 3D layout reconstruction methods are categorized into two main types: constraint-based and deep learning-based approaches.

The constraint-based 3D layout reconstruction approach extracts structural features from images using geometric constraints to estimate 3D layout. These constraints draw from prior knowledge and include vanishing points, parallelism, coplanarity, orthogonality, and perspective relationships (Hoiem et al. 2007). Among these, the Manhattan world assumption is frequently applied in single-image 3D layout reconstruction algorithms. This assumption posits that most artificial environments align object lines and planes with one of the three axes of the Cartesian coordinate system (Coughlan and Yuille 1999). Researchers have widely utilized this assumption for

single-image 3D layout reconstruction, particularly in architectural applications, where it allows the segmentation of indoor scenes into floors, walls, and ceilings, thereby reconstructing the 3D model of the scene. Additionally, several scholars have expanded the Manhattan World model to improve its applicability in real-world environments, enhancing its accuracy and effectiveness in diverse scenarios (Straub et al. 2018).

The deep learning-based 3D layout reconstruction approach constructs a target training dataset to map semantic labels to scene geometry using large volumes of data. This method effectively predicts a room's geometric structure through end-to-end learning techniques. LayoutNet (Zou et al. 2018) inputs a pre-processed Manhattan line map into a U-Net structure consisting of an encoder and decoder, generating a probability map for wall corners and boundaries between walls, floors, and ceilings, marking the first deep learning approach to perform 3D layout reconstruction as an end-to-end learning task. Dula-Net (Yang et al. 2019) introduces a deep learning framework for predicting Manhattan-world 3D room layouts from a single RGB panorama, using a dual-branch architecture with equirectangular and perspective ceiling views connected through a feature fusion scheme to enhance accuracy in 2D-floor plans and layout height predictions. LED2-Net (Wang et al. 2021) conducts a differentiable layout-to-depth transformation, which reformulates 360-degree room layout estimation as a 360-degree depth estimation problem, enabling end-to-end training and improving model generalizability by integrating geometric information. HorizonNet (Sun et al. 2019) introduces a 1D representation for 3D room layout estimation from panoramic images, using end-to-end learning and Pano Stretch Data Augmentation to reduce computational complexity and accurately recover complex room shapes. Its advantages over existing methods include lower computational cost and more straightforward implementation, improving performance and generalizability across various indoor environments.

2.3 3D Layout Reconstruction For Construction Management

Information gathering on construction sites traditionally relies on manual inspections and record-keeping, resulting in low levels of digitization and making it difficult for managers to grasp comprehensive site information effectively. The advent of 360-degree capture devices, which can document an entire scene in a single shot, has enabled the effective use of panoramic images for recording construction site conditions.

Research has conducted a comprehensive review of 360° panoramic visualization technologies in the AEC industry, highlighting key applications and benefits for enhancing construction education, monitoring, visualization, and safety training (Shinde et al. 2023). Panoramic image-based 3D layout reconstruction has been applied in construction management scenarios, including automatic progress assessment of interior construction sites (Fang et al. 2023), detection of indoor functional elements like dome lights and outlets (Pintore et al. 2018), partial construction space reconstruction virtual reality applications (Feng et al. 2018). However, existing researches have focused on generating 3D models from panoramic images captured in stationary positions, with limited exploration of mobile technology for capturing panoramic images to reconstruct complete interior construction spaces. This paper explores 3D layout reconstruction of interior construction sites through mobile inspections using a portable 360-degree panoramic camera, ensuring minimal disruption to daily operations.

3. METHODOLOGY

3.1 Inspection Trajectory Mapping

This paper utilizes the OpenVSLAM (Sumikura et al. 2019) algorithm to analyze and process panoramic inspection videos to obtain inspection trajectories, recording the spatial positions and location information at each moment during the inspection process. As an indirect method that extracts ORB feature points from target scenes, OpenVSLAM operates through three modules: tracking, mapping, and global optimization.

To obtain the position information of the inspection trajectory from camera to floor plan, an affine transformation is applied to associate the pixel coordinate system C_p with the camera trajectory coordinate system C_t . This process involves fundamental operations, including scaling, rotation, and shearing. By abstracting these fundamental operations, we simplify the task of matching the path to the floor plan into an affine transformation process. Specifically, from the path generated by OpenVSLAM, three trajectory coordinates requiring relocation are recorded as (x_i^t, y_i^t) . Given the existing construction floor plan, three reference points are selected on the plan and recorded as (x_i^p, y_i^p) . Next, the affine matrix M calculates the pixel coordinates of the inspection trajectory points on the floor plan, as shown in Equation 1, where a is the Linear Transformation Matrix and t is the Translation Vector. Finally, the pixel coordinates of all trajectory points are plotted on the floor plan, illustrating the actual inspection path and trajectory within the plan.

$$C_p = C_t M = \begin{bmatrix} x'_i & y'_i & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x'_i & y'_i & 1 \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{12} \\ t_1 \\ a_{21} \\ a_{22} \\ t_2 \end{bmatrix} \quad (1)$$

3.2 Capture Point Determination

After mapping the inspection trajectory onto the floor plan, the system divides it into continuous capture points based on the building's layout. The target interior space is segmented into multiple rectangles, with each rectangle's center assumed to be the best capture point. The system calculates and calibrates the center coordinates for each segment. The optimal capture point is the one closest to the center of each rectangle, ensuring a comprehensive view of the interior space.

The system then constructs a KD Tree using the camera's capture points along the trajectory for an efficient nearest neighbor search (Bentley 1975). The global optimal capture point search algorithm traverses the KD Tree, as shown in Equation (2), and identifies the closest capture point to each calibrated center based on the shortest distance. This method locates the point along the trajectory nearest the target space's center, ensuring the optimal viewpoint for capturing the panoramic image and extracting the corresponding video frame.

$$Nearest(q) = \arg \min_{p_i \in P} d(q, p_i) \quad \text{where} \quad d(q, p_i) = \sqrt{(x_q - x_i)^2 + (y_q - y_i)^2} \quad (2)$$

3.3 3D Layout Reconstruction

Since the inspection recording process involves handheld shooting, the resulting panoramic images in the form of equirectangular projection contain large areas of occlusion (e.g., the body and recording device), leading to the loss of indoor layout corner points and boundary details. This section uses the Large Mask Inpainting (LaMa) (Suvorov et al. 2021) algorithm to remove occlusions and restore boundary information. Using the original RGB image and the mask image from the panoramic inspection video as inputs, the algorithm, based on Fourier Convolutions, processes the image by leveraging the global structure and contextual information to generate natural and coherent inpainting results. This method retrieves the complete background and restores the full boundary contours of the interior room space.

After Recovering scene boundaries and corner points from indoor panoramic images, HorizonNet reconstructs the room layout into a point cloud format. The panoramic image captured by the camera is an equidistant rectangular image projected from a square relative to the camera's coordinate system. To ensure that the final layout aligns with the world coordinate system, the panoramic image must be re-projected before being input into the HorizonNet neural network. The camera's pose is determined using its 6-axis IMU data and the Mahony algorithm. The py360convert tool then uses the Euler angles derived from the pose calculation to re-project the panoramic image into the world coordinate system.

The image data processing based on HorizonNet can be divided into three stages: pre-processing, processing, and post-processing. In the pre-processing stage, the algorithm starts with panoramic images as input, where the data is aligned to ensure consistency under the equirectangular projection. This alignment helps simplify the process of detecting vertical wall boundaries, a critical step for accurately estimating the room layout. In the Processing stage, ResNet-50 is used to extract relevant multi-scale features. HorizonNet employs a 1D representation that predicts floor-wall, ceiling-wall, and wall-wall boundaries for each image column. A bidirectional LSTM processes these columns sequentially, capturing long-range dependencies and geometric patterns across the layout. In the Post-processing stage, the predicted 1D boundaries are projected into 3D space using the assumption of a 1.8-meter camera height for real-world scaling. HorizonNet enhances accuracy through Pano Stretch Data Augmentation, which generates diverse training samples. The final result is a detailed point cloud model that captures the room's geometric structure, suitable for both cuboid and non-cuboid layouts. The flowchart of 3D layout reconstruction for interior space is shown in Figure 2.

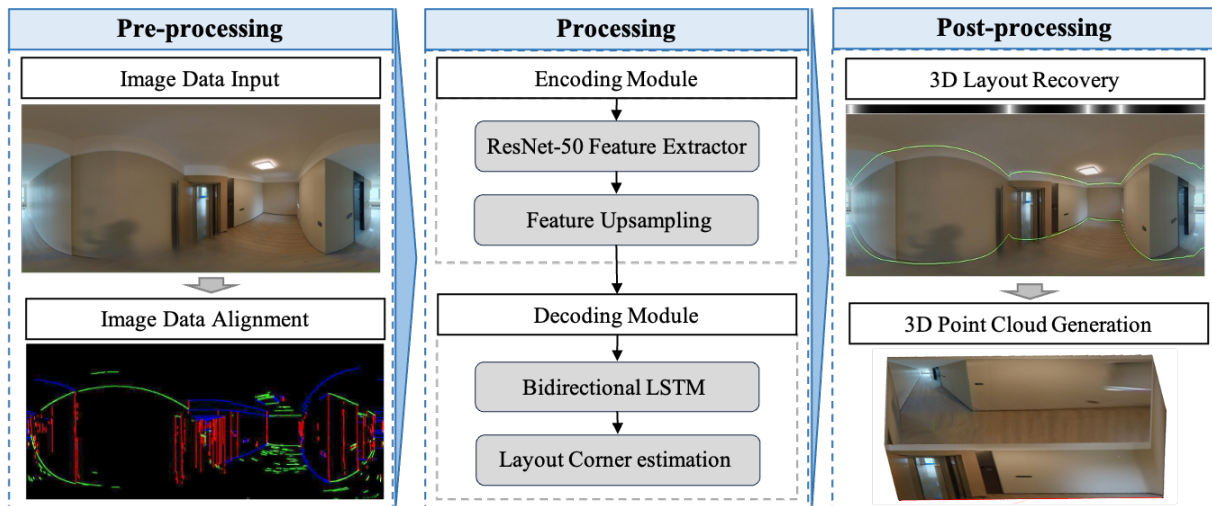


Figure 2. Flowchart of HorizonNet

4. EXPERIMENT

The proposed method was validated on a concrete Modular Integrated Construction (MiC) project during the final interior finishing phase. A handheld Insta360 One RS was used for inspection, recording an 87-second panoramic video. After equidistant cylindrical projection, the OpenVSLAM algorithm reconstructed the inspection trajectory, which was projected onto known construction blueprints using affine transformation, as shown in Figure 3. The trajectory was discretized into inspection points, and the KD-tree search identified the optimal points nearest to the center of each space, as shown in Figure 4. Frames from these points were extracted for further image processing and layout reconstruction.

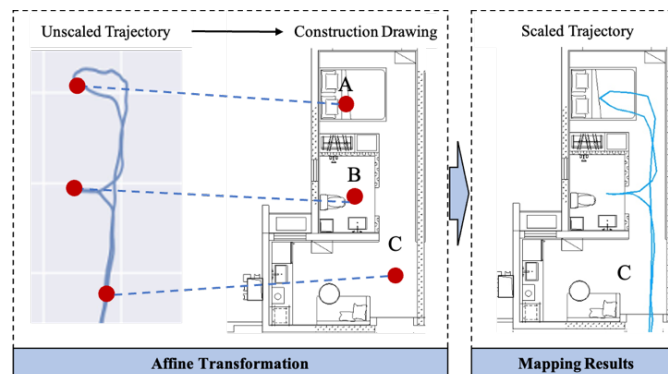


Figure 3. Matching process of inspection path trajectory with the floor plan

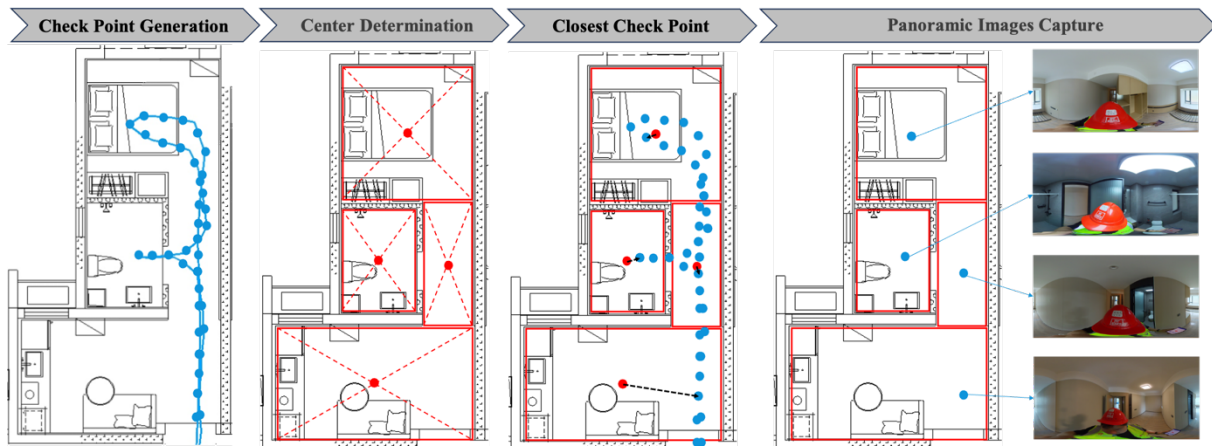


Figure 4. Determination of optimal capture point from inspection trajectory

Before image processing, IMU data calibrated the camera's pose to ensure that all panoramic images used for layout reconstruction had a unified orientation. This calibration step was essential for aligning the images correctly in 3D space. Using the LAMA algorithm for key frame occlusion removal, the process begins by generating a mask for the occluded regions in the original image. The original image and its corresponding mask are then input into the LAMA algorithm, which inpaints the occluded areas and restores corner points along the boundaries of the panoramic image, such as where the floor meets the walls. The steps for removing occlusions in panoramic images are shown in Figure 5.

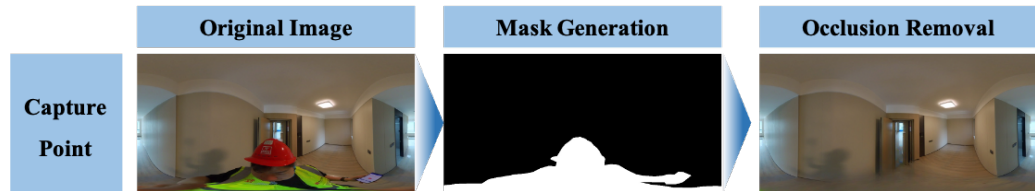


Figure 5. Process of occlusion removal for panoramic image

Our research employed HorizonNet to reconstruct the target room layouts from panoramic images. The model was trained on a computing system equipped with an Intel® Xeon® CPU E5-2683 v4 @ 2.10GHz and two Nvidia GeForce RTX 3090 GPUs, running on the Ubuntu 20.04.06 operating system. The training dataset consisted of a total of 335 panoramic images collected from three different construction sites. Of these, 88 images contained occlusions, while 247 images were free from occlusions. The dataset encompassed various room types, including living rooms, dining rooms, bedrooms, restrooms, kitchens, and hallways. The dataset was partitioned into a training set (268 images, 80%) and a testing set (67 images, 20%) to ensure robust model training and evaluation. The model was fine-tuned using pre-trained weights from the Structure3D model, with a learning rate of 0.0003, a batch size of 16, and training conducted over 100 epochs. The fine-tuning process took approximately 2.5 hours to complete. To evaluate the performance of the trained model, we used two key metrics: 3D Intersection over Union (IoU) and Corner Error. The 3D IoU metric was employed to assess the overlap between the 3D room layout predicted by HorizonNet and the corresponding ground truth, achieving an IoU score of 77.2%. Additionally, the Corner Error, calculated as 0.82%, quantifies the Euclidean distance between the predicted and actual corner points of the room layout, normalized by the diagonal length of the image. These two metrics provide valuable insights into the model's accuracy and precision in predicting room layouts.

The complete process of room layout reconstruction, including all the steps involved, is visually summarized in Figure 6. However, it is important to note that dynamic elements present on construction sites—such as moving objects, workers, or temporary structures—can significantly influence layout accuracy. These elements may cause occlusion of key layout features, such as corners, during the image capture process, thereby introducing inaccuracies in the reconstruction. To mitigate this issue, images with occluded areas larger than one-third of the image resolution should be excluded from the dataset, as they may lead to unreliable predictions and degrade model performance. This consideration is crucial for maintaining the integrity of the dataset and ensuring the quality of the room layout reconstruction.

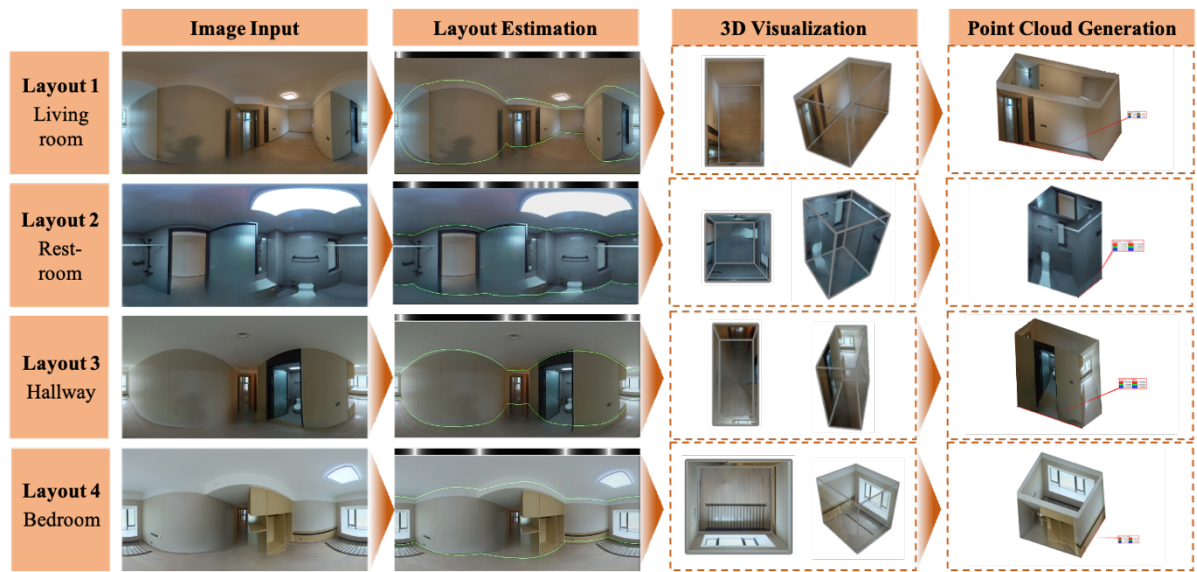


Figure 6. The process of room layout reconstruction

In HorizonNet, a camera height of 1.8 meters is assumed for projecting the predicted 2D layout boundaries (floor-wall and ceiling-wall) into a 3D space. This assumption allows for the accurate calculation of floor-to-ceiling distances and wall positions relative to the camera, facilitating realistic room reconstruction from panoramic image inputs. Following the 3D layout reconstruction, one boundary from each reconstructed layout was selected and compared with the as-designed floor plan. The average discrepancy between the two was approximately 0.3 meters, as depicted in Figure 7. However, the study is subject to two limitations regarding layout estimation and real-scale recovery. In terms of layout estimation, the average prediction accuracy is constrained by the limited number of panoramic images available from the construction site. Regarding real-scale recovery, measurement errors may arise from fluctuations in camera height (± 5 cm) during inspections, which limit the precision of the 3D reconstruction. Future research could focus on augmenting panoramic image data through image stretching techniques. Additionally, the adoption of a fixed-height camera system may enhance the accuracy of future studies.

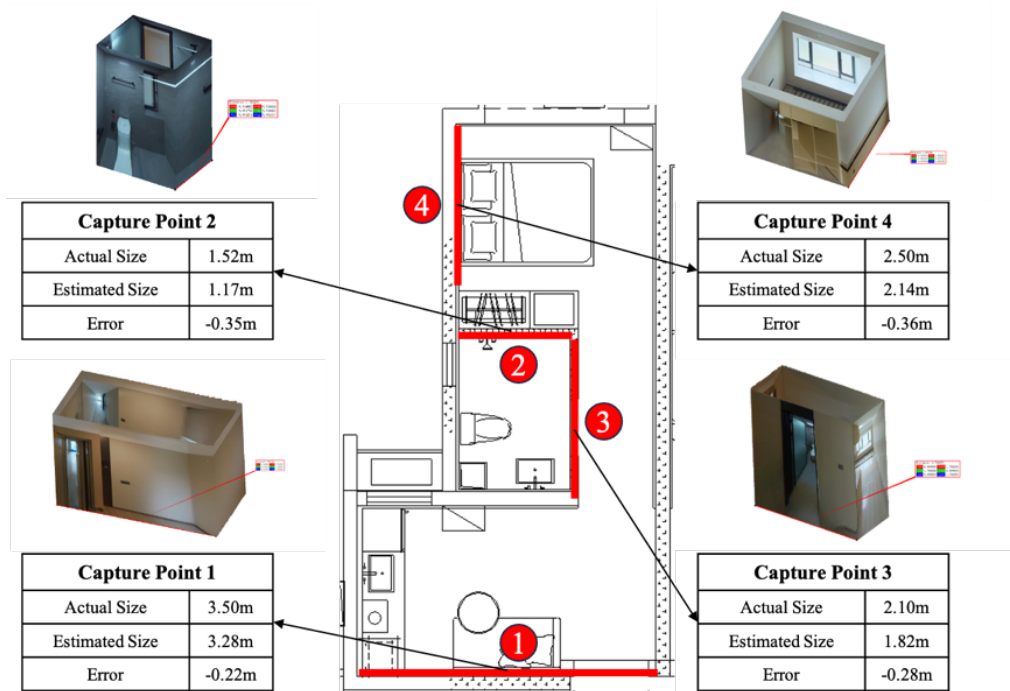


Figure 7. Size evaluation of proposed method for room layout reconstruction

5. CONCLUSIONS

This paper introduces a new strategy for 3D layout reconstruction of interior construction sites using a portable 360-degree panoramic camera. The three key contributions are inspection trajectory mapping, capture point determination, and 3D layout reconstruction. It addresses challenges in capturing optimal panoramic data through an OpenVSLAM-based path mapping and capture point retrieval algorithm. The Kd-tree is applied to locate points along the trajectory nearest to the center of each target space, ensuring the best viewpoint for capturing panoramic images and extracting key video frames. Using these key images, HorizonNet reconstructs 3D layouts, achieving 77.2% 3D IoU and 0.82% Corner Error, with a 0.3-meter boundary error compared to as-designed floor plans. This method serves as an alternative to stationary layout estimation, offering valuable insights into intelligent site inspections and precise spatial modeling.

This research can stand out as a flexible tool for future construction site monitoring, enhancing accuracy and real-time capability in dynamic environments. By incorporating interior layout reconstruction with higher accuracy, construction professionals can improve both the efficiency and completeness of site inspections, ensuring more reliable data collection and spatial analysis throughout the construction process.

ACKNOWLEDGMENTS

The study was supported by the National Key Research & Development Program of China (2022YFC3801700), The Science Foundation for the Science and Technology Commission of Shanghai Municipality (22dz1207100 and 22dz1207800), Fundamental Research Funds for the Central Universities (2024-

1-ZD-02).

6. REFERENCE

- Bentley, J. L. 1975. "Multidimensional binary search trees used for associative searching." *Commun. ACM*, 18 (9): 509–517. <https://doi.org/10.1145/361002.361007>.
- Cinnamon, J., and L. Jahiu. 2023. "360-degree video for virtual place-based research: A review and research agenda." *Computers, Environment and Urban Systems*, 106: 102044. <https://doi.org/10.1016/j.compenvurbsys.2023.102044>.
- Coughlan, J. M., and A. L. Yuille. 1999. "Manhattan World: compass direction from a single image by Bayesian inference." *Proceedings of the Seventh IEEE International Conference on Computer Vision*, 941–947 vol.2. Kerkyra, Greece: IEEE.
- Cui, Y., S. Schuon, D. Chan, S. Thrun, and C. Theobalt. 2010. "3D shape scanning with a time-of-flight camera." *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1173–1180. San Francisco, CA, USA: IEEE.
- Durrant-Whyte, H., and T. Bailey. 2006. "Simultaneous localization and mapping: part I." *IEEE Robot. Automat. Mag.*, 13 (2): 99–110. <https://doi.org/10.1109/MRA.2006.1638022>.
- Fang, X., H. Li, H. Wu, L. Fan, T. Kong, and Y. Wu. 2023. "A fast end-to-end method for automatic interior progress evaluation using panoramic images." *Engineering Applications of Artificial Intelligence*, 126: 106733. <https://doi.org/10.1016/j.engappai.2023.106733>.
- Feng, Z., V. A. González, L. Ma, M. M. A. Al-Adhami, and C. Mourgues. 2018. "Rapid 3D Reconstruction of Indoor Environments to Generate Virtual Reality Serious Games Scenarios." arXiv. <https://doi.org/10.48550/ARXIV.1812.01706>.
- Furukawa, Y., and C. Hernández. 2015. "Multi-View Stereo: A Tutorial." *FNT in Computer Graphics and Vision*, 9 (1–2): 1–148. <https://doi.org/10.1561/06000000052>.
- Geng, J. 2011. "Structured-light 3D surface imaging: a tutorial." *Adv. Opt. Photon.*, 3 (2): 128. <https://doi.org/10.1364/AOP.3.000128>.
- Guo, F., C. T. Jahren, Y. Turkan, and H. David Jeong. 2017. "Civil Integrated Management: An Emerging Paradigm for Civil Infrastructure Project Delivery and Management." *J. Manage. Eng.*, 33 (2): 04016044. [https://doi.org/10.1061/\(ASCE\)ME.1943-5479.0000491](https://doi.org/10.1061/(ASCE)ME.1943-5479.0000491).
- Hoiem, D., A. A. Efros, and M. Hebert. 2007. "Recovering Surface Layout from an Image." *Int J Comput Vis*, 75 (1): 151–172. <https://doi.org/10.1007/s11263-006-0031-y>.
- Khan, F., S. Salahuddin, and H. Javidnia. 2020. "Deep Learning-Based Monocular Depth Estimation Methods—A State-of-the-Art Review." *Sensors*, 20 (8): 2272. <https://doi.org/10.3390/s20082272>.
- Lu, Y., S. Wang, S. Fan, J. Lu, P. Li, and P. Tang. 2024. "Image-based 3D reconstruction for Multi-Scale civil and infrastructure Projects: A review from 2012 to 2022 with new perspective from deep learning methods." *Advanced Engineering Informatics*, 59: 102268. <https://doi.org/10.1016/j.aei.2023.102268>.
- Pintore, G., R. Pintus, F. Ganovelli, R. Scopigno, and E. Gobbetti. 2018. "Recovering 3D existing-conditions of indoor structures from spherical images." *Computers & Graphics*, 77: 16–29. <https://doi.org/10.1016/j.cag.2018.09.013>.
- Shinde, Y., K. Lee, B. Kiper, M. Simpson, and S. Hasanzadeh. 2023. "A Systematic Literature Review on 360° Panoramic Applications in Architecture, Engineering, and Construction (AEC) Industry." *ITcon*, 28: 405–437. <https://doi.org/10.36680/j.itcon.2023.021>.
- Straub, J., O. Freifeld, G. Rosman, J. J. Leonard, and J. W. Fisher. 2018. "The Manhattan Frame Model—Manhattan World Inference in the Space of Surface Normals." *IEEE Trans. Pattern Anal. Mach. Intell.*, 40 (1): 235–249. <https://doi.org/10.1109/TPAMI.2017.2662686>.
- Sumikura, S., M. Shibuya, and K. Sakurada. 2019. "OpenVSLAM: A Versatile Visual SLAM Framework." *Proceedings of the 27th ACM International Conference on Multimedia*, 2292–2295. Nice France: ACM.
- Sun, C., C.-W. Hsiao, M. Sun, and H.-T. Chen. 2019. "HorizonNet: Learning Room Layout With 1D Representation and Pano Stretch Data Augmentation." *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1047–1056. Long Beach, CA, USA: IEEE.
- Suvorov, R., E. Logacheva, A. Mashikhin, A. Remizova, A. Ashukha, A. Silvestrov, N. Kong, H. Goka, K. Park, and V. Lempitsky. 2021. "Resolution-robust Large Mask Inpainting with Fourier Convolutions." arXiv.
- Ullman, S. 1979. "The interpretation of structure from motion." *Proc. R. Soc. Lond. B.*, 203 (1153): 405–426. <https://doi.org/10.1098/rspb.1979.0006>.
- Verykokou, S., and C. Ioannidis. 2023. "An Overview on Image-Based and Scanner-Based 3D Modeling Technologies." *Sensors*, 23 (2): 596. <https://doi.org/10.3390/s23020596>.
- Wang, F.-E., Y.-H. Yeh, M. Sun, W.-C. Chiu, and Y.-H. Tsai. 2021. "LED²-Net: Monocular 360° Layout Estimation via Differentiable Depth Rendering." *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 12951–12960. Nashville, TN, USA: IEEE.
- Yang, S.-T., F.-E. Wang, C.-H. Peng, P. Wonka, M. Sun, and H.-K. Chu. 2019. "DuLa-Net: A Dual-Projection

361 Network for Estimating Room Layouts From a Single RGB Panorama.” *2019 IEEE/CVF Conference on*
 362 *Computer Vision and Pattern Recognition (CVPR)*, 3358–3367. Long Beach, CA, USA: IEEE.
 363 Zou, C., A. Colburn, Q. Shan, and D. Hoiem. 2018. “LayoutNet: Reconstructing the 3D Room Layout from a
 364 Single RGB Image.” *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2051–
 365 2059. Salt Lake City, UT: IEEE.
 366

A REAL-WORLD VISUAL SLAM DATASET FOR INDOOR CONSTRUCTION SITES

Wenyu Li¹, Xinyu Chen², and Yantao Yu³

1) Research Assistant, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: wenyuli@ust.hk

2) Ph.D. Candidate, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: xchengl@connect.ust.hk

3) Ph.D., Assoc. Prof., Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: ceyantao@ust.hk

Abstract: This paper presents a RGBD slam construction dataset with a mounted platform, designed to collect the unique challenges encountered in construction sites. An Ouster OS0-128 LiDAR is utilized as the sensor of LiDAR SLAM, working as the ground truth for localization. Our dataset records various construction settings with different stages of building materials and structures, such as concrete, brick, plaster, and putty, providing a comprehensive benchmark for training and evaluating SLAM algorithms. Through testing on current SLAM algorithms, we demonstrate the limitations of traditional approaches in these environments and provide a VINS based algorithm as the benchmark. This dataset serves as a valuable resource for researchers aiming to enhance SLAM performance in the real construction environments. The detailed information of the dataset is available at <https://github.com/WenyuLWY/HCIC-Construction-VSLAM-Dataset.git>

Keywords: Visual SLAM, Construction Robot, Dataset

1. INTRODUCTION

In the field of automated construction, the deployment of construction robots has gained significant traction for the automation of tasks such as surveying, inspection, and material handling. Compared to human workers, construction robots can work continuously without fatigue, reducing project costs, while also improving the overall quality and consistency of different construction tasks. A key factor in enabling these robots to operate effectively in complex construction environments is accurate localization. Without precise positioning, robots cannot operate reliably across different sites, which is critical for ensuring safe and efficient automation.

Indoor localization presents greater challenges compared to outdoor environments due to the absence of GPS and other external positioning systems. To address these challenges, Simultaneous Localization and Mapping (SLAM) has developed as a foundational technology in the field of construction robotics, playing an essential role in enabling robots to autonomously navigate and localize within complex construction environments.

Although LiDAR-based SLAM is commonly used in construction, it faces limitations such as high costs and sensor weight. Visual SLAM (V-SLAM), by contrast, has gained prominence for its advantages. It uses cameras, which are more cost-effective and lightweight, and can capture rich environmental details such as texture, color, and structure, providing comprehensive information for mapping and localization. With recent advancements in computer vision and image processing, visual SLAM has also become more robust, offering greater flexibility in challenging construction environments.



Figure 1. Typical indoor working conditions for construction robots (Figures are from Bright Dream Robotics(BDR) company in Foshan, China, which provides us the building sites for data collecting.)

Using V-SLAM datasets is an effective method for testing and evaluating the performance of localization algorithms. However, most existing SLAM benchmark have been collected in finished indoor environments, such as offices or homes (Sturm et al., 2012), or in outdoor urban and autonomous driving scenarios(Geiger et al., 2013). Recently, more challenging general indoor scene datasets have proposed (Shi et al, 2020; Yin et al, 2022). These

popular datasets are typically ideal environments for SLAM experiments. Traditional visual SLAM systems built on these datasets perform well in such scenarios, where stable features such as corners, edges are reliable for feature extraction and matching. However, they will experience a series of challenges when applied to construction environments. A construction site is usually defined as low textural, structural, and with large surfaces. Figure 1 provides real examples for illustrating the working conditions of construction robots. The extreme setting results in poor performance and inaccurate localization of traditional feature point based SLAM algorithms. We directly evaluate ORBSLAM (Campos et al., 2021), a widely used SLAM algorithm known for its effectiveness in general localization tasks, as shown in Figure 2. In Figure 2(a), although the system detects some feature points only on a relatively rough wall, it still estimates its pose. However in Figure 2(b), it can not detect any feature from the image, therefore the system drifts with imu and fails soon.

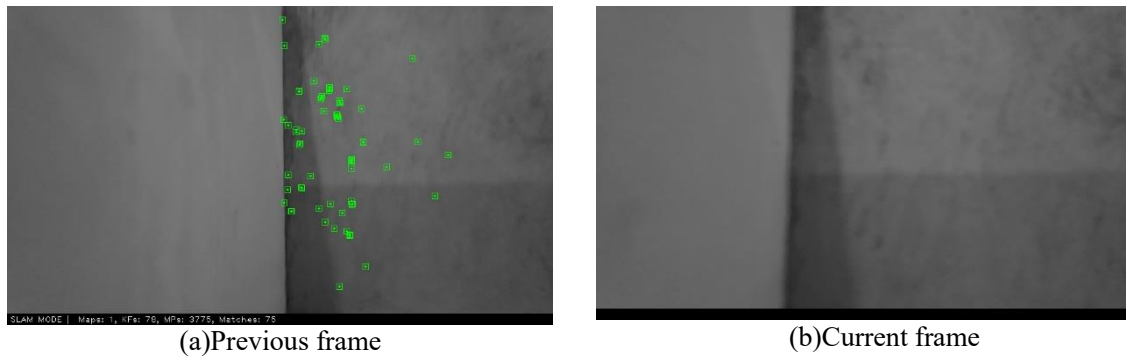


Figure 2. ORB feature extraction in our construction dataset(From ORBSLAM3)

This indicates that SLAM algorithms developed based on general datasets are unsuitable for real construction sites. The primary reason for this limitation is the lack of specialized SLAM datasets tailored to construction environments, which hinders the advancement of SLAM algorithms designed for construction robots. To address these issues, this paper introduces a novel dataset specifically created to capture the complexities of construction settings. This dataset serves as a comprehensive benchmark for training and evaluating SLAM algorithms, enabling them to better tackle the challenges inherent in real-world construction environments.

2. DATASET AND Methodology

2.1 Overview

The dataset was collected in indoor construction sites with various types of walls, ceilings, and other architectural characteristics typical of construction environments, such as unfinished structures, exposed pipes, and scattered materials. These scenes are complex and diverse, containing numerous geometric features and dynamic elements such as construction workers and moving equipment. This setting provides a unique challenge for evaluating visual SLAM algorithms.

For the core visual images in our dataset, we chose to capture RGB-D images instead of monocular or stereo images, which could directly provide depth information for each pixel. We have compared these cameras and finally make the decision for some reasons: First, in real-world navigation tasks, construction robots normally require absolute scales for accurate positioning and mapping, while the monocular slam cannot provide this absolute scale, making it unsuitable for reliable navigation in practical environment. Second, visual-inertial systems heavily depend on the Inertial Measurement Unit (IMU) to obtain scale. However, for ground robots typically move at a constant velocity on a 2D plane, the IMU will not be fully excited in each direction, therefore there will be barely no useful measurements. This can lead to divergence, large drift error and scale drift. Third, though stereo vision estimates the depth by matching features between the two camera images, it faces challenges in low-texture or changing lighting conditions commonly found in construction environments, making it difficult for stereo cameras to recover reliable scale and depth.

Finally, our robust solution is to provide an RGB-D camera for the visual SLAM benchmark. The usage of this kind of cameras could help reducing computational load and improving data processing efficiency, and focusing on the development of new visual SLAM algorithms. We have compared our dataset with several current construction slam datasets, the comparison is shown in Table 1.

Table 1. Comparison of current construction SLAM datasets

Dataset	Motion	Visual Sensor Type	LiDAR Sensor
Hilti (Helmberger et al., 2022)	Handheld	Stereo/Grayscale	OS0-64/Livox MID70
ConSLAM (Trzeciak et al., 2023)	Handheld	RGB/NIR	Velodyne VIP-16
Ours	Mounted	RGB-D	OS0-128

We noticed that we are the first that provide RGBD images. The Hilti dataset (Helmberger et al., 2022) might be the most well-known construction SLAM dataset, but it only collects 10hz stereo images using Alphasense cameras, due to the limitation by the image resolution. The ConSLAM dataset (Trzeciak et al., 2023) is primarily a lidar slam dataset, so it only uses a simple RGB camera. Our dataset will be an ideal choice to test and develop visual slam in real construction sites.

2.2 Ground Truth

It is challenging to generate accurate reference positions for SLAM performance evaluation in indoor construction sites. GPS is often unreliable indoors due to weak signals and multipath effects, failing to provide the required precision. Some research employs Total Stations or Terrestrial Laser Scanners (TLS) to generate highly accurate ground truth data (Helmberger et al., 2022; Trzeciak et al., 2023). However, these methods are frequently constrained by the complexity and slowness of operation, making them less suitable for real-time and changing indoor settings.

While some research has tried to use advanced lidar slam as the ground truth. TIERS(Qingqing et al., 2022) provides a reference location from lidar slam when motion capture system is unavailable in largescale environment. ConPR (LEE et al., 2023) fuses lidar slam and GPS measurement to generate global absolute trajectories for place recognition. Therefore, high-precision lidar systems can generate stable and accurate positioning data. In our study, we adopt a similar approach to directly obtain ground truth. An advanced lidar-inertial odometry Fast-LIO2(Xu et al., 2022) is selected, with some modifications to ensure compatibility with our lidar. But users are allowed to use their own lidar slam algorithm as reference. Figure 3 shows the mapping and localization result after running the lidar slam in our lab, where the colored lines in the pictures represent the ground truth trajectories.

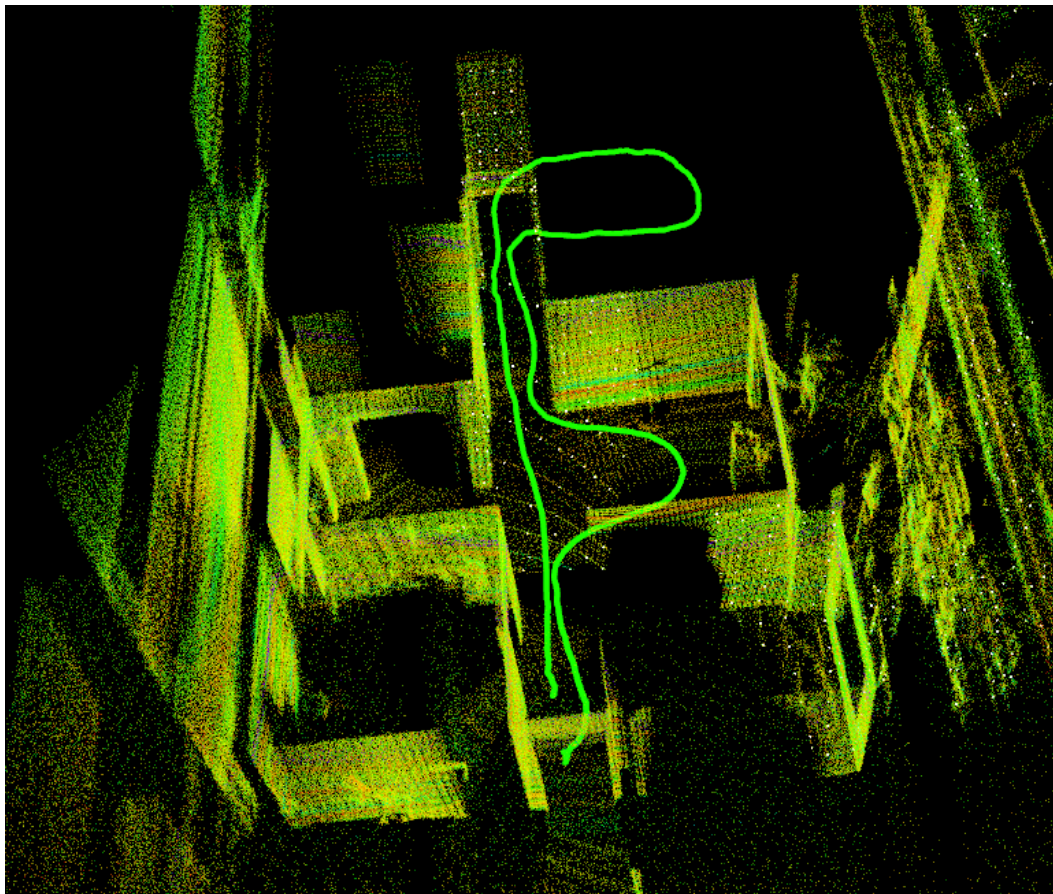


Figure 3. Results of the lidar slam on a test sequence

2.3 Hardware

Data collection was conducted using an Intel RealSense L515 camera, an Ouster OS0-128 LiDAR sensor, and both equipped with their built-in imu modules. The Ouster OS0-128 is a spinning high precision LiDAR sensor that offers a full 360-degree horizontal field of view and a 90-degree vertical range, making it ideal for comprehensive scene capture. With a range of up to 50 meters and a 10 Hz operating rate, this sensor effectively

detects distant objects and provides accurate data for SLAM applications, aided by its integrated IMU for real-time positioning. The Intel RealSense L515 detects highly precise depth measurements up to 9 meters with a field of view of $70^\circ \times 55^\circ$, tailored for detailed indoor applications. Its higher capture rate of 30 Hz enables it to track rapid changes in dynamic environments. Additionally, the RGBD camera is also regarded as a solid-state lidar, due to its extremely accurate depth measurement. However, there is some debate about whether its use falls under the category of visual SLAM (Li et al., 2022 ;Qingqing et al., 2022). The general consensus is that if the SLAM algorithm primarily relies on RGB images for visual feature extraction and uses depth images as supplementary input, it is still considered as visual RGBD SLAM (Xu et al., 2023). On the other hand, methods that directly use point clouds generated from the depth images for localization are typically regarded as LiDAR SLAM (Wang et al., 2021). Since the proposed dataset primarily utilizes RGB images and depth images, it aligns with the definition of visual SLAM and our dataset is categorized as a visual SLAM dataset. The specific configurations of the sensors are listed in Table 2.

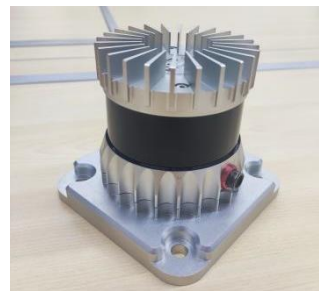
Table 2. Sensor configurations on the platform

Sensor	Type	IMU	FoV	Range	Rate
OS0-128	spinning	built-in	$360^\circ \times 90^\circ$	50 m	10hz
RealSense L515	solid-state/RGBD	built-in	$70^\circ \times 55^\circ$	9 m	30hz

The data is recording through a laptop, using Robot Operating System (ROS) with Ubuntu 20.04. The camera is placed on a tripod for stability (Figure 4(a)). The lidar, as shown in Figure4(b), is powered by an external power source and is connected to the laptop via an Ethernet cable. All devices are mounted on a manually operated platform, which allows for flexible movement and data collection.



(a)RGBD camera



(b)LiDAR sensor

Figure 4. Sensor configuration and data collection platform

2.4 Sensor Data and Software

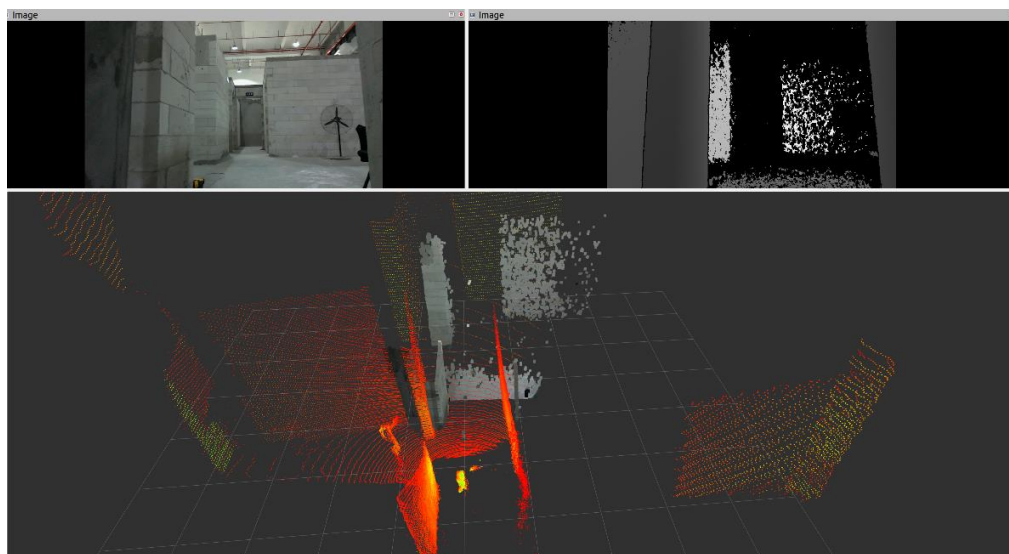


Figure 5. sensor data of lidar and RGBD camera

Figure 5 is an example of sensor data in one scanning. The above two images are RGB and depth image respectively. Each pixel in the depth image indicates the distance of the corresponding pixel in the RGB image. The red pointcloud is from OS0-128 lidar. It is used to register current scan to the pointcloud map. The colored

pointclouds are generated by the L515 camera. Benefitting from the RGB-D mode, it is convenient to obtain the rgb pointclouds or receive depth images aligned with the RGB images. Since RGB pointclouds can be recovered from depth images, and the transmission rate of images is also much faster than that of pointclouds, we chose to use depth images rather than directly providing colored pointclouds in our dataset. In Figure 5 the colored pointclouds are provided by some third-party utilities for visualization.

To transform the depth image to pointcloud, we can use the camera intrinsics(given in /camera/color/camera_info topic, as listed in Figure 5) by Equation (1):

$$\begin{aligned} x &= \frac{(j - cx)z}{fx} \\ y &= \frac{(i - cy)z}{fy} \\ z &= \text{depth_value} * \text{depth_factor} \end{aligned} \quad (1)$$

where the depth_value is from the depth image at (i,j); the depth_factor is set to 0.001 for this camera; fx,fy,cx,cy are the camera intrinsics provided by the realsense driver; x,y,z are the 3D coordinates respectively.

Some RGBD SLAM directly use pointcloud for Iterative Closest Point (ICP) matching rather than operate on the depth image (Gu & Meng, 2023; Xu et al., 2023; Wang et al., 2021), and shows impressive performances. However, due to limitations in on board computation, they need to downsample the depth maps multi threading parallelization before further proceeding.

In this dataset we provide a fast implementation for converting the depth image, which is mainly realized through the utilization of OpenCV cv::Mat class and continuous memory read/write operations. Compared with OpenCV official implementation, our approach allows additional channels like RGB or inverse depth, and supports the conversion from cv::Mat to the PointCloud2 message format in ROS. Additionally, it can be convert to PCL point clouds with further customization. This enhanced flexibility and efficiency make our dataset well-suited for real-time SLAM applications.

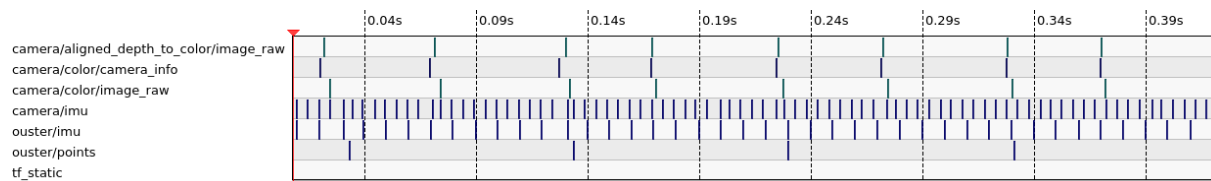


Figure 6. Rosbag information in rqt bag

The sensor data is recorded in ROS bag format using the official ROS drivers with their default synchronization schemes enabled. For the RealSense camera, IMU data is first aligned to the image messages by matching the closest timestamps. Since the point cloud and images are captured by two different sensors, their timestamps are not perfectly matched, resulting in some misalignment, as shown in Figure 6. However, the RGB and depth images are synchronized in practice when parsing the data. It is just caused by timestamp precision. The Ouster LiDAR is set to the same time settings as the camera and is synchronized with its built-in IMU. As the LiDAR data is used as the ground truth, we did not align the LiDAR points with the image data. After processing the entire ROS bag and saving the estimated poses, the trajectories can be matched and evaluated for localization performance using public tools like EVO (Rebecq et al., 2017).

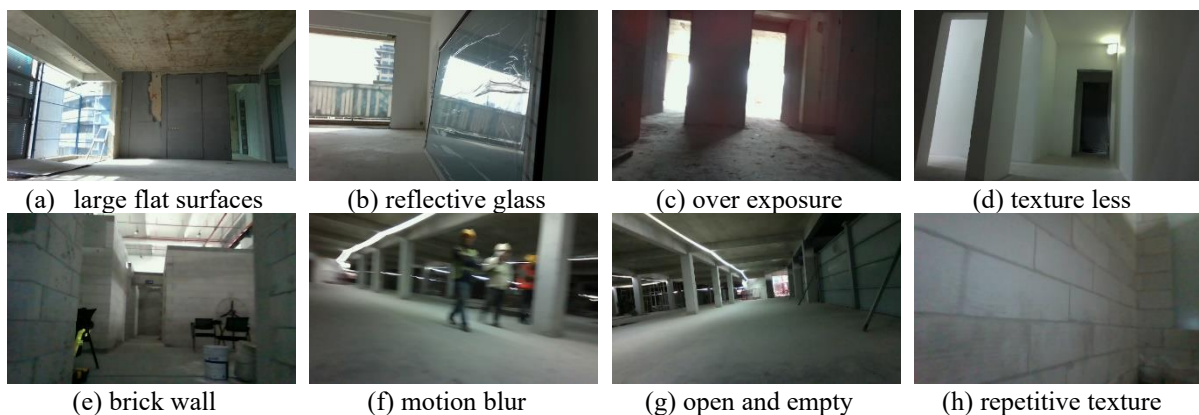


Figure 7. Representative example frames in our construction dataset

Figure 7 are some selected representative pictures from the dataset, which reflect real-world ongoing construction conditions. The motivation for collecting these scenes comes from recent SLAM research specifically focused for construction environments (Chen & Yu, 2023; Chen & Yu, 2024; Yarovoi & Cho, 2024; Yang & Cai, 2024). We capture the diversity of construction settings, including various stages of building materials and structures, such as concrete, brick, plaster, and putty, which often have low or nearly no texture. Some sequence may also involve other auxiliary materials, equipment, workers and machinery. These factors contribute to a more realistic and challenging testing environment for SLAM algorithms.

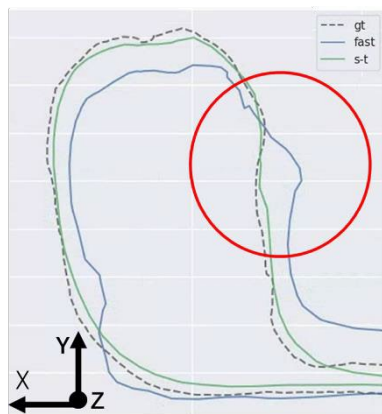
Table 3 provides a list of the selected sequences from the raw collected data. Most sequences are captured in static environments to simulate a construction robot operating independently. We have also included a dynamic sequence (Floor1) to reflect the presence of workers around the robot. Additionally, our dataset includes a closed-loop sequence (Building B1). However, it should be noted that the robot's localization must rely on real-time odometry, and the impact of loop closure correction is limited.

Table 3. Selected sequences in our dataset

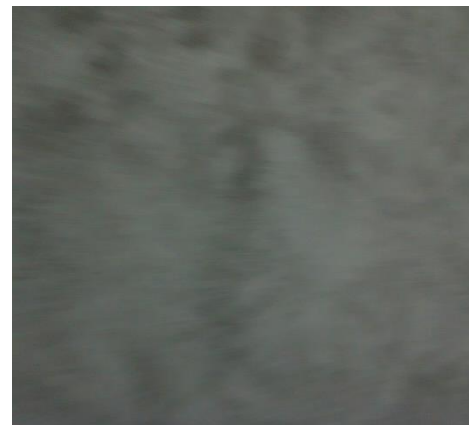
Sequence	Duration	Description
Building A1, static	85s	Texture-less, multiple rooms
Building B1, static	78s	Repetitive texture, looped trajectory
Floor14, static2	56s	Over-exposure
Floor4, static1	35s	Glass, static workers
Floor4, static3	24s	Texture-less, single room
Floor1, dynamic	55s	Motion blur, dynamic workers

3. Benchmark Results and Discussion

We evaluated several visual SLAM systems with different frontends on our dataset, as illustrated in Figure 8 (a). The "Fast" frontend refers to the fast corner detection combined with optical flow tracking, implemented by DVINS(Liu et al., 2022). VINS-MONO (Qin et al., 2018) utilizes Shi-Tomasi (S-T) corner points with optical flow tracking, but for this experiment, we selected the RGB-D version of VINS (Shan et al., 2019). ORBSLAM (Campos et al., 2021) was also tested but failed to complete the entire trajectory, as shown in Figure 2. It struggled to detect features when the robot approached a gray wall (Figure 8 (b)) closely, leading to totally drift. This demonstrates the difficulties feature matching based methods in construction environments with low texture or flat surfaces.



(a) Localization results



(b) The corresponding area to the red circle

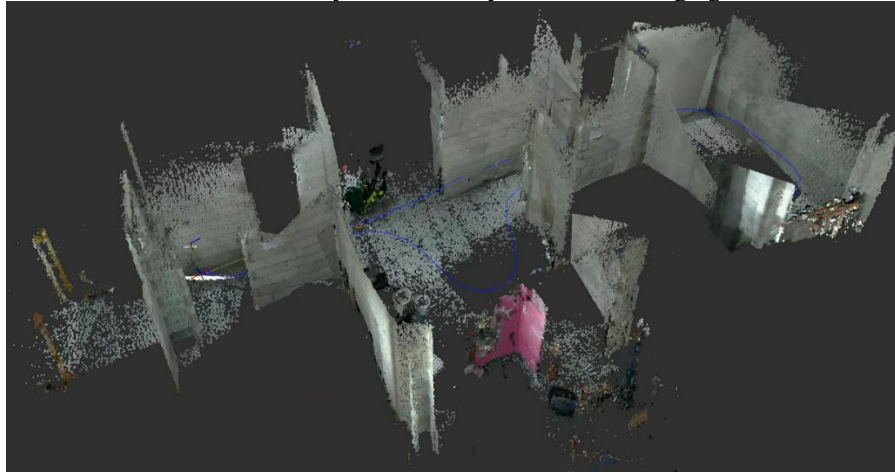
Figure 8. A comparison between several visual frontends

The mentioned visual SLAM algorithms are highly representative methods and are widely regarded as benchmarks by most researchers. However, our simple tests revealed that VINS-based improved methods demonstrate greater robustness. Therefore, we recommend using the VINS-based visual SLAM algorithm as a benchmark for our dataset. For visualization, we utilize RTABMAP (Labbé and Michaud, 2019) as the dense mapping backend, replacing the default visual odometry with VINS RGB-D (Shan et al., 2019). RTABMAP is particularly known for its versatility and modular design, making it a popular choice for both research and practical applications. In our experiment, it is used for mapping and visualization to support different visual SLAM algorithms.

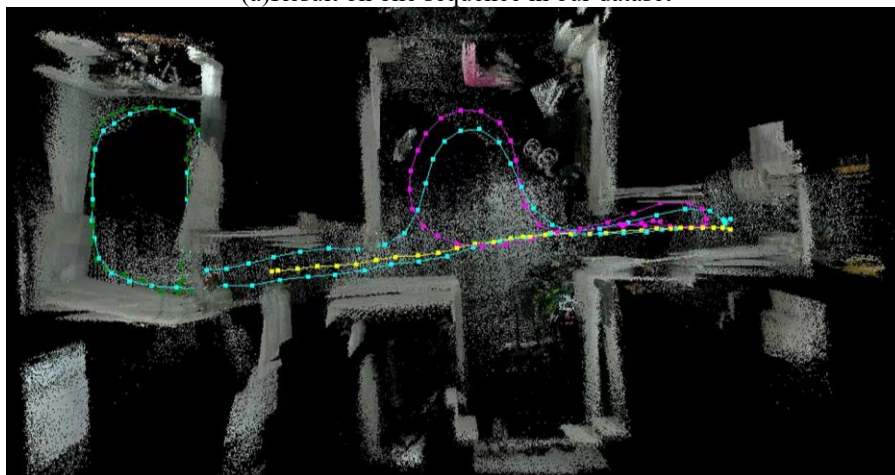
The final results of mapping and localization are shown in Figure 9. Figure 9(a) illustrates the outcome of a single run, where the thin blue line represents the estimated trajectory. Figure 9(b) shows the results from different runs in the same environment, with the point cloud map merging each sub-map across sessions. The

differently colored lines correspond to individual robot paths, and the node points indicate keyframe poses from the visual odometry. This experiment aims to provide a simple benchmark and a visualization evaluation based on the collected dataset.

The mapping results in Figure 9(a) and Figure 9(b) are not perfect, with some mismatches observed, particularly in the multiple mapping results shown in Figure 9(b). This highlights the challenges faced by current visual localization and mapping methods in real construction environments. Future work could focus on developing more robust SLAM algorithms that better leverage prior knowledge, such as 2D drawing, BIM models and structure information to enhance accuracy and reliability in these challenging construction scenarios.



(a) Result on one sequence in our dataset



(b) Results from multiple experiments within the same room

Figure 9. Mapping and localization results by combining VINS RGBD and RTABMAP

4. CONCLUSIONS

This study introduces an RGB-D visual SLAM dataset specifically collected in real construction sites, which is often overlooked in existing benchmarks. It uniquely includes the dynamic and complex nature of real construction environments, which pose significant challenges for traditional SLAM algorithms. By offering detailed environmental data and realistic settings, we hope this dataset will be a useful resource for construction automation and robotics community, and can facilitate advancements in SLAM localization systems tailored for construction sites.

ACKNOWLEDGMENTS

This work was supported by the HKUST-BDR Joint Research Institute (Grant No. OKT23EG01), the HKUST Bridge Gap Fund (Grant No. BGF.012.2022), the Collaborative Research Fund (C6044-23GF) supported by the University Grants Committee of the Government of the Hong Kong Special Administrative Region, and the 30 for 30 Scheme (3030007) funded by The Hong Kong University of Science and Technology.

REFERENCES

Campos, C., Elvira, R., Rodríguez, J. J. G., Montiel, J. M., & Tardós, J. D. (2021). ORB-SLAM3: An Accurate

- Open-Source Library for Visual, Visual-Inertial, and Multimap SLAM. *IEEE Transactions on Robotics*.
- Chen, X., & Yu, Y. (2023). HLE-SLAM: SLAM for overexposed construction environment. In ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction (Vol. 40, pp. 585-588). IAARC Publications.
- Chen, X., & Yu, Y. (2024). An unsupervised low-light image enhancement method for improving V-SLAM localization in uneven low-light construction sites. *Automation in Construction*, 162, 105404.
- Geiger, A., Lenz, P., Stiller, C., & Urtasun, R. (2013). Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research*, 32(11), 1231–1237.
- Gu, P., & Meng, Z. (2023). S-VIO: Exploiting Structural Constraints for RGB-D Visual Inertial Odometry. *IEEE Robotics and Automation Letters*, 8(6), 3542–3549.
- Helmberger, M., Morin, K., Berner, B., Kumar, N., Cioffi, G., & Scaramuzza, D. (2022). The Hilti SLAM Challenge Dataset, *IEEE Robotics and Automation Letters*, 7(3), 7518–7525.
- Labbe, M., & Michaud, F. (2019). RTAB-Map as an open-source lidar and visual simultaneous localization and mapping library for large-scale and long-term online operation. *Journal of Field Robotics*, 36(2), 416–446.
- LEE, D., Jung, M., & Kim, A. (2023). ConPR: Ongoing construction site dataset for place recognition. *IROS 2023 workshop on closing the loop on localization: What are we localizing for, and how does that shape everything we should do?*
- Li, P., Cai, K., Saputra, M. R. U., Dai, Z., & Lu, C. X. (2022, October). Odombeyonvision: An indoor multi-modal multi-platform odometry dataset beyond the visible spectrum. In 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 3845-3850). IEEE.
- Liu, J., Li, X., Liu, Y., & Chen, H. (2022). RGB-D Inertial Odometry for a Resource-Restricted Robot in Dynamic Environments. *IEEE Robotics and Automation Letters*, 7(4), 9573–9580.
- Qingqing, L., Xianjia, Y., Queralta, J. P., & Westerlund, T. (2022). Multi-Modal Lidar Dataset for Benchmarking General-Purpose Localization and Mapping Algorithms. *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 3837–3844.
- Qin, T., Li, P., & Shen, S. (2018). VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator. *IEEE Transactions on Robotics*, 34(4), 1004–1020.
- Rebecq, H., Horstschaefer, T., Gallego, G., & Scaramuzza, D. (2017). EVO: A Geometric Approach to Event-Based 6-DOF Parallel Tracking and Mapping in Real Time. *IEEE Robotics and Automation Letters*, 2(2), 593–600.
- Shan, Z., Li, R., & Schwertfeger, S. (2019). RGBD-Inertial Trajectory Estimation and Mapping for Ground Robots. *Sensors*, 19(10), Article 10.
- Shi, X., Li, D., Zhao, P., Tian, Q., Tian, Y., Long, Q., ... & She, Q. (2020, May). Are we ready for service robots? the openloris-scene datasets for lifelong slam. In 2020 IEEE international conference on robotics and automation (ICRA) (pp. 3139-3145). IEEE.
- Sturm, J., Engelhard, N., Endres, F., Burgard, W., & Cremers, D. (2012). A benchmark for the evaluation of RGB-D SLAM systems. 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, 573–580.
- Trzeciak, M., Pluta, K., Fathy, Y., Alcalde, L., Chee, S., Bromley, A., Brilakis, I., & Alliez, P. (2023). ConSLAM: Construction Data Set for SLAM, *Journal of Computing in Civil Engineering*, 37(3), 04023009.
- Wang, H., Wang, C., & Xie, L. (2021). Lightweight 3-D Localization and Mapping for Solid-State LiDAR. *IEEE Robotics and Automation Letters*, 6(2), 1801–1807.
- Xu, J., Li, R., Huang, S., Zhao, X., Qiu, S., Chen, Z., & Zhao, L. (2023). R2DIO: A Robust and Real-Time Depth-Inertial Odometry Leveraging Multi-Modal Constraints for Challenging Environments. *IEEE Transactions on Instrumentation and Measurement*, 1–1.
- Xu, W., Cai, Y., He, D., Lin, J., & Zhang, F. (2022). FAST-LIO2: Fast Direct LiDAR-Inertial Odometry. *IEEE Transactions on Robotics*, 38(4), 2053–2073.
- Yarovoi, A., & Cho, Y. K. (2024). Review of simultaneous localization and mapping (SLAM) for construction robotics applications. *Automation in Construction*, 162, 105344.
- Yang, L., & Cai, H. (2024). Enhanced visual SLAM for construction robots by efficient integration of dynamic object segmentation and scene semantics. *Advanced Engineering Informatics*, 59, 102313.
- Yin, J., Li, A., Li, T., Yu, W., & Zou, D. (2021). M2dgr: A multi-sensor and multi-scenario slam dataset for ground robots. *IEEE Robotics and Automation Letters*, 7(2), 2266-2273.

SEMANTIC SEGMENTATION OF MEP WITH DEEP LEARNING AND SYNTHETIC POINT CLOUDS

Hongzhe Yue¹, Qian Wang²

1) Ph.D. Candidate, School of Civil Engineering, Southeast University, Nanjing, China. Email: yuehongzhe@seu.edu.cn

2) Professor, School of Civil Engineering, Southeast University, Nanjing, China. Email: qianwang@seu.edu.cn

Abstract: Semantic segmentation of point clouds with deep learning (DL) heavily relies on large datasets for training. However, there is a significant scarcity of datasets for Mechanical, Electrical, and Plumbing (MEP) scenes. To address this gap, this study proposes a method namely the ray-based laser scanning and intersection algorithm (RBLSIA) for automatically generating synthetic point clouds for MEP from Building Information Modeling (BIM) models. Based on RBLSIA, this study conducted totally 25 groups of comparative experiments, investigating the semantic segmentation performance on MEP scenes with different training datasets and different generation approaches for synthetic point clouds. The results show that: 1) the mean Intersection over Union (mIoU) with synthetic point clouds produced by the RBLSIA method is on average 3.32% higher than that by the uniform sampling method; 2) increasing the number of synthetic point cloud samples further improved both the OA and mIoU for semantic segmentation, even surpassing the training accuracy achieved with real point clouds.

Keywords: Synthetic point cloud; MEP; Deep learning; Semantic segmentation

1. INTRODUCTION

Building Information Modeling (BIM) has become a crucial tool in managing and maintaining building facilities. BIM provides comprehensive semantic and geometric information, which is essential for effective facility management. As-built BIM models, which capture the actual state of a facility, play a key role in improving maintenance and monitoring processes, thereby enhancing operational efficiency (Xiong et al., 2013).

Mechanical, Electrical, and Plumbing (MEP) systems are critical components of a facility. These systems require regular maintenance and renovation, particularly in older buildings. BIM models, which offer detailed information on MEP components, are highly valuable for these tasks. However, many existing MEP systems, especially in older facilities, do not have corresponding BIM models. Moreover, as-built conditions often differ from the original design plans, making frequent site surveys necessary to keep BIM models updated.

In recent years, three-dimensional (3D) laser scanning, especially terrestrial laser scanning (TLS), has demonstrated high accuracy and efficiency in capturing detailed geometric data, making it an effective method for 3D reconstruction (Wang et al., 2022). A 3D laser scanner emits infrared laser beams and measures the distance to objects based on the reflected signals. By capturing distances from multiple angles, the scanner generates a complete set of point cloud data that accurately represents the 3D geometry of the scanned environment. However, point cloud data, while rich in geometric detail, often lacks semantic information, necessitating further processing to identify and classify objects within the data. Traditional object recognition methods, such as those based on geometric shape descriptors, hard-coded knowledge, supervised learning, or BIM-vs.-Scan (Wang et al., 2020), rely on predefined human knowledge and extensive parameter tuning. These methods become inadequate when dealing with MEP components that have extensive occlusions and irregular shapes.

Over the past few years, deep learning (DL) algorithms such as PointNet and PointNet++ have shown significant potential in processing point clouds for tasks such as classification and segmentation (Yue et al., 2024). DL algorithms significantly reduce reliance on manually designed features by autonomously learning and extracting complex data features from large labeled training datasets, achieving better performance (Yin et al., 2021; Zhang et al., 2022). However, DL often requires large amounts of sample data for training. Compared to images, point cloud data acquisition relies on scanning at multiple stations with a laser scanner, which is extremely time-consuming and labor-intensive. Additionally, labeling point cloud data is also more challenging than labeling image data. Therefore, obtaining a large amount of labeled point cloud data is crucial for DL-based model reconstruction.

Recently, BIM-based synthetic point cloud generation has been employed to automatically produce labeled point cloud data from BIM models, providing a more cost-efficient way to acquire data for DL model training (Tang et al., 2022; Tang et al., 2023). Most current research focuses on generating synthetic point clouds for building indoor scenes and bridges, achieving improvements in semantic segmentation accuracy (Lamas et al., 2024; Tang et al., 2023; Won et al., 2020; Zhai et al., 2022; Zhang and Zou, 2023). However, there has been no research on generating synthetic point clouds specifically for MEP scenes. It remains unclear whether synthetic point clouds for MEP systems can improve the accuracy of DL-based semantic segmentation. Traditional methods of synthetic point cloud generation based on uniform sampling may not accurately reflect real-world occlusions in point clouds, underscoring the need for a tailored approach to generating synthetic point clouds for MEP systems.

To address this gap, this study proposes a method to generate synthetic point clouds from BIM models

specifically for MEP scenes by simulating the real laser scanning process. To validate the proposed method, this study compares the semantic segmentation performance on MEP scenes with different approaches for generating synthetic point clouds, and different training datasets.

2. METHOD

2.1 Proposed Method for Generating Synthetic Point Clouds for MEP Scenes

This study introduces a method for generating synthetic point clouds for MEP scenes using BIM models. The overall workflow is illustrated in Figure 1. Initially, the BIM model is converted into the OBJ format, which is then processed using the Trimesh library in Python. Following this, synthetic point clouds are generated using the Ray-based Laser Scanning and Intersection Algorithm (RBLISA). The details of RBLISA involve three key components: Simulating Laser Scanner Positions, Setting Scanning Parameters and Error Simulation, and Intersection with Triangle Mesh.

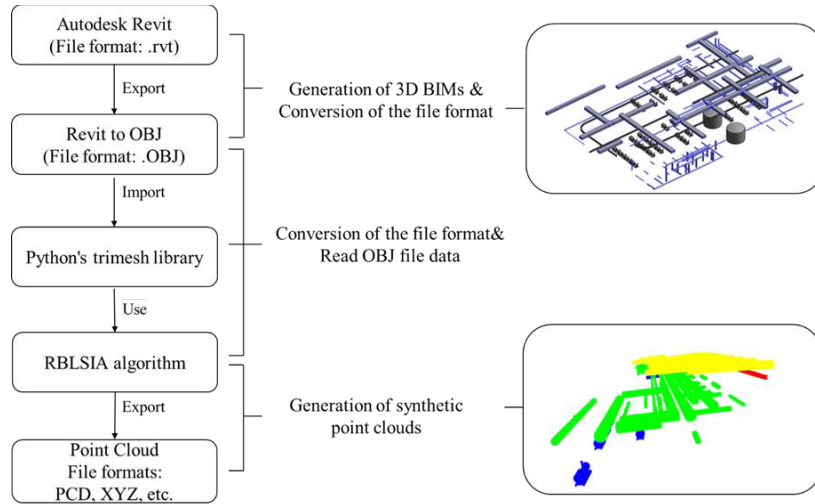


Figure 1. Process of generating synthetic point clouds for MEP scene from BIM.

Simulating Laser Scanner Positions: The RBLISA begins by simulating the positions of the laser scanner based on human-accessible areas. The scanner positions are chosen to avoid areas occupied by equipment such as pumps and tanks, while ensuring sufficient spacing between each scanning position.

Setting Scanning Parameters and Error Simulation: Next, the scanning parameters, such as vertical and horizontal resolution, are determined based on the actual specifications of the laser scanner. To simulate real-world conditions, measurement errors are also introduced. Synthetic laser rays are then generated accordingly.

Intersection with Triangle Mesh: The core of the RBLISA is the Ray Intersections with Triangle Mesh (RITM) algorithm. This algorithm checks for intersections between each laser ray and the scene's triangle mesh. The closest intersection points are recorded and used to generate the synthetic point clouds. These points include relevant data such as material, color, and category. The pseudo-code for the RITM this method is shown in Algorithm 1.

Algorithm 1: Ray Intersections with Triangle Mesh algorithm

-
- $intersection_points \leftarrow []$
 - For each $direction \in D$:
 - $d_{min} \leftarrow \infty$
 - For each $mesh_part \in scene.geometry$:
 - $h \leftarrow mesh_part.ray.intersects_location(direction)$
 - If h :
 - $d \leftarrow \|h - P_s\|$
 - If $d < d_{min}$:
 - $d_{min} \leftarrow d$
 - $closest_intersection \leftarrow (h, material, color, category)$
 - If $closest_intersection$:
 - $intersection_points.append(closest_intersection)$
 - return $intersection_points$
-

Figure 2 shows an example of the synthetic point cloud generation process using the RBLSIA method. Firstly, the virtual laser scanner position is determined, as indicated by the green mark in Figure 2 (a). Then, the simulated laser rays are generated and the intersections between laser rays and the model's triangular triangle mesh are calculated, as shown in Figure 2 (b). Eventually, the synthetic point clouds of the MEP scene are generated, and the 3D view and plane view are shown in Figure 2 (c) and Figure 2 (d), respectively.

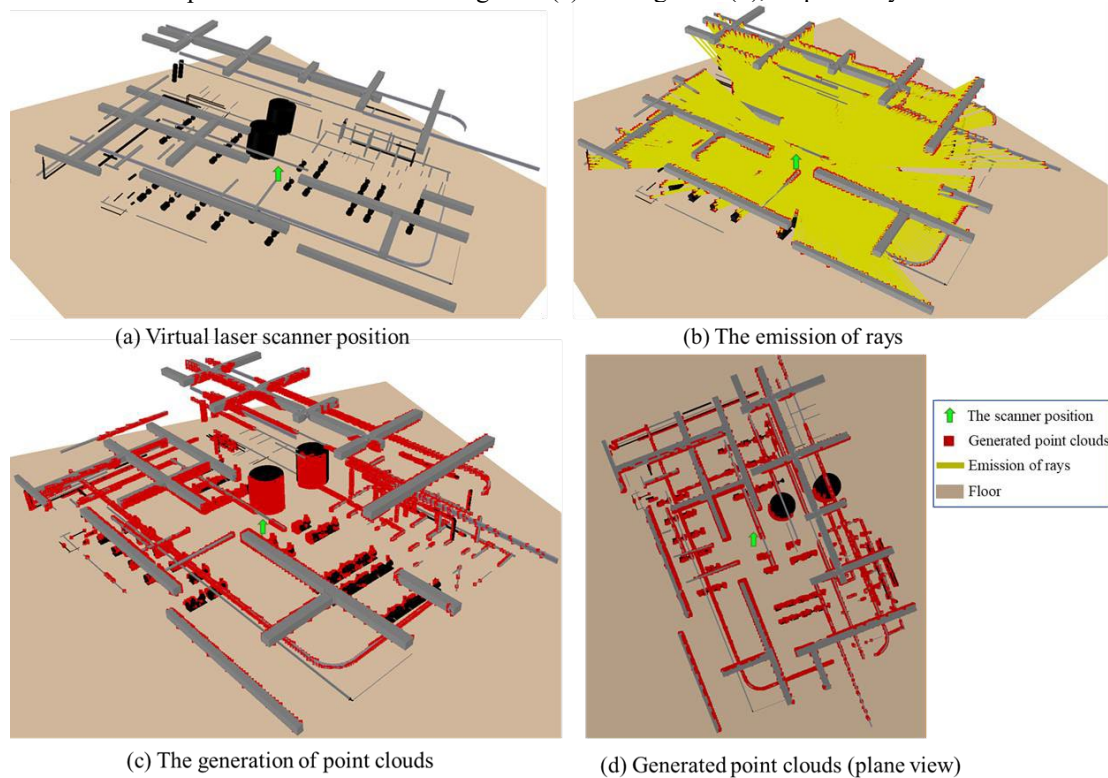


Figure 2. Synthetic point cloud generation process using the proposed RBLSIA method.

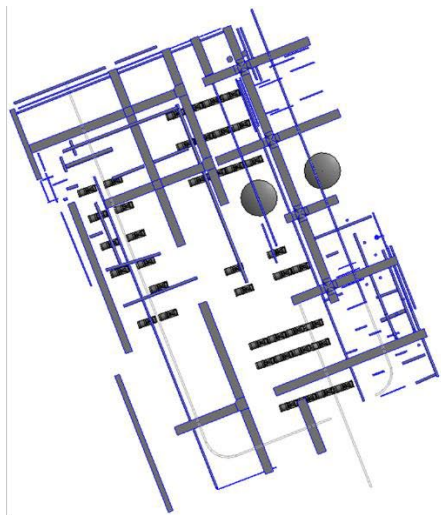
2.2 Data Preparation

The PSNET5 dataset (Yin et al., 2023) is used in this study to validate the proposed approach. The initial scene includes 500 million raw data points, covering approximately 3700 square meters, and consists of four areas with different scenes, as shown in Table 1.

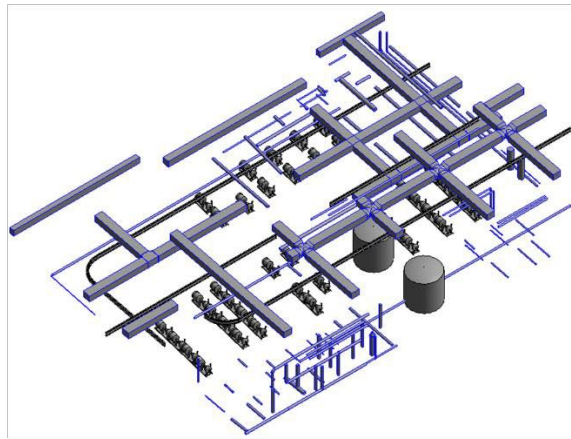
Table 1. The number of points of each semantic category of the PSNET5 dataset.

Area	Scene	ibeam	pipe	pump	rbeam	tank	Total
Area1	CH	359,610	22,799,721	1,478,449	5,821,433	0	30,459,213
Area2	OSCG	467,782	5,712,622	5,712,622	2,488,327	5,455,825	14,404,173
Area3	SPH	3,023,003	8,408,907	2,455,212	9,055,639	1,523,331	24,466,092
Area4	WRT2	656,071	3,052,706	836,554	7,040,629	0	11,585,960
Total		4,506,466	39,973,956	5,049,832	24,406,028	6,979,156	80,915,438

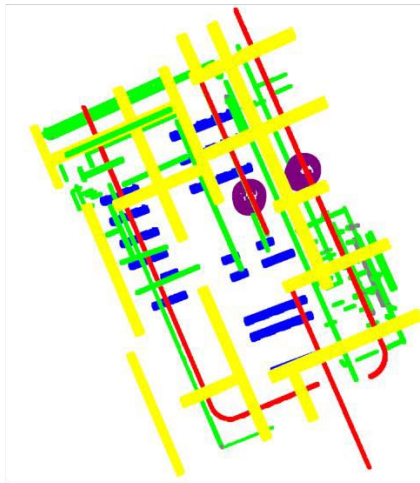
In this study, synthetic point clouds are generated from the raw point clouds of Area1, Area2, and Area3. The raw point clouds are imported into Revit, where MEP components are modeled to create BIM models for each area. Subsequently, synthetic point clouds are generated from these BIM models. Figure 3 illustrates the BIM model and generated point clouds for Area3.



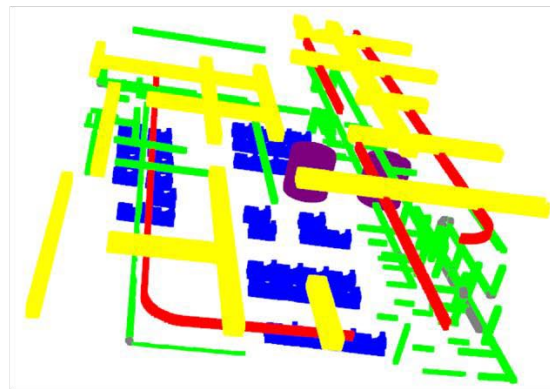
(a) Top view of the BIM model of Area3



(b) 3D view of the BIM model of Area3



(c) Top view of the generated point clouds of Area3



(d) 3D view of the generated point clouds of Area3

Figure 3. The BIM model and generated point clouds of Area3.

2.3 Experimental Settings

This study employed ResPointNet++ as the DL algorithm (Yin et al., 2023). Each experiment was conducted over 100 epochs, utilizing the SGD optimizer with an initial learning rate of 0.01 and cross-entropy as the loss function. All experiments were performed on a single NVIDIA GTX 4090 GPU, with each set taking approximately 12 hours to complete. Due to the inconsistent color standards and varying reflections in MEP components, the input channels were limited to XYZ coordinates, excluding color information.

To evaluate the performance of DL models, three metrics were used: Overall Accuracy (OA), Intersection over Union (IoU), and mean IoU (mIoU). OA provides an overall measure of accuracy by indicating the proportion of correctly predicted points relative to the total. IoU assesses segmentation quality for each class by measuring the overlap between predicted and actual segments. In contrast, mIoU offers a balanced evaluation across all N classes, highlighting the model's consistent performance across different object types. The formulas for these metrics are presented in Equations (1), (2), and (3).

$$OA = \frac{\text{Number of correctly classified points}}{\text{Total number of points}} \quad (1)$$

$$IoU = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive} + \text{False Negative}} \quad (2)$$

$$mIoU = \frac{1}{N} \sum_{i=1}^n IoU_i \quad (3)$$

2.4 Experiment Design

The validation experiments include three parts. First, the performance of the uniform sampling method and the RBLSIA sampling method synthetic point clouds was compared. Second, the performance of semantic segmentation using with real point clouds and synthetic point clouds as training data was compared. Finally, the effect of increasing the number of synthetic point cloud samples on the semantic segmentation results was examined.

3. RESULTS

3.1 Comparison of Different Generation Methods for Synthetic Point Clouds

This section aimed to compare the performance of two different generation methods for synthetic point clouds: uniform sampling method and RBLSIA sampling method. As shown in Table 2, five comparison groups (CGs) (A to E) were conducted with different training sets and test sets, where some used purely synthetic data for training while some others used a mix of real and synthetic data for training.

According to the comparison results, it is clear that using synthetic point clouds generated by the RBLSIA method could always yield better model performance than using data from the uniform sampling method. On average, the RBLSIA method could achieve an improvement of 2.13% and 3.32% in OA and mIoU, respectively. Among all IoU metrics, the tank category showed the most significant improvement of 38.38%, followed by pump of 5.47%, rbeam of 2.41%, pipe of 1.55%, and ibeam of -1.76%. The substantial improvements in the tank and pump categories may be attributed to these being minority classes. For minority classes, more realistic synthetic point cloud features help in learning more accurate characteristics, thereby improving the precision of these categories. The average performance of the RBLSIA method on ibeam was slightly worse than that of uniform sampling. This is possibly because ibeam is relatively slender, making it difficult for the RBLSIA method to include its complete features, whereas uniform sampling is more likely to include complete features. Overall, the results demonstrate that the RBLSIA sampling method provides better model performance compared to uniform sampling method.

Table 2. Comparison between uniform sampled synthetic point clouds and RBLSIA sampled synthetic point clouds

CG	Exp. No.	Training set	Test set	OA (%)	mIoU (%)	IoU (%)				
						ibeam	pipe	pump	rbeam	tank
A	1	Area2 RBLSIA synthetic	Area3	64.62	45.46	56.45	48.74	0.02	69.23	64.62
	2	Area2 uniform synthetic	Area3	63.19	42.85	66.73	54.08	0	67.18	26.24
B	3	Area3 RBLSIA synthetic	Area4	85.52	68.05	98.30	50.58	26.14	97.19	/
	4	Area3 uniform synthetic	Area4	83.64	63.56	97.76	45.53	21.32	89.64	/
C	5	Area1 RBLSIA synthetic, Area3 real	Area4	95.22	89.10	95.97	85.22	78.76	96.46	/
	6	Area1 uniform synthetic, Area3 real	Area4	93.28	83.09	97.05	75.26	63.92	96.15	/
D	7	Area1+Area3 RBLSIA synthetic	Area4	96.30	90.06	95.23	84.89	81.41	94.79	/
	8	Area1+Area3 uniform synthetic	Area4	93.76	89.64	93.23	88.65	79.91	96.78	/
E	9	Area2+Area3 RBLSIA synthetic	Area1	81.69	74.38	68.67	87.04	74.03	67.79	/
	10	Area2+Area3 uniform synthetic	Area1	78.81	71.33	68.63	85.18	67.83	63.67	/

3.2 Comparison of Training with Real and Synthetic Point Clouds

This section compared the performance of training with real point clouds and synthetic point clouds of the same areas. As shown in Table 3, a total of seven CGs (F to L) were carried out. In each CG, DL models were trained separately with real point clouds and synthetic point clouds of certain areas, and the two trained models

were tested on the real point clouds of another area.

According to the results, in all the CG except CG-L, training with synthetic point clouds resulted in lower OA and mIoU compared to training with real point clouds. On average, OA was decreased by 6.35% and mIoU was decreased by 10.84% when training with synthetic point clouds. Therefore, in most cases, the training effectiveness of synthetic point clouds was inferior to that of real point clouds, reflecting the differences between synthetic and real point clouds.

Table 3. Comparison between training with synthetic point clouds and real point clouds

CG	Exp. No.	Training set	Test set	OA	mIoU (%)	IoU (%)				
						ibeam	pipe	pump	rbeam	tank
F	11	Area3 real	Area4	92.85	80.31	97.06	73.23	53.89	97.05	/
	3	Area3 synthetic	Area4	85.52	68.05	98.30	50.58	26.13	97.19	/
G	12	Area3 real	Area1	90.70	90.98	79.76	96.87	97.54	89.77	/
	13	Area3 synthetic	Area1	77.21	49.74	45.58	74.11	28.12	51.15	/
H	14	Area1 real	Area4	91.55	69.30	39.13	79.56	66.40	92.11	/
	15	Area1 synthetic	Area4	82.49	56.15	39.29	47.73	50.25	87.34	/
I	16	Area1+Area2+Area3 real	Area4	96.27	95.25	97.01	91.96	95.06	96.98	/
	17	Area1+Area2+Area3 synthetic	Area4	88.22	75.49	64.72	79.43	61.61	96.19	/
J	18	Area1 real	Area3	77.00	64.82	73.85	56.59	49.75	79.08	/
	19	Area1 synthetic	Area3	70.99	56.11	56.58	41.89	50.09	75.87	/
K	20	Area2 real	Area4	71.57	35.37	42.63	23.01	0.00	75.64	/
	21	Area2 synthetic	Area4	51.48	29.17	45.78	28.02	0.00	42.89	/
L	22	Area2 real	Area3	45.06	20.06	17.88	0.00	0.00	82.42	0.00
	1	Area2 synthetic	Area3	64.62	45.47	56.46	48.74	0.00	69.23	52.92

3.3 Comparison of Training with Real and Synthetic Point Clouds with Different Data Sizes

Results in the previous section showed that, with the same training data size, training with synthetic point clouds would achieve worse performance than training with real point clouds. However, one major advantage of using synthetic point clouds is its lower cost of generating labeled dataset. Therefore, this section aimed to examine whether it is possible to improve the model performance by increasing the training data size when using synthetic point clouds.

As shown in Table 4, totally five CGs (M to Q) were implemented. In each CG, the training set contained only one area when training with real point clouds, while the training set contained one or two areas when training with synthetic point clouds. In all the five CGs, it is found that training with synthetic point clouds of two areas could generate better performance than training with synthetic point clouds of only one area, with an average improvement of 13.91% for OA and 24.58% for mIoU. Hence, it is concluded that increasing the training data size could effectively improve the model performance when training with synthetic point clouds.

Among all the five CGs, three of them (CG-M, CG-N, and CG-O) showed that training with synthetic point clouds of two areas could yield better performance than training with real point clouds of only one area. This indicated that, with a larger training data size, purely using synthetic point clouds was able to achieve better performance than using real point clouds.

Table 4. Comparison between uniform sampled synthetic point clouds and RBLSIA sampled synthetic point clouds

CG	Exp. No.	Training set	Test set	OA(%)	mIoU (%)	IoU (%)				
						ibeam	pipe	pump	rbeam	tank
M	14	Area1 real	Area4	91.55	69.30	39.13	79.56	66.4	92.11	/
	15	Area1 synthetic	Area4	82.49	56.15	39.29	47.73	50.25	87.34	/
	23	Area1+Area3 synthetic	Area4	96.30	90.06	95.23	84.89	81.41	94.79	/
N	11	Area3 real	Area4	92.85	80.31	97.06	73.23	53.89	97.05	/
	3	Area3 synthetic	Area4	85.52	68.05	98.30	50.58	26.13	97.19	/

	23	Area1+Area3 synthetic	Area4	96.30	90.06	95.23	84.89	81.41	94.79	/
O	20	Area2 real	Area4	71.57	35.37	42.63	23.01	0	75.64	/
	21	Area2 synthetic	Area4	51.48	29.17	45.78	28.02	0.00	42.89	/
	24	Area2+Area3 synthetic	Area4	89.52	70.84	97.11	56.83	30.94	98.47	/
P	18	Area1 real	Area3	77.00	64.82	73.85	56.59	49.75	79.08	/
	19	Area1 synthetic	Area3	70.99	56.11	56.58	41.89	50.09	75.87	/
	25	Area1+Area2 synthetic	Area3	73.42	56.77	46.75	55.13	53.13	72.09	56.77
Q	12	Area3 real	Area1	90.70	90.98	79.76	96.87	97.54	89.77	/
	13	Area3 synthetic	Area1	77.21	49.74	45.58	74.11	28.12	51.15	/
	9	Area2+Area3 synthetic	Area1	81.69	74.38	68.67	87.04	74.03	67.79	/

4. DISCUSSION

The proposed method provides a cost-efficient approach to generating training point cloud data and enhancing semantic segmentation for MEP scenes, which can play an important role in the maintenance and renovation of existing MEP systems in not only industrial plants but also other types of facilities. Without any real point cloud data, simply using synthetic point clouds generated from BIM models can obtain semantic segmentation models with certain accuracies. Specifically, according to the experimental results, training with synthetic point clouds of only one area could achieve an OA of 50% to 86% and an mIoU of 30% to 68% (refer to Table 3), and training with synthetic point clouds of two areas could further achieve an OA of 73% to 96% and an mIoU of 56% to 90% (refer to Table 4).

5. CONCLUSIONS

This study proposed the RBLSIA to automatically generate synthetic point clouds for MEP from BIM models. A total of 25 comparative experiments were carried out to evaluate the semantic segmentation performance across different methods for synthetic point cloud generation, and different training datasets. The experimental results indicated that: 1) the mIoU with synthetic point clouds produced by the RBLSIA method is on average 3.32% higher than that by the uniform sampling method; 2) increasing the number of synthetic point cloud samples further improved both the OA and mIoU for semantic segmentation, even surpassing the training accuracy achieved with real point clouds.

The results also reveal a gap between synthetic and real point clouds, particularly for components with significant differences between BIM models and real objects (such as pumps). Future research should focus on reducing these discrepancies to further enhance the quality of synthetic point clouds. Additionally, future studies can expand the range of MEP components (e.g., valves and fittings) and utilize different point cloud synthesis algorithms (e.g., diffusion models) to investigate the impact of various synthesis methods and component types on semantic segmentation accuracy.

ACKNOWLEDGMENTS

This work was supported by the National Key R&D Program of China (No. 2023YFC3804300), Start-up Research Fund of Southeast University (No. RF1028623126) and Jiangsu Provincial Science and Technology Program (Hong Kong, Macao, and Taiwan Science and Technology Cooperation Project), Grant No. BZ2024058.

REFERENCES

- Lamas, D., Justo, A., Soilán, M., Riveiro, B., 2024. Automated production of synthetic point clouds of truss bridges for semantic and instance segmentation using deep learning models. *Automation in Construction* 158, 105176.
- Tang, S., Li, X., Zheng, X., Wu, B., Wang, W., Zhang, Y., 2022. BIM generation from 3D point clouds by combining 3D deep learning and improved morphological approach. *Automation in Construction* 141, 104422.
- Tang, S.J., Huang, H.S., Zhang, Y.J., Yao, M.M., Li, X.M., Xie, L.F., Wang, W.X., 2023. Skeleton-guided generation of synthetic noisy point clouds from as-built BIM to improve indoor scene understanding.

- Automation in Construction* 156, 105076.
- Wang, Q., Li, J.J., Tang, X.Y., Zhang, X.C., 2022. How data quality affects model quality in scan-to-BIM: A case study of MEP scenes. *Automation in Construction* 144, 104598.
- Wang, Q., Tan, Y., Mei, Z.Y., 2020. Computational Methods of Acquisition and Processing of 3D Point Cloud Data for Construction Applications. *Archives of Computational Methods in Engineering* 27, 479-499.
- Won, J., Czerniawski, T., Leite, F., 2020. Semantic segmentation of point clouds of building interiors with deep learning: Augmenting training datasets with synthetic BIM-based point clouds. *Automation in Construction* 113, 103144.
- Xiong, X.H., Adan, A., Akinci, B., Huber, D., 2013. Automatic creation of semantically rich 3D building models from laser scanner data. *Automation in Construction* 31, 325-337.
- Yin, C., Wang, B., Gan, V.J., Wang, M., Cheng, J.C., 2021. Automated semantic segmentation of industrial point clouds using ResPointNet++. *Automation in Construction* 130, 103874.
- Yin, C., Yang, B., Cheng, J.C.P., Gan, V.J.L., Wang, B.Y., Yang, J., 2023. Label-efficient semantic segmentation of large-scale industrial point clouds using weakly supervised learning. *Automation in Construction* 148, 104757.
- Yue, H., Wang, Q., Zhao, H., Zeng, N., Tan, Y., 2024. Deep learning applications for point clouds in the construction industry. *Automation in Construction* 168, 105769.
- Zhai, R.M., Zou, J.G., He, Y.F., Meng, L.Y., 2022. BIM-driven data augmentation method for semantic segmentation in superpoint-based deep learning network. *Automation in Construction* 140, 104373.
- Zhang, H.X., Zou, Z.B., 2023. Quality assurance for building components through point cloud segmentation leveraging synthetic data. *Automation in Construction* 155, 105045.
- Zhang, Z., Ji, A., Wang, K., Zhang, L., 2022. UnrollingNet: An attention-based deep learning approach for the segmentation of large-scale point clouds of tunnels. *Automation in Construction* 142, 104456.

HTUNNEL-HLS: A 3D POINT CLOUD DATASET FOR BENCHMARKING CONSTRUCTION HIGHWAY TUNNEL SEGMENTATION METHODS

Li-zhuang Cui¹, Lei Kou², Hanming Zhang³, Yusen Wang⁴, Feng Guo⁵, and Jian Liu^{6,*}

- 1) Ph.D. Candidate, School of QILU Transportation, Shandong University, Jinan, China. Email: cvlizhuang@163.com
- 2) Ph.D., School of QILU Transportation, Shandong University, Jinan, China. Email: lei.kou@sdu.edu.cn
- 3) Ph.D., School of QILU Transportation, Shandong University, Jinan, China. Email: hmzhang619@sdu.edu.cn
- 4) MEng., The Hong Kong University of Science and Technology, Hong Kong, China. Email: 765236976@qq.com
- 5) Ph.D., Assoc. Prof. School of QILU Transportation, Shandong University, Jinan, China. Email: fenggg@sdu.edu.cn
- 6) Ph.D., Prof., School of QILU Transportation, Shandong University, Jinan, China. Email: liujianshanda@gmail.com

Abstract: 3D scene understanding is revolutionising tunnel engineering. However, deep learning algorithms are data-hungry, which means the application of scene understanding on tunnel engineering requires a customized point cloud dataset in the construction field. In this paper, we introduce a new point cloud dataset called HTunnel-HLS, specifically designed for construction highway tunnel environment. HTunnel-HLS aims to establish a new database for developing semantic segmentation, and importantly, construction highway tunnel scene. Besides, the dataset provides both point-level semantic labelling along with a large range of types of semantic instance labels categorized into support structures, mechanical facilities, and others. Data have been acquired by the Hand Laser Scanning (HLS) system Hovermap and contains 28 scenes, over 1.58 billion 3D points, correspond to a 9 km long tunnel section. This paper also provides the performance of several representative baseline methods. The impact of scale on model performance is analyzed from the perspective of grid size, and outlines potential future works and challenges for fully exploiting this dataset.

Keywords: LiDAR, 3D Point cloud, Semantic segmentation, Tunnel construction.

1. INTRODUCTION

Construction safety and quality are key performance indicators for mechanized tunneling, especially in light of the rising promise of unmanned construction (Rachmawati, 2022). In order to ensure safety, long term stability and quality control in modern tunneling operations, the acquisition of geotechnical information about encountered rock conditions (Yi, 2023), detailed installed support (Wang, 2023) and rate of advancement information is required. Data collection as the tunnel progresses must make use of fast and effective technology in order to surmount operational constraints (Fekete, 2010). And it is important to inspect and evaluate the structures during construction (Cui, 2024).

Point clouds, as the most direct representation of real-world 3D digitalization, have been applied to tunnel construction and management (Zhao, 2023). Point clouds acquired by laser scanning are valuable digital assets for highway tunnels, which can be used for tunnel profile control (Kim, 2019), joint surface analysis (Xu, 2022), lining quality assessment (Lato, 2014), tunnel deformation monitoring (Deng, 2024), and 3D model reconstruction (Duan, 2021). However, A prerequisite for automatic extraction of present-day kinds of tasks is the identification of individual structures from the background scene. Conventional, heuristic algorithms suffer from several drawbacks in efficiency and generalization (Chi, 2023; Xie, 2020).

The emergence of deep learning technology has already shown considerable promise for improving the efficiency and accuracy of tunnel management. However, most recent researches have explored the application of 3D deep learning networks to the point cloud segmentation task of shield tunneling (Lin, 2024; Xie, 2023; Xie, 2017) rather than drill-and-blast highway tunnel. Empirically, data collection at drill-and-blast tunnel construction sites is more challenging than conventional scanning operations. Although emerging datasets for semantic segmentation have been proposed (A, 2013; Munoz, 2009; Ye, 2020; Zolanvari, 2019; Tong, 2020), few of them provide point-wise instance annotation for construction highway tunnel scene on a large scale. At the drill-and-blast tunnel construction site, the structural complexity of elongated spaces and the procedural continuity of dynamic construction pose significant challenges to the acquisition and annotation of 3D laser point clouds. This is why it is difficult to find annotated point cloud datasets related to tunnel engineering in popular open-source datasets, especially for construction tunnels. In contrast, there is a distinct lack of annotated datasets for highway tunnel point clouds which presents a significant obstacle for 3D deep learning segmentation. Thus, specialized datasets and their corresponding benchmarks are urgently required for the development of 3D deep learning for highway tunnel point clouds.

To that end, this paper introduces a new and richly annotated tunnel point cloud dataset, 'HTunnel-HLS', to enable automated semantic segmentation using 3D deep learning. Data experiments are presented to explore the influences of grid size in data preprocessing on the 3D deep learning model performance. This study also provides corresponding benchmarks by applying typical 3D deep learning networks to the SDU-Tunnel3D dataset. Data will be acquired and processed soon.

2. NEW DATASET: HTUNNEL-HLS

2.1 Data acquisition

This study is based on a highway tunnel under construction. The highway tunnel is a three-lane road tunnel with a triple-circle curved wall lining structure, where the New Austrian Tunneling Method (NATM) is adopted. Considering factors such as tunnel structure, face stability, and geological conditions, this tunnel is constructed by the drill-and-blast method, which is specifically constructed by the bench cut method. The bench cut method involves excavating the designed profile in two stages: upper and lower sections, with the upper section excavated first followed by the lower section.

Table 1. The key technical specifications of the HLS.

Indicator item	Parameter	Indicator item	Parameter
Scan distance	0.4m~100m	LiDAR accuracy	+/- 30mm
Mapping accuracy	+/- 15mm	Angular field of view	360°×360°
Maximum data capture travelling speed	2 m/s	Data acquisition speed	Up to 3×10^5 points/sec

Observations at the tunnel construction site, Figure 1 (a), reveal narrow and confined working spaces, with tight operations of machinery and vehicles. The acquisition has been carried out by the handheld laser scanners (HLS) named Hovermap, and produced by Emesent, as shown in Figure 1 (b). Table 1 summarizes the key technical specifications of the device. The procedure used to collect our LiDAR point cloud dataset is similar to previously published work (Cui, 2023). In addition to the tunnel structural elements, each point cloud also includes cables, pipes, loose rocks, construction vehicles, and other non-structural elements, Figure 1 (c). Hence, all possible elements and noise that may affect the point cloud segmentation are considered in the present dataset. The point clouds in HTunnel-HLS are acquired by HLS and have similar point density distributions, which differs from the feature that point clouds acquired using TLS decreases with distance to the scanner.



Figure 1. (a) The tunnel construction site; (b) Tunnel 3D point cloud data is being collected using HLS on the site ; (c) Complex point cloud data of tunnel construction site (Cui, 2023).

2.2 Data annotation

To achieve rigorous model training and evaluation, the data annotation needs to be comprehensive, consistent, and accurate for the specified labels. Instead of pixel-wise or voxel-wise annotation, every single point in HTunnel-HLS is labelled individually. This exhaustive manual annotation of the ground truth dataset avoids any biases introduced by particular segmentation algorithms.

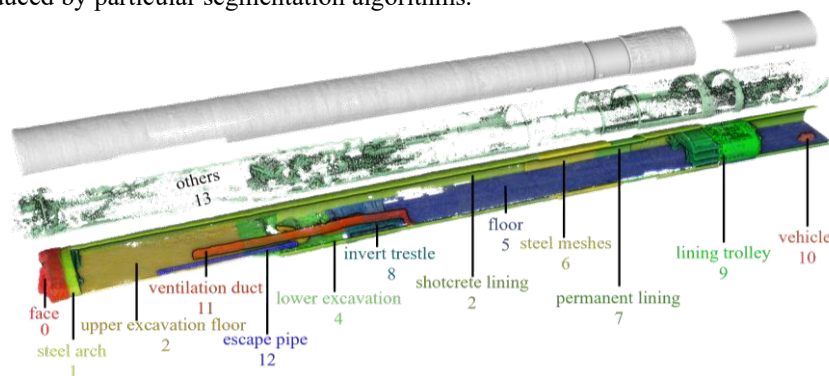


Figure 2. A representative example of labels and classes on a 3D point cloud.

In general, we categorize instances in construction tunnels into 3 classes: support structures, mechanical facilities, and others, with a total of 14 semantic labels. In the class “others” include noise points, water and electrical pipelines, ducts, loose rocks, personnel, debris, trailing shadows, etc. As objects from real-world scans are often cluttered with background and/or are partial due to occlusions. To prevent confusion, Fig. 2 illustrates a representative example of labels and classes on a 3D point cloud. A color represents a category and corresponds to a semantic label. It is important to note that by default, if an unknown class is encountered, it will be treated as 'others'.

Our annotation process does not entail labeling every point in the point cloud. Instead, we employ 3D clipping to segment all point clouds into different instances and assign a category to each segmented object, which are then individually saved in separate files. All point cloud annotations are hand-labeled by professional assessors and undergo manual cross-checking. We manually scrutinize each object, rectify inconsistent labels, and discard ambiguous, low-quality reconstruction, unlabeled, sparsely populated, and insufficient instances to form a category for training. This ensures the consistency and high quality of the annotation work. The point clouds are segmented manually using CloudCompare software. Unlike the cluttered spatial relationships of instances, the spatial distribution of instances inside tunnels is traceable.

2.3 Description of the dataset

The resulting HTunnel-HLS dataset is analysed to provide quantitative insights into its key characteristics. The dataset consists of 28 tunnel segments categorized into 14 classes, as a result, 586 objects are annotated. This extensive scale provides novel avenues for training data-intensive algorithms. The point cloud proportion of each class is presented in Fig. 3, showing that the HTunnel-HLS dataset is ‘class-imbalanced’ where label ‘ShotcreteLining’ accounts for the greatest proportion of the point clouds. While mechanical facilities e.g., escape pipe, invert trestle, and vehicle have fewer points, which is consistent with the situation of the construction sites. In addition, point clouds captured by the Hovermap scanner include not only x, y, z Cartesian coordinates but also real-world dimensions such as intensity, GPS time, return number, range, and ring etc. In order to better examine the performance and generalization of deep learning algorithms, only the original x, y, z Cartesian coordinates of each point is retained in the dataset to ensure data consistency and fair comparison. From Fig. 2, it can be observed that there is critically shape similarity between the shotcrete lining and permanent lining, as well similarity in the scale. Comparing the results for upper excavation floor and floor, shape similarity is also observed. This similarity highlights the difficulty in identifying individual segments using 3D deep learning. The training set and test set are split artificially 23:5 (respectively).

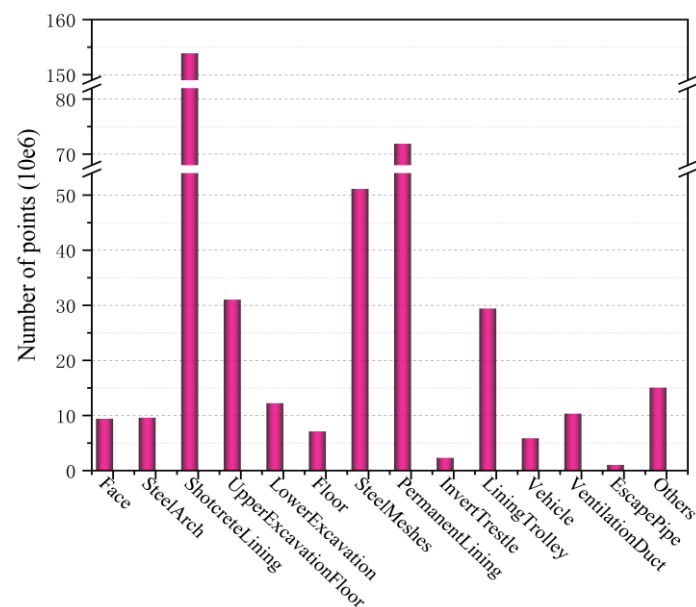


Figure 3. The point cloud proportion of each class.

3. EXPERIMENTS AND DISCUSSION

3.1 Baseline approaches for semantic segmentation

We only selected algorithms that have published results and available codes, and three popular point-based deep learning models for semantic segmentation were tested on the proposed dataset as baseline approaches: DGCNN (Wang, 2018) suggests that local geometric features are important to 3D recognition tasks, and

the model constructs graphs to extract local geometric features from local neighborhoods, and applies EdgeConv as a convolution-like operation. EdgeConv is isotropic about input features with convolutional operations on graph nodes and their edges.

RandLA-Net (Hu, 2020) utilizes random point sampling and introduces a local feature aggregation module to achieve rapid processing of large-scale point clouds while effectively preserving geometric details.

Point Transformer V3 (Wu, 2024), a stride towards overcoming the traditional trade-offs between accuracy and efficiency in point cloud processing, redefines the framework of Point Transformer. PTv3 reverts to utilizing dot-product attention. By prioritizing efficiency over the accuracy of less impactful mechanisms, PT v3 harness the power of scale, leading to enhanced performance. Simply put, the model becomes stronger by being made simpler and faster.

3.2 Evaluation metrics and configurations

For semantic segmentation, OA (overall accuracy; Equation (1)), IoU (interaction over union; Equation (2)) and mIoU (mean interaction over union; Equation (3)) are employed to evaluate model performance. In particular, mIoU presents the most reasonable evaluation for class-imbalanced datasets and is thus adopted as the priority metric.

$$OA = \frac{\sum TP_n}{\text{Total number of points}} \quad \text{Equation (1)}$$

$$IoU_n = \frac{TP_n}{TP_n + FP_n + FN_n} \quad \text{Equation (2)}$$

$$mIoU = \frac{\sum IoU_n}{N} \quad \text{Equation (3)}$$

where N is the total number of labels; n is the nth label in N; TP, FP and FN represent numbers of points of true positives, false positives, and false negatives of the predictions, respectively. OA and mIoU evaluate the overall quality of semantic segmentation; and IoU of each class measures the performance of each class.

In this study, the results were for baseline illustration purpose only, and better results could be potentially achieved with further tuning. Batch sizes were adjusted accordingly, and the remaining hyperparameters within the original model were not modified. The network structures and parameter settings of these algorithms may not be directly comparable, and parameter tuning does not guarantee the fairness of comparison. In this study, the results are for baseline illustration purpose only, and better results could be potentially achieved with further tuning. All models were trained and tested on a NVIDIA RTX 2080Ti with 11G of RAM. And batch sizes were adjusted accordingly.

3.3 Performance of baseline approaches

The results for semantic segmentation baseline approaches using HTunnel-HLS are shown in Table 2. PT V3 is on the top spot of HTunnel-HLS benchmark at the moment, and it achieved the highest OA, mACC, and mIoU among the tested baseline algorithms. DGCNN performed the worst in terms of OA, mACC, and mIoU in our dataset. And PT V3 has a significant performance improvement in the segmentation of these 14 types of objects. Since DGCNN uses KNN for construction of graphs to capture local features, it may not perform well in this dataset with varying point density and Class Imbalance Problem. The IoU of all three models is relatively low in steel arch class. It is worth mentioning that RandLA-Net is the first to complete the entire training process of 100 epochs, PT V3 is the second, and DGCNN takes the longest time.

Table 2. Semantic segmentation results of different methods (%).

Methods	OA	mIoU	mACC	0	1	2	3	4	5	6	7	8	9	10	11	12	13
DGCNN	0.8221	0.5867	0.7066	0.3637	0.3207	0.7193	0.7965	0.5725	0.8788	0.7652	0.5724	0.5169	0.7601	0.6183	0.6699	0.3692	0.2897
RandLA-Net	0.8397	0.7154	0.8520	0.5709	0.3603	0.7160	0.5785	0.6664	0.8311	0.9201	0.7157	0.7985	0.9324	0.7857	0.8635	0.8771	0.3990
PT V3	0.9700	0.9099	0.9481	0.9403	0.7615	0.9651	0.9184	0.9089	0.9586	0.9558	0.9852	0.9501	0.9582	0.8697	0.9789	0.9534	0.6345

4. DISCUSSION

As mentioned in Point Transformer V3, model performance is more profoundly influenced by scale than by complex design intricacies. The scale of the tunnel point cloud dataset constructed in this paper is significantly larger than that of indoor scenes and also exceeds that of common road scenes in terms of both the number and density of points. It is both common and necessary to the input point clouds preprocessing before deep learning

training and evaluation to improve computational efficiency, whilst also minimizing the effect on accuracy. In general, the density of the input point cloud is reduced through sampling, and memory consumption is reduced. Appropriate sampling resolution is very important for the test results. For public data set S3DIS (indoor scene) (Armeni, 2016), the sampling resolution is 0.04 by default, and for semantic3d (Hackel, 2017) and SemanticKITTI (outdoor street scene) (Behley, 2019), the sampling resolution is 0.06 by default. When the input size is kept constant, the lower input sampling resolution represents a larger receptive field, higher performance and faster training efficiency. Of course, the corresponding price to be paid is the possible loss of local information. Therefore, it is necessary to determine the appropriate grid size for the newly established tunnel scene dataset.

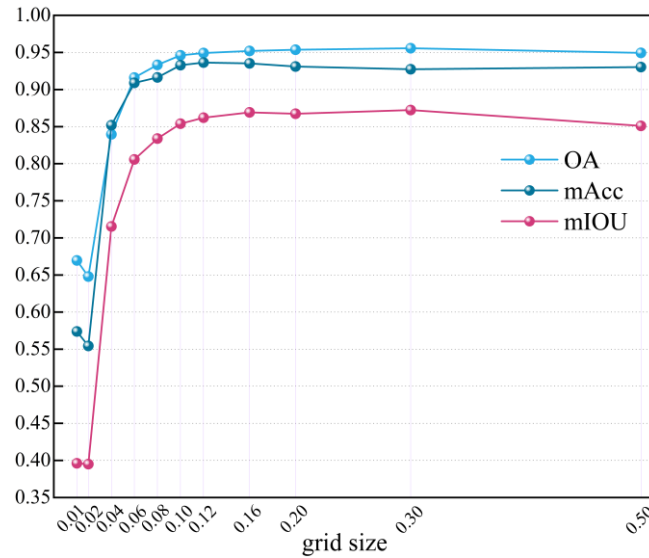


Figure 4. The performance curve obtained after data preprocessing with different grid sizes.

Table 3. Semantic segmentation results of different grid size.

Models	Grid size	mIoU	0	1	2	3	4	5	6	7	8	9	10	11	12	13
G1	0.01	0.3961	0.1612	0.2823	0.5273	0.3370	0.3567	0.5921	0.8634	0.5117	0.3281	0.7192	0.0243	0.6221	0.0000	0.2202
G2	0.02	0.3951	0.2791	0.1736	0.5288	0.1891	0.2232	0.5826	0.9005	0.4294	0.5607	0.7213	0.0227	0.7017	0.0000	0.2190
G3	0.04	0.7154	0.5709	0.3603	0.7160	0.5785	0.6664	0.8311	0.9201	0.7157	0.7985	0.9324	0.7857	0.8635	0.8771	0.3990
G4	0.06	0.8058	0.6808	0.4534	0.8758	0.8061	0.7400	0.8973	0.9380	0.8967	0.8796	0.9536	0.9284	0.8854	0.8883	0.4580
G5	0.08	0.8339	0.7234	0.5557	0.9175	0.8215	0.8010	0.8995	0.9355	0.9252	0.8853	0.9630	0.9143	0.9280	0.9056	0.4988
G6	0.10	0.8540	0.7490	0.5819	0.9395	0.8955	0.8836	0.9177	0.9340	0.9606	0.8887	0.9655	0.8861	0.9171	0.9297	0.5076
G7	0.12	0.8621	0.7699	0.6179	0.9449	0.9002	0.8811	0.9225	0.9274	0.9658	0.9109	0.9651	0.8698	0.9431	0.9330	0.5173
G8	0.16	0.8693	0.8443	0.5440	0.9333	0.9194	0.9033	0.9256	0.9386	0.9735	0.9262	0.9796	0.9167	0.9390	0.8920	0.5350
G9	0.20	0.8674	0.8051	0.6090	0.9410	0.9128	0.8899	0.9366	0.9395	0.9708	0.9011	0.9755	0.9099	0.9439	0.8704	0.5377
G10	0.30	0.8724	0.8519	0.6079	0.9511	0.9201	0.8996	0.9355	0.9375	0.9726	0.8954	0.9259	0.8842	0.9599	0.9378	0.5347
G11	0.50	0.8510	0.8001	0.6076	0.9454	0.9074	0.8385	0.9147	0.9368	0.9646	0.8651	0.9637	0.8815	0.9373	0.8538	0.4972

Preprocess the raw point cloud and eleven models are trained using different input sizes, as listed in Table 3. RandLA-Net (Hu, 2020) containing a random sampling module is adopted here considering the remarkably computation and memory efficiency. Although RandLA-Net is no longer limited to $1\text{m} \times 1\text{m}$ blocks compared with pointnet++ (Qi, 2017), it can effectively learn the overall geometric structure of objects. However, for a brand-new semantic segmentation scenario, it is crucial to know how to set up the hyperparameters. The results show that when the grid size exceeds 0.12, there is a plateau in model performance (Figure 4). The mIoU of Model G6 reaches 86.21%. In the process of random sampling, number of input points, i.e., the number of nearest neighbors of the query point, is set to 40960, so a larger grid size means a larger receptive field, as shown in Figure 5. As a result, this means that the model is able to learn the overall geometry of the object more efficiently during training. As a larger range of point clouds are added to the training dataset, there is a notable increase in both OA and, in particular, mIoU. However, it is reasonable to expect a degradation of the model's performance when the grid size is too large, such as over 0.3, as this will discard the key features. The point cloud is downsampled with a four-fold decimation ratio.

During the training process, a loss function is typically used to intuitively show the trend of the difference

between the model's predictions and the ground truth as the number of iterations increases. Under the same learning rate and batch size settings, the loss curves of models with different grid sizes during the training process are plotted as shown in Figure 6. Ideally, the loss curve should show a gradual decline and tend to stabilize. This indicates that as training progresses, the model's performance on the training set continues to improve. It can be observed that from G7 onward, this trend is evident. However, when the grid size is less than 0.12, the loss curve remains oscillatory and noisy throughout the training process, indicating that the loss function fails to converge. Conversely, when the grid size exceeds 0.12, the loss curve decreases rapidly in the early training stages and stabilizes, suggesting that the model converges quickly. This provides a parameter reference for achieving good performance in scenarios involving data such as tunnels. A visual comparison of semantic segmentation results of G7 (Grid size = 0.12 m) and ground truth is shown in Figure 7.

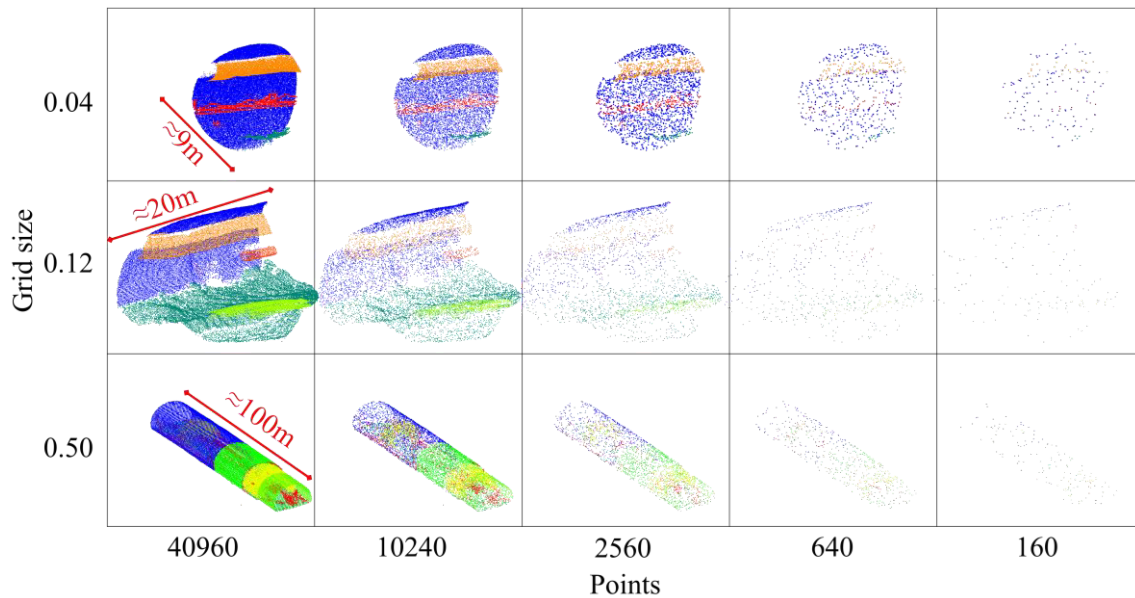


Figure 5. The area of points entered for each layer after setting different grid sizes.

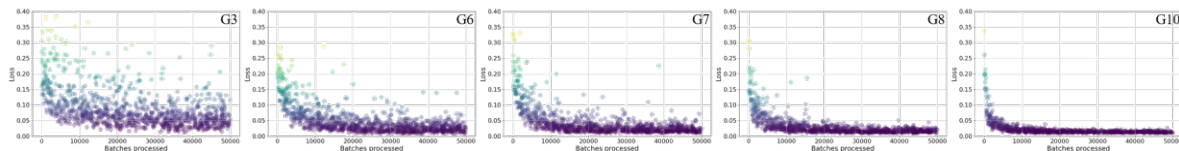


Figure 6. The loss curves of models with different grid sizes during the training process.

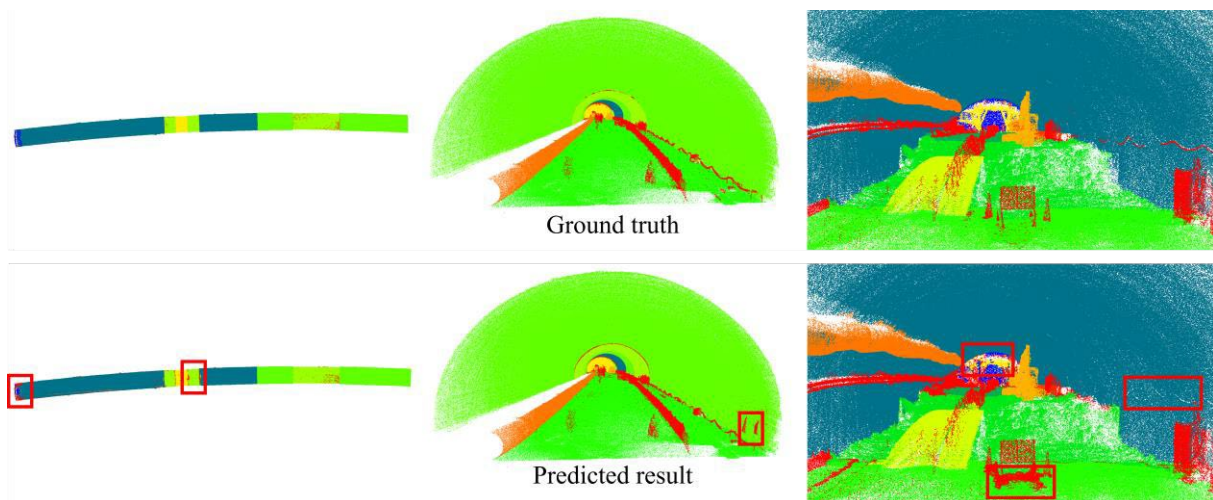


Figure 7. Visual comparison of results of semantic segmentation with different grid sizes.

4. CONCLUSIONS

We have presented a new 3D point cloud database manually annotated, specifically designed for construction highway tunnel environment. The dataset covers approximately 9 km of tunnel section with over 1.58 billion 3D points. Each 3D point has been classified into 14 categories, resulting in a list of (x, y, z) points. Three popular end-to-end point cloud semantic segmentation algorithms were tested as baselines for this dataset. We analyzed the impact of grid size on model performance and determined that a grid size of 0.12 is suitable for tunnel scenarios.

The intention of presenting this new tunnel point cloud dataset is to encourage developing creative deep learning models. This will promote the adoption and application of 3D point clouds in the tunnel engineering industry, thereby accelerating the digital transformation of traditional sectors. The labels of this new dataset will be improved and updated with feedback from the research community.

ACKNOWLEDGMENTS

REFERENCES

- A, M. H., A, M. A., and B, U. S. (2013). Change detection in urban areas by object-based analysis and on-the-fly comparison of multi-view data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 86(12), 52-64.
- Armeni, I., Sener, O., Zamir, A., Jiang, H., Brilakis, I., and Fischer, M. (2016). 3d semantic parsing of large-scale indoor spaces. *IEEE Computer Society*.
- Behley, J., Garbade, M., Milioto, A., Quenzel, J., Behnke, S., and Stachniss, C. (2020). SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE.
- Chi, L., Yuhang, L. and Yuhua, C., (2023). A State-of-the-Practice Review of Three-Dimensional Laser Scanning Technology for Tunnel Distress Monitoring. *Journal of Performance of Constructed Facilities*, 37(2), pp.1-22.
- Cui, L., Zhou, L., Xie, Q., Liu, J., Han, B. (2023). Direct generation of finite element mesh using 3D laser point cloud. *Structures*. V47 1579-1594.
- Cui, L.Z., Liu, J., Luo, H., Wang, J., Zhang, X., Lv, G. and Xie, Q., (2024). Deformation Measurement of Tunnel Shotcrete Liner Using the Multiepoch LiDAR Point Clouds. *Journal of Construction Engineering and Management*, 150(6), p.04024049.
- Deng, H., Zhu, H. and Guo, J., (2024). Research on Deformation Monitoring of Roadway Surrounding Rock Based on Mobile 3D Laser Scanning Technology. *Academic Journal of Science and Technology*, 10(3), pp.155-162.
- Duan, D.Y., Qiu, W.G., Cheng, Y.J., Zheng, Y.C. and Lu, F., (2021). Reconstruction of shield tunnel lining using point cloud. *Automation in Construction*, 130, p.103860.
- Fekete, S., Diederichs, M. and Lato, M., (2010). Geotechnical and operational applications for 3-dimensional laser scanning in drill and blast tunnels. *Tunnelling and underground space technology*, 25(5), pp.614-628.
- Hackel, T., Savinov, N., Ladicky, L., Wegner, J.D., Schindler, K. and Pollefeys, M., (2017). Semantic3d. net: A new large-scale point cloud classification benchmark. arxiv preprint arxiv:1704.03847.
- Hu, Q., Yang, B., Xie, L., Rosa, S., and Markham, A. (2020). RandLA-Net: Efficient Semantic Segmentation of Large-Scale Point Clouds. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE.
- Kim, Y. and Bruland, A., (2019). Analysis and evaluation of tunnel contour quality index. *Automation in Construction*, 99, pp.223-237.
- Lato, M.J. and Diederichs, M.S., (2014). Mapping shotcrete thickness using LiDAR and photogrammetry data: Correcting for over-calculation due to rockmass convergence. *Tunnelling and Underground Space Technology*, 41, pp.234-240.
- Lin, W., Sheil, B., Zhang, P., Zhou, B., Wang, C., and Xie, X. (2024). Seg2tunnel: a hierarchical point cloud dataset and benchmarks for segmentation of segmental tunnel linings. *Tunnelling and Underground Space Technology incorporating Trenchless Technology Research*, 147.
- Munoz, D., Bagnell, J. A., Vandapel, N., and Hebert, M. (2009). Contextual classification with functional max-margin markov networks. *IEEE*.
- Qi, C. R., Su, H., Mo, K., and Guibas, L. J. (2017). Pointnet: deep learning on point sets for 3d classification and segmentation. *IEEE*.
- Qi, C. R., Yi, L., Su, H., and Guibas, L. J. (2017). Pointnet++: deep hierarchical feature learning on point sets in a metric space.

- Rachmawati, T.S.N. and Kim, S., (2022). Unmanned Aerial Vehicles (UAV) integration with digital technologies toward construction 4.0: A systematic literature review. *Sustainability*, 14(9), p.5708.
- Thomas, H., Qi, C. R., Deschaud, J. E., Marcotegui, B., Goulette, F., and Guibas, L. (2020). KPConv: Flexible and Deformable Convolution for Point Clouds. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE.
- Tong, G., Li, Y., Chen, D., Sun, Q., and Xiang, G. (2020). Cspc-dataset: new lidar point cloud dataset and benchmark for large-scale semantic segmentation. *IEEE Access*, PP (99), 1-1.
- Wang Z, Hu M, Lai J, et al. (2023). Reliability analysis method for tunnel structural design: Brief review and relevant prospects. *Structures*, 55: 1894-1905.
- Wang, Y., Sun, Y., Liu, Z., Sarma, S. E., Bronstein, M. M., and Solomon, J. M. (2018). Dynamic graph cnn for learning on point clouds. *ACM Transactions on Graphics*, 38(5).
- Wu, X., Jiang, L., Wang, P. S., Liu, Z., Liu, X., and Qiao, Y. (2024). Point Transformer V3: Simpler, Faster, Stronger. *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
- Xie, J., Huang, X., Zhang, Z., and Jin, G. (2023). Cohesive zone model-based analyses of localized leakage of segmentally lined tunnels. *Frontiers of structural and civil engineering*.
- Xie, X., and Lu, X. (2017). Development of a 3d modeling algorithm for tunnel deformation monitoring based on terrestrial laser scanning. *Underground Space*, 2(1), 16-29.
- Xie, Y., Tian, J., and Zhu, X. X. (2020). Linking points with labels in 3d: a review of point cloud semantic segmentation. *Geoscience and remote sensing* (4), 8.
- Xu, Z.H., Yu, T.F., Lin, P., Wang, W.Y. and Shao, R.Q., (2022). Integrated geochemical, mineralogical, and microstructural identification of faults in tunnels and its application to TBM jamming analysis. *Tunnelling and Underground Space Technology*, 128, p.104650.
- Ye, Z., Xu, Y., Huang, R., Tong, X., and Stilla, U. (2020). Lasdu: a large-scale aerial lidar dataset for semantic labeling in dense urban areas. *International Journal of Geo-Information*, 9(7), 450.
- Yi, X., Feng, W., Wang, D., Yang, R., Hu, Y., and Zhou, Y. (2023). An efficient method for extracting and clustering rock mass discontinuities from 3D point clouds. *Acta Geotechnica*, 18(7), 3485-3503.
- Zhao, H., Jiang, L., Fu, C. W., and Jia, J. (2019). PointWeb: Enhancing Local Neighborhood Features for Point Cloud Processing. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE.
- Zhao, Y., Zhu, Z., Liu, W., Zhan, J. and Wu, D., (2023). Application of 3D laser scanning on NATM tunnel deformation measurement during construction. *Acta Geotechnica*, 18(1), pp.483-494.
- Zolanvari, S. M. I., Ruano, S., Rana, A., Cummins, A., Da Silva, R. E., and Rahbar, M. (2019). Dublincity: annotated lidar point cloud and its applications.

A SCAN-TO-BIM APPROACH FOR RENOVATING EXISTING BUILDING ROOMS USING SURFACE RECONSTRUCTION AND DEEP LEARNING

May Thin Zar Soe¹, Nobuyoshi Yabuki², Tomohiro Fukuda³, and Yoshiro Hada⁴

1) Master Course Student, Division of Sustainable Energy and Environmental Engineering, Osaka University, Osaka, Japan. Email: soe@it.see.eng.osaka-u.ac.jp

2) Ph.D., Prof., Division of Sustainable Energy and Environmental Engineering, Osaka University, Osaka, Japan. Email: yabuki@see.eng.osaka-u.ac.jp

3) Ph.D., Assoc. Prof., Division of Sustainable Energy and Environmental Engineering, Osaka University, Osaka, Japan. Email: fukuda.tomohiro.see.eng@osaka-u.ac.jp

4) D. Eng., Group Leader, Advanced Digital Technology Laboratory, Institute of Technology, Tokyu Construction Co., Ltd. Email: hada.yoshirou@tokyu-cnst.co.jp

Abstract: Since global energy consumption is a critical issue and the building sector plays a significant role in high energy demand, renovating existing buildings is crucial. Building Information Modelling (BIM) represents the physical and functional characteristics of a building or structure with detailed information. It allows all the architecture, engineering, and construction (AEC) industry stakeholders to work collaboratively. As-built BIM models reflect the modifications of existing conditions of buildings and fully automated generation of as-built BIM models remains a major challenge. The Scan-to-BIM process is widely used for the renovation and documentation of existing buildings. This process includes capturing the physical conditions of a building or structure using 3D laser scanning technology and converting it into a BIM model. Originally, 3D laser scanners had a very high cost, however, free 3D scanning applications that use Light Detection and Ranging (LiDAR) technology can be easily compatible with mobile phones or tablets nowadays. This paper proposes to contribute to renovating existing buildings by developing a new approach to scan-to-BIM combining surface reconstruction from point cloud data and object detection with deep learning. A 3D free scanning application generated the point cloud data of the interior room of the building and the surface model was reconstructed. Moreover, the location of the air terminals on the ceiling was investigated by developing the object detection model with deep learning and installing the air terminals on the surface reconstructed model. The finalised surface model with the air terminals was exported to Industry Foundation Classes (IFC) format. The reconstructed IFC model developed in this research can be used for Computational Fluid Dynamics (CFD) analysis with appropriate property sets.

Keywords: Building Information Modelling, Surface Reconstruction, Object Detection, Deep Learning, Air Terminals, Indoor Environment Simulation

1. INTRODUCTION

Building energy consumption accounts for about 40% of global energy consumption and the energy-saving retrofit of existing buildings provides a significant opportunity to reduce global energy consumption (Huang et al., 2022). A BIM model is a digital representation of facilities that records all information relevant to a building's life cycle from construction to demolition (Patraean et al., 2015). Building owners, facilities management groups, and governments are leading efforts to use BIM for new construction, renovation, or refurbishment of various facilities (Jung et al., 2018). In addition, it is important to note that the BIM created in the design stage of a facility is called as-designed BIM, and the BIM that reflects a facility in its as-built condition is called as-built BIM (Bosche et al., 2015).

As-built BIM are the key aspects in the renovation, maintenance, or further facility management as they support documenting the existing buildings by modifying the existing conditions. However, the traditional generating process flow of as-built models has several challenges such as the necessary for a professional background for field surveys and measurements, taking time, and the low accuracy of manual collection of data and remodelling (Varajic, 2020). Various technologies such as photogrammetry and laser scanning can generate point cloud data as the digital representation of the object or the scene. Therefore, combining modern technologies, such as laser scanning, and LiDAR with BIM can be a powerful tool for optimising the process, providing accurate data in less time (Qiu et al., 2021). Many efforts have been made to automatically develop BIM with these immediate survey datasets, such as a 3D point cloud or 3D mesh model with specific algorithms (Tang et al., 2022). However, the automated generation of the BIM model of indoor environments is challenged by the inherent noise and incompleteness of the data and requires further investigation (Tran et al., 2020).

Furthermore, object detection is a fundamental visual recognition problem in computer vision. It recognises object categories and predicts each object's location by a bounding box (Wu et al., 2020). Deep learning (DL) has become the most widely used computational approach in the field of Machine Learning (ML) with its ability to achieve outstanding results on several complex cognitive tasks. One of the benefits of DL is the ability to learn massive amounts of data (Alzubaidi et al., 2021). Object detection with deep learning combines computer vision and neural networks to detect and localise images or videos automatically.

Based on these backgrounds, it is known that combining deep learning technology and Scan-to-BIM for indoor building rooms remains underexplored. Therefore, the research attempts to directly reconstruct the surface model from the point cloud and get the locations of the air terminals by object detection with deep learning. The main purpose of this research is to develop a new system of reconstructing the interior room of the building including the air terminals and use it for CFD analysis for further energy renovation processes.

2. RELATED WORKS

Indoor environment simulation of existing buildings is a key tool for improving comfort, air quality, and energy efficiency. BIM models can enhance simulations' accuracy, efficiency, and reliability, contributing to better-informed decisions not only for buildings but also for heating, ventilation, and air conditioning (HVAC) systems. Matsuba et al. (2023) developed a new IFC data model for BIM, sensors, and indoor environment simulation. However, their research exists an original 3D model of the building of interest. This study will reconstruct the BIM model of the existing building by directly scanning the existing building room.

Many researchers have developed various approaches to reconstruct indoor spaces from point clouds. Varghese et.al. (2016) investigated the building boundary tracing and regularisation from the LiDAR point cloud. Jung et al. (2018) developed a 3D volumetric reconstruction of multiple-room building interiors for as-built BIM. Xie et al. (2017) reconstructed indoor buildings from mobile laser scanning data. However, the boundary generation method used in their studies started with point cloud preprocessing such as noise reduction or clustering. Their boundary extraction is based on floor segmentation. The boundary generation in our study will overcome noise removal of the point cloud model by directly plane segmentation to the raw point cloud data and detecting the boundary points. The ceiling will be chosen for the plane segmentation so that difficulties in getting point cloud data of the exact floor shape due to many furniture on the floor can be avoided. Moreover, their boundary regularisation used contour regularisation which can lead to over-smoothing and irregular shapes. Our boundary regularisation method will focus on each corner point and boundary line.

Ishikawa et al. (2021) directly unevenly spaced Cartesian grids from laser-scanned point clouds without creating any 3D model. Their method needs ten scanned point clouds to register to generate the grid and has to manually assign the boundary condition for the placement of the air inlet and outlet. This study will create object detection datasets for air terminals to automatically figure out the locations of the air terminals of the existing building rooms.

3. DEVELOPMENT OF THE PROPOSED METHOD

3.1 Overview of the Method

The proposed method focuses on the 3D reconstruction of the interior room of the building for CFD analysis with three main steps: surface reconstruction, air terminal detection and IFC modelling. Firstly, the point cloud model of the interior room is scanned using Scaniverse, a free application that uses LiDAR technology. The room's boundary points were extracted using plane segmentation and boundary detection algorithms. The concave hull algorithm changed the detected point cloud boundary to the line segments and they were regularised. The regularised boundary was regarded as the floor boundary line and floor-to-ceiling height was estimated by the histogram approach. The surface model was reconstructed based on the regularised floor boundary and estimated floor-to-ceiling height. As for the object detection part, the air conditioning and air-conditioning vents datasets were prepared and trained using the deep learning model. The finalised object detection model identifies the locations of the air conditioner and air conditioning vents on the ceiling using x and y coordinates relative to the leftmost top corner point. Based on these coordinates, air terminals were installed on the reconstructed surface model and the finalised model was exported to IFC format. IFC is an open and standardised file format for sharing and exchanging BIM models across different software platforms developed by buildingSMART. The overall process flow is shown in Figure 1.

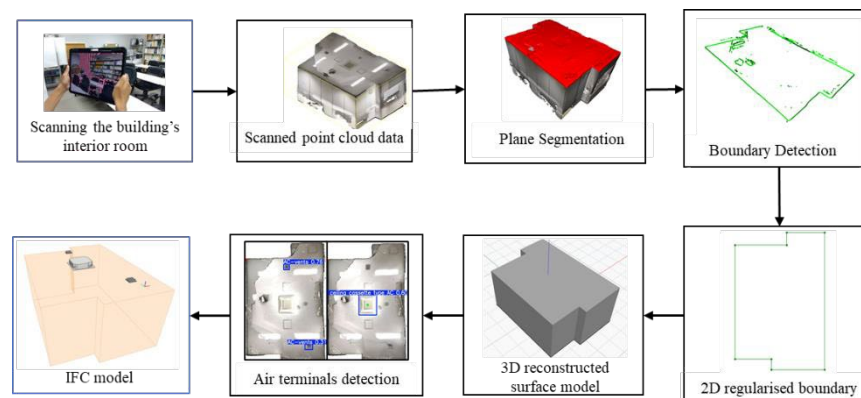


Figure 1. Overall process flow

Our research is supposed to generate the IFC model of the interior room of the building from the point cloud by the combination of surface reconstruction and object detection with deep learning. The plane detection and extraction on the ceiling allow the avoidance of manual processing of the noisy point cloud data due to the interior furniture. A user-friendly free 3D scanning application was used to get the point cloud data. The image of the point cloud model does not have enough resolution if compared with the normal images and custom datasets were created for that. The final IFC model including the air terminals can be used for CFD analysis in future work.

3.2 Generation of the Surface Model from Point Cloud

Figure 2 shows the detailed process of generating a surface model from the raw point cloud. The red colour shows our original tasks and the blue colour shows the existing tasks from the previous research.

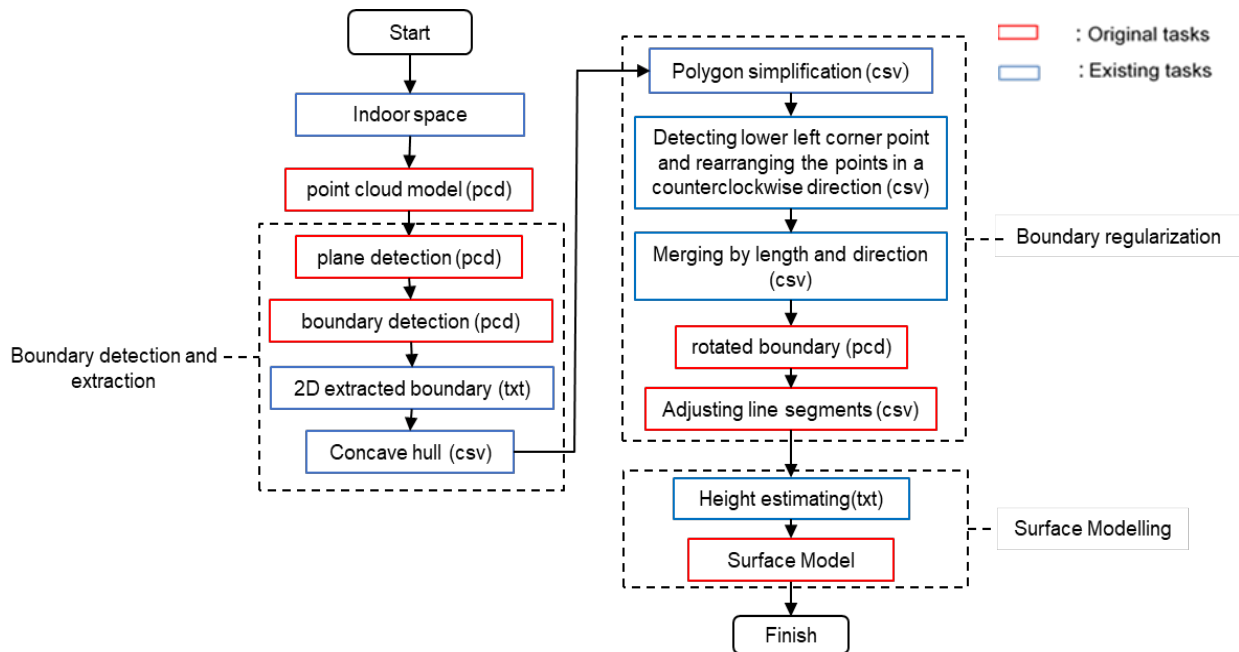


Figure 2. Process flow of wall-surface boundary generation

3.2.1 Boundary Detection and Extraction

The interior room of the building was scanned by Scaniverse 2.0.0 (Niantic, 2024) on iPad and exported into point cloud format as shown in Figure 3 (a). The largest planar segment, the ceiling as shown in Figure 3 (b), is identified using the Random Sample Consensus (RANSAC) algorithm (Fischler & Bolles, 1981) implemented in the Open3D library. In this step, the ceiling was chosen to be segmented (Figure 3 (c)) instead of the floor as the exact boundary shape of the floor is difficult to capture due to a lot of noises caused by furniture on the floor. The boundary points of the segmented ceiling were detected (Figure 3 (d)) by using the boundary detection algorithm (Mineo et al., 2019) of the Open3D library, which finds the boundary points among an unordered point cloud by analysing the angle among the normal of a point and its neighbours.

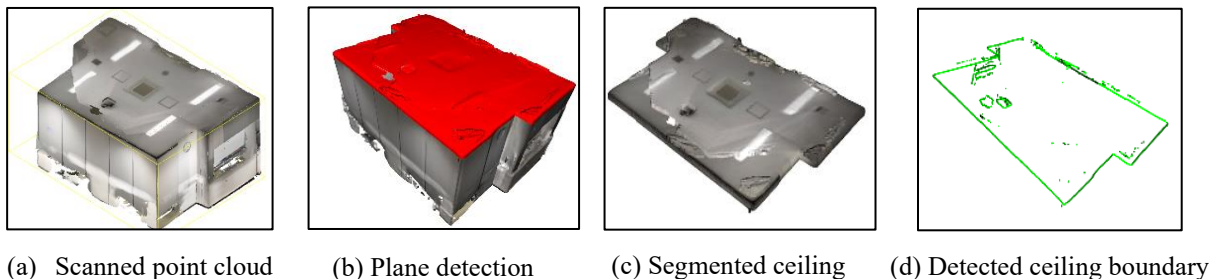


Figure 3. Ceiling boundary detection and extraction

3.2.2 Boundary Regularisation

The detected 3D boundary points were changed to 2D boundary points as shown in Figure 4 (a) and connected with line segments by a 2D concave hull algorithm as shown in Figure 4 (b). The concave hull was simplified as shown in Figure 4 (c) using the Douglas-Peucker Algorithm (Douglas & Peucker, 1973) which can reduce the number of points in a poly-line approximation of a curve while preserving its essential shape. However, the simplified boundary still has curves on the main corners of the boundary. As a result, the boundary lines were merged based on their lengths and directional differences. Line segments which are shorter than a length threshold and have directional differences within a threshold are combined. One important thing to note before going to this step is that the sequence of the points directly extracted from the point clouds is messy, and the right length and directional differences of the adjacent line segments cannot be determined. To address this, the lower left point was detected, and all the points were arranged in an anticlockwise direction. The boundary merged by length and direction difference can be seen as shown in Figure 4 (d). As for the automatic rotation of the boundary line, the angle of deviation of the first line segment was calculated by trigonometric function. The rotation matrix was written based on the negative of the angle of deviation and applied to all the start and end points. The new rotated boundary is shown in blue as shown in Figure 4 (e). For the final regularisation of the boundary, the direction for adjusting the line segment was determined by comparing the absolute difference in x, and y-coordinates. For example, if the absolute difference in x-coordinates is smaller than in y-coordinates, the segment is more horizontally aligned and needs to be adjusted vertically. Following this, a vertical line was drawn from the start point's x-coordinate to the end point's y-coordinate by keeping the x-coordinate constant for a vertical adjustment, and a horizontal line was drawn from the start point's y-coordinate to the end point's x-coordinate by keeping the y-coordinate constant. The final regularised boundary can be seen as shown in Figure 4 (f).

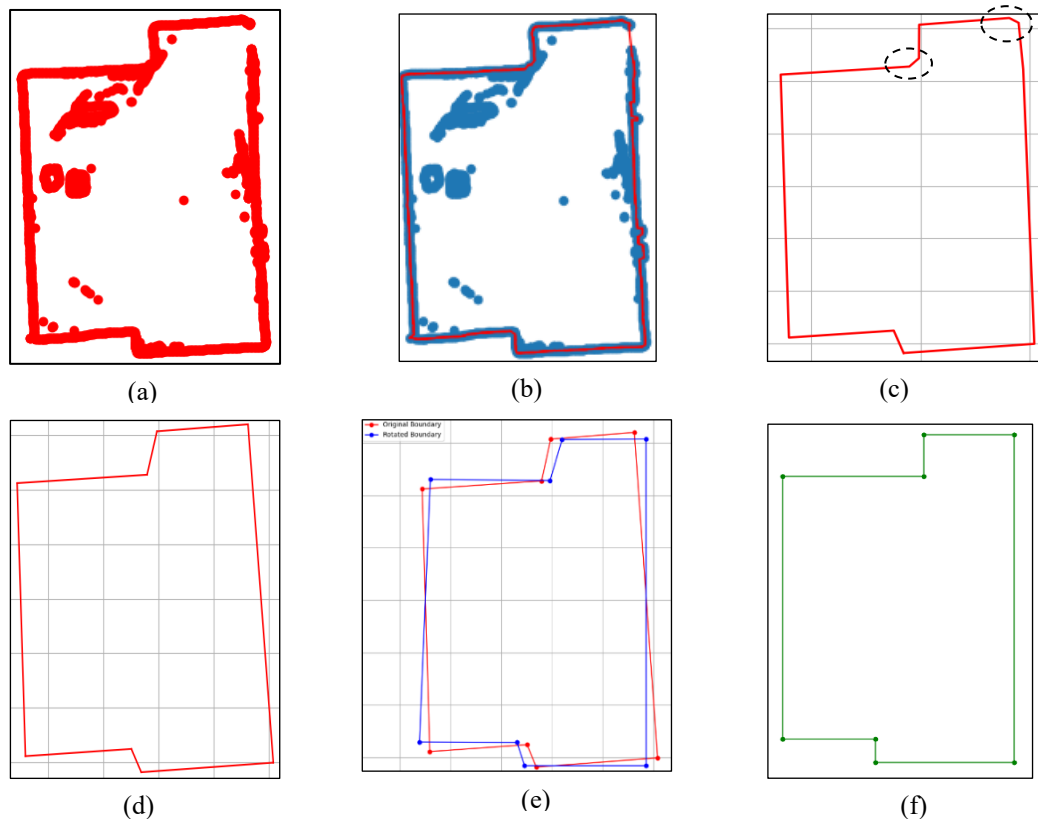


Figure 4. Boundary regularisation

3.2.3 Surface Modelling

Before proceeding to the surface modelling stage, the floor-to-ceiling height estimation was developed first. Many researchers have demonstrated the effectiveness of estimating floor-to-ceiling height using histograms (Turner et al., 2014) and we used this method for estimation of the height of our proposed model. The z-coordinates of the point cloud model were extracted, and the mean z-values were computed. The low mean z-values were defined as the floor, and the high mean z-values as the ceiling. The height difference between the two peaks shows the floor-to-ceiling height. The developed histogram is shown in Figure 5.

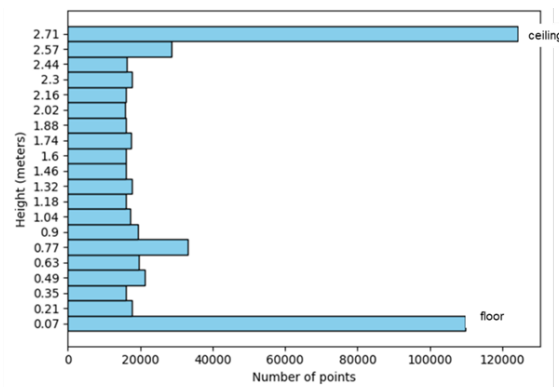


Figure 5. Histogram representing the number of scan points and their z-values

Afterwards, the regularised boundary in the last stage was imported to the Revit Dynamo by Python Script. The imported line segments as shown in Figure 6 (a) were joined as poly curves and the floor surface was created as shown in Figure 6 (b). The wall surfaces were created by extruding the boundary line segments with the z-direction value of the estimated height by the histogram as shown in Figure 6 (c). The ceiling surface was created by translating geometry from the floor with the z-translation. The reconstructed surface model of the room is shown in Figure 6 (d).

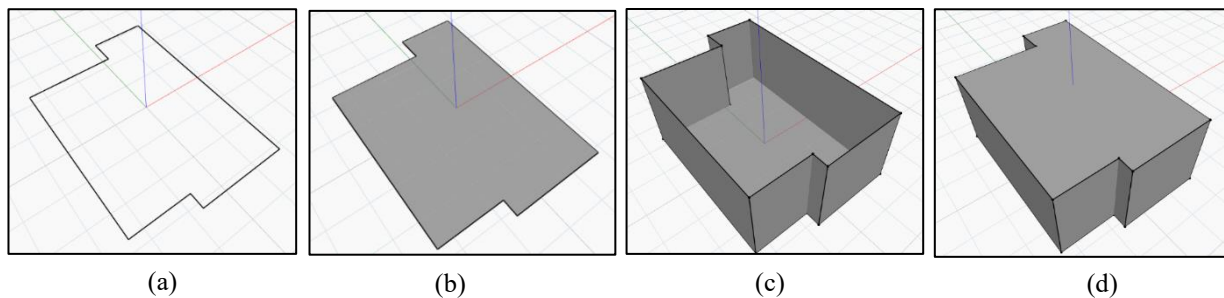


Figure 6. Surface modelling in Revit Dynamo

3.3 Detection of Air Terminals

3.3.1 Dataset Details and Training Conditions

This research developed two object detection datasets for detecting air conditioners (AC) and air conditioning vents (AC vents) with 218 and 281 images respectively. The datasets' details are shown in Table 1. The collected images were manually annotated with Roboflow (Dwyer et al., 2024), an end-to-end platform for building, training, and deploying computer vision models. The target image for detection is the image of the point cloud ceiling and AC vents have strikes that are difficult to be captured well in the point cloud. To deal with this, augmentation was added while creating the AC vents dataset.

You Only Look Once Version 5 (YOLOv5) (Jocher et al., 2020) is the state-of-the-art object detection model designed for real-time detection of objects within images and videos. The datasets created were trained with YOLOv5 and the training results for each dataset are shown in Table 2. The development environment is shown in Table 3.

Table 1. Datasets details

	AC dataset	AC vents dataset
Number of images	218 (trained: 153, valid: 33, test: 17)	677 (trained: 594, valid: 54, test: 29)
Collection method	Canon Powershot SX710 HS, downloaded from pj Dataset	Canon Powershot SX710 HS
Image size	640 x 640 px	640 x 640 px
Annotation Method	manually	manually
Classes	2 (4-way ceiling cassette-type AC, wall-mounted type AC)	1 (AC vent)
Augmentation	N.A.	grayscale 30%, blur 25px, noise 5%

Table 2. Training results

OS	Windows 11 Education 23H2
CPU	12 th Gen Intel(R) Core (TM) i7-12700
RAM	32.0 GB
GPU	NVIDIA GeForce RTX 3060 Ti

Table 3. Development environment (desktop PC)

	AC dataset	AC vent dataset
Precision (P)	1	0.986
Recall (R)	0.972	1
mAP ₅₀	0.994	0.993
mAP ₅₀₋₉₅	0.875	0.851

3.3.2 Detection Results

For the testing of the prepared object detection model, preprocessing of the input image is necessary. Initially, the point cloud of the segmented ceiling as shown in Figure 7 (a) was converted to an RGB image as shown in Figure 7 (b) with Open3D and Numpy libraries. To ensure that the object was properly centred in the image, the blank space around the main object was trimmed using a custom function with Python Imaging Library (PIL). Subsequently, the main object inside the image was rotated using the same method in the boundary regularisation stage. The edges of the objects in the image were detected by Canny edge detection (Canny, 1987) and vertical lines were identified by Hough line transform (Hough, 1962), focusing on those with angles between 80 and 100 degrees. Then, the inclination angle of the first detected vertical line was calculated and the object in the image was rotated as shown in Figure 7 (c). The detection results with the prepared object detection model on different images can be seen in Figure 8.

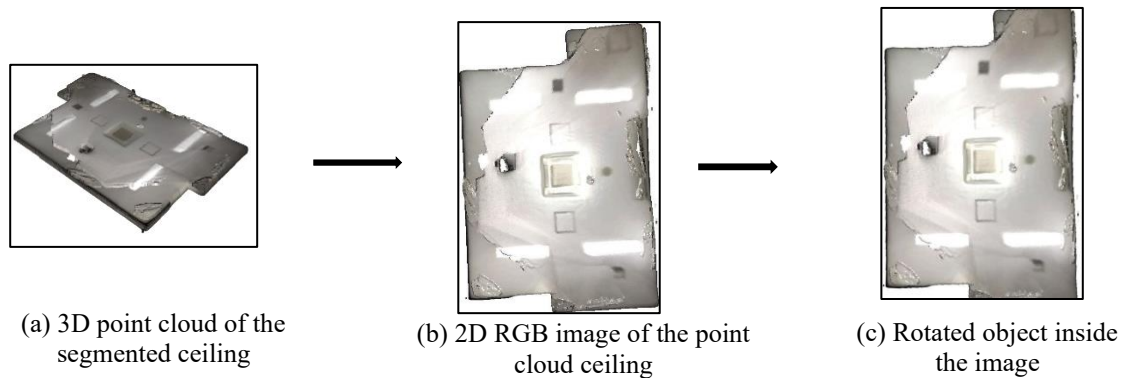


Figure 7. Preparation of image for object detection

Input	Output	
	AC dataset	AC vents dataset
		
		
		
		
		

Figure 8. Detection results

3.4 Modelling of AC Terminals and Generation of IFC Model

Firstly, the centres of the bounding boxes of the detected objects were computed. Then, the left-top corner point was detected by the Shi-Tomasi Corner Detection Method (Shi & Tomasi, 1994) with the OpenCV library. The distances between that corner point and the bounding box centres were evaluated. Finally, the pixel coordinates were converted to real-world coordinates by multiplying by the pixel size. The centre points of the AC and AC vents were added to the ceiling of the surface model considering the left-top corner point as the reference, via the Point.ByCoordinates Dynamo node as shown in Figure 9.

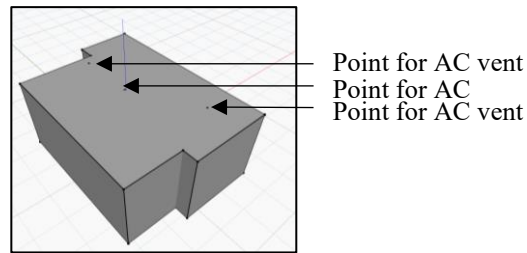


Figure 9. Locating the points of AC and AC vents in Dynamo Revit

The AC and AC vents were added to the surface model by FamilyInstance.ByPoint Dynamo node, which requires the family type and point as inputs. The AC family, downloaded from the bimobject website and loaded into Revit, was connected to the family input. The AC location points calculated from the last stage were connected to the point input node. As a result, the AC and AC vents were automatically added and aligned on the ceiling of the surface model as shown in Figure 10.

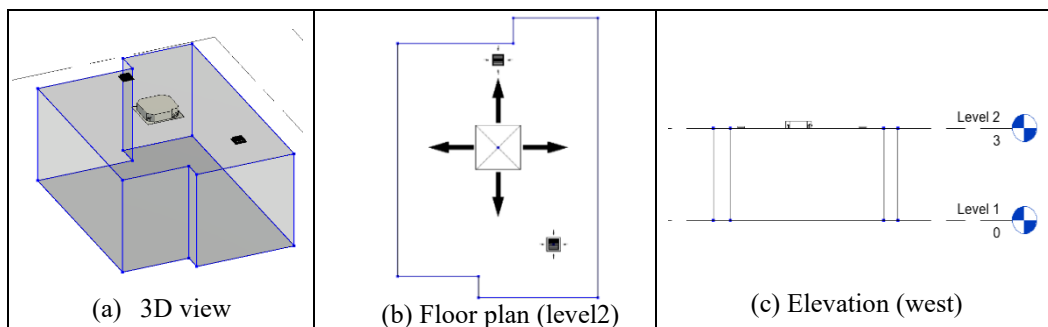


Figure 10. Different views in Revit

The reconstructed surface model was exported to the IFC4 schema as shown in Figure 11 using Autodesk Revit. IFC4 offers more comprehensive property sets for HVAC elements like air terminals by supporting detailed attributes for flow rates, pressure losses, and performance data which are vital for CFD analysis.

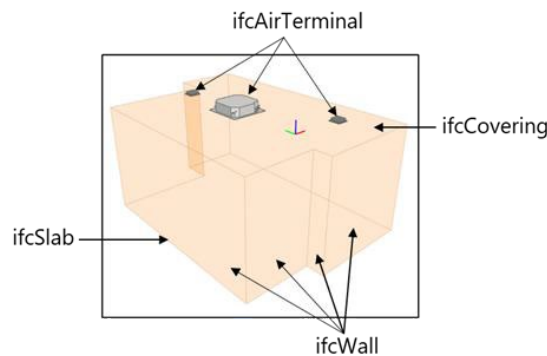


Figure 11. IFC model viewed on BIM Vision

4. ACCURACY ASSESSMENT

Intersection over Union (IoU) method was used to check the accuracy of the reconstructed surface model. IoU is a powerful metric for comparing two shapes based on their overlap and the IoU of the developed model in this research is calculated as the area of the intersection divided by the area of the union of the concave hull from the extracted boundary and final regularised boundary. Details can be seen in Table 4. The comparison between the ground truth model and the reconstructed model is shown in Figure 12.

Table 4. IoU Result

	Area (m ²)	Area of Intersection (m ²)	Area of Union (m ²)	IoU
Concave hull from the extracted boundary	24.58	23.28	25.01	0.931
Regularised boundary	23.71			

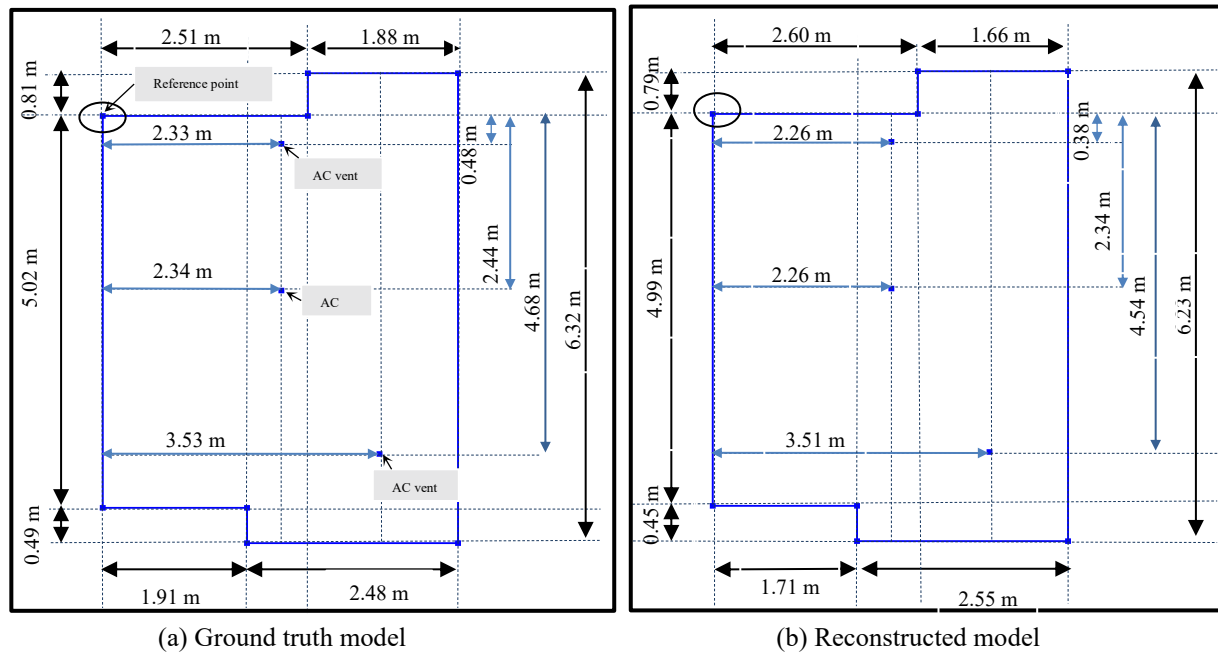


Figure 12. Comparison between the ground model and reconstructed model

5. DISCUSSION

This research developed a new method of generating an IFC model of the existing building rooms from point cloud data. The scanning device used in this research can be compatible with mobile phones or tablets and easily accessible by users without a professional background. Unlike most previous studies which emphasise extracting the boundary from the floor, our method addresses the challenges of noise removal and getting the exact boundary shape directly from the point cloud. The boundary detection in our study is achieved by segmenting the ceiling through plane detection and there is no need to worry about not getting the exact boundary shape due to a lot of furniture on the floor or the tall bookshelves against the walls. This research developed two custom datasets aimed at detecting the air terminals inside the images of the point clouds which have lower resolutions compared to normal images.

However, certain stages in boundary line generation from the detected point cloud boundary in this research require the user to manually set the parameter thresholds. This can lead to time consumption and data instability. Developing a program that selects the appropriate parameter thresholds for different input data would enhance productivity. Additionally, this research's boundary regularisation method focuses on points and straight-line segments, limiting its applicability to interior rooms with curved boundaries. The openings such as doors and windows, were not considered in this research. In detecting the air terminals, particularly the AC vents, faced significant challenges due to the lines on them, which were difficult to identify because of their low visibility in the point cloud images. As a solution, two datasets were developed separately for AC and AC vents with different augmentations. Future research should explore more efficient methods for linking the point cloud data to the reconstructed model directly, or how to generate a unified dataset able to detect all air terminals in point cloud images.

6. CONCLUSIONS

In conclusion, this study demonstrates a new approach to Scan-to-BIM by combining surface reconstruction and object detection with deep learning. The free application on tablet generated the point cloud data used for this research and can be easily accessible by everyone who lacks specialised knowledge and experience. The boundary detection and regularisation methods introduced in this research enable the rapid and accurate generation of exact boundary shapes, closely matching the real room's boundary shape. All the steps from plane detection to boundary regularisation were completed by Python script without using any manual or paid software. The surface reconstruction from the boundary line segments and installing the air terminals using the location obtained from object detection were accomplished with the Revit Dynamo. As for the air terminal detection, deep learning-based object detection was utilised and custom datasets specifically focused on the point cloud image were developed.

Future work will expand the object detection part by considering the openings, such as doors and windows, as well as material types. Additionally, the IFC model developed in this research will be enhanced by integrating the custom property sets prepared for CFD analysis.

REFERENCES

- Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaria, J., Fadhel, M.A., AL-Amidie, M., and Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions, *J Big Data*, 8(1), 53.
- Bosche, F.N., O’Keeffe, S. (2015). The need for Convergence of BIM and 3D imaging in the open world, *Proceedings in the CItA BIM Gathering Conference*, Dublin, Ireland, 109-116.
- Canny, J. (1987). A Computational Approach to Edge Detection, *Reading in Computer Vision*, 184-203.
- Douglas, D.H., and Peucker, T.K. (1973). Algorithms for the Reduction of the Number of Points Required to Represent a Digitized Line or its Caricature, *THE CANADIAN CARTOGRAPHER*, 10(2), 112-122.
- Dwyer, B., Nelson, J., Hansen, T., et al. (2024). Roboflow (Version 1.0) [Software]. Available from <https://roboflow.com.computer.vision>.
- Fischler, M.A. and Bolles, R.C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, *Communications of the ACM*, 24(6), 381-395.
- Huang, H., Wang, H., Hu, Y.J., Li, C., and Wang, X. (2022). The development trends of existing building energy conservation and emission reduction—A comprehensive review, *Energy Report*, 8, 13180.
- Hough, P.V.C. (1962). Method and means for recognizing complex patterns, *U.S. Patent*, 3,069,654.
- Ishikawa, T., Kanai, S., and Date, H. (2021). Direct Generation of Cartesian Grid for As-built CFD Analysis from Laser-scanned Point Clouds, *Computer-Aided Design & Applications*, 18(6), 1341-1358.
- Jocher, G., et al. (2020). YOLOv5 by Ultralytics [Computer software]. Available at: <https://github.com/ultralytics/yolov5>
- Jung, J., Stachniss, C., Ju, S., and Heo, J. (2018). Automated 3D volumetric reconstruction of multiple-room building interiors for as-built BIM, *Advanced Engineering Informatics*, 38, 811-825.
- Mineo, C., Pierce, S.G., and Summan, R. (2019). Novel algorithms for 3D surface point cloud boundary detection and edge reconstruction, *Computational Design and Engineering*, 6(1), 81-91.
- Matsuba, K., Yabuki, N., Fukuda, T., and Hada, Y. (2023). A building information model for sensing and simulation toward net-zero energy building renovation, *Proceedings of the 5th International Conference on Civil and Building Engineering Informatics*, Bangkok, Thailand, pp.153-159.
- Patraean, V., Armeni, I., Nahangi, M., Yeung, J., Brilakis, I., and Haas, C. (2015). A state of research in automatic as-built modelling, *32nd International Symposium on Automation and Robotics in Construction*, 162-171.
- Qiu, Q., Zhou, X., Zhao, J., Yang, Y., Tian, S., Wang, J., Liu, J., and Liu, H. (2021). From sketch BIM to design BIM: An element identification approach using industry foundation classes and object recognition, *Building and Environment*, 188, 107423.
- Shi, J. and Tomasi. (1994). Good features to track, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, pp.1063-6919.
- Scaniverse (Version 2.0.0) [Software]. (2024). Niantic. <https://apps.apple.com/us/app/scaniverse-3d-scanner/id1541433223>
- Turner, E., Cheng, P., and Zakhor, A. (2015). Fast, automated, scalable generation of textured 3D models of indoor environments, *IEEE Journal of Selected Topics in Signal Processing*, 9(3), 409-421.
- Tran, H. and Khoshelham, K. (2020). Procedural Reconstruction of 3D Indoor Models from Lidar Data Using Reversible Jump Markov Chain Monte Carlo, *Remote Sensing*, 12(5), 838.
- Tang, S., Li, X., Zheng, X., Wu, B., Wang, W., and Zhang, Y. (2022). BIM generation from 3D point clouds by combining 3D deep learning and improved morphological approach, *Automation in Construction*, 141, 104422.
- Varajic. A. (2020). *As-builts explained and why you should consider it for your next project*. Retrieved from Cortland Design website: <https://www.cortlanddesign.com/blog/architecture-design/as-builts-explained-and-why-you-need-them>
- Varghese, V., Shajahan, D.A., and Nath, A. (2016). Building boundary tracing and regularization from LiDAR point cloud, *International Conference on Emerging Technological Trends (ICETT)*, Kollam, India, 7873645.
- Wu, X., Sahoo, D., and Hoi, S.C.H. (2020). Recent advances in deep learning for object detection, *Neurocomputing*, 396, 39-64.
- Xie, L. and Wang, R. (2017). Automatic indoor building reconstruction from mobile laser scanning data, *Remote Sensing and Spatial Information Sciences*, 42(2), 417-422.

VIRTUAL SCANNING METHOD BASED ON BIM AND RAY TRACING FOR OCCLUSION EVALUATION AND GEOMETRY EXTRACTION

Siwei Lin¹, Liping Duan², Bin Jiang³ and Jincheng Zhao⁴

1) Ph.D. Candidate, School of Ocean and Civil Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China. Email: linsiwei@sjtu.edu.cn

2) Ph.D., Assoc. Prof., State Key Laboratory of Ocean Engineering, School of Ocean and Civil Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China. Email: duanliping@sjtu.edu.cn

3) Ph.D. Candidate, School of Ocean and Civil Engineering, Shanghai Jiao Tong University, Shanghai 200240, China; China Ship building NDRI Engineering Co., Ltd., Shanghai, 200063, China. Email: 2021jiangbin@sjtu.edu.cn

4) Ph.D., Prof., State Key Laboratory of Ocean Engineering, School of Ocean and Civil Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China. Email: jczhao@sjtu.edu.cn

Abstract: This paper presents a virtual scanning method based on Building Information Model (BIM) which can simulate the real scan process and occlusion condition. The method converts the geometry of BIM into triangular mesh and point cloud. The multi-level data encoding is performed for the triangular mesh and point cloud. An octree is created based on point cloud to accelerate query speed. Two stage intersection tests (ray-octree leaf node and ray-triangle intersection tests) are organized to minimize computational cost. The scanning process of the scanner is parameterized and returns the nearest point with a component label along the emitted ray. The virtual scanning method is validated on a steel structure with over 800,000 triangular meshes., the calculations of over 100 million rays are completed within 10 minutes with serial computation. The visibility ratio of every component is obtained by compare the number of points between virtual scan component and complete component with the same point cloud spacing. Furthermore, the method is validated by comparing the consistency between the virtual scan cross-sections and real scan cross-sections. The real component axis extracted by the registration between virtual scan cross-sections and real scan cross-sections demonstrates better stability.

Keywords: Point cloud, Virtual scan, BIM, Ray tracing, Geometry extraction

1. INTRODUCTION

Laser scanning technology, with its high precision, high resolution, and efficiency, has been widely applied in geometric measurement and digital transformation (Wang and Kim 2019). In the field of civil engineering, compared to the single-point measurements performed by total stations, terrestrial laser scanning (TLS) can quickly acquire large amounts of high-precision point cloud data, making them more suitable for measurements in large-scale spaces and complex structures (Jiang et al. 2021). However, due to the complex arrangement of components and the terrain limitations of TLS, issues such as scanning occlusion and data loss are inevitable. The presence of occlusion significantly complicates object and state recognition, which are crucial for facility management and infrastructure maintenance (Agapaki and Brilakis 2020). Additionally, occlusion also impairs the accuracy of geometric extraction and model reconstruction (Yan and Hajjar 2022). Therefore, the degree evaluation of occlusion should be prioritized when performing scanning.

Previous studies have focused on scan planning to minimize the ratio of occlusion as much as possible (Chen et al. 2022; Li et al. 2022). However, visibility detection can be time-consuming and resource-intensive when dealing with large-scale or complex scenes (Noichl et al. 2024). Some simplified approaches often come at the cost of reduced accuracy in visibility detection. Virtual scanning technology simulates the process of a terrestrial laser scanner emitting laser rays and returning the coordinates of the nearest obstacles the laser reaches, which can maximally reflect the process of real-world scanning. Building Information Modeling (BIM) offers strong capabilities for 3D visualization and information integration in civil engineering sector (Liu et al. 2024). Combining virtual scanning technology with BIM makes it possible to simulate building laser scanning and occlusion assessment. This paper proposes a virtual scanning method based on BIM and ray tracing for large-scale scenes. The overall process of virtual scanning and applications are shown in Figure 1. The method converts the geometry of BIM to OBJ geometry and point cloud. Multi-level encoding of triangular meshes and point clouds is performed to support intersection detection and virtual scan point labeling. An octree is created to accelerate the query of ray-triangle intersections. Two-stage intersection detection: The intersection detection between rays and octree leaf nodes, as well as the intersection detection between rays and the triangles contained in the octree nodes, effectively reduces the query complexity. By setting virtual scanning parameters according to the scanner resolution and scanning process, the method can effectively simulate actual scans and multi-station scanning. In this paper, virtual scanning tests are conducted on a steel structure with over 800,000 triangular meshes. With serial computation of over 100 million rays, the calculations were completed within 10 minutes. Additionally, by analyzing the ratio of virtual scan point clouds to the full component point cloud, the occlusion degree of each component can be assessed. The paper compares the consistency between the virtual scan and real scan cross-sections, validating the correctness of the method. Furthermore, the component axis lines extracted from the

registration of virtual scan cross-sections and real scan cross-sections demonstrate better stability.

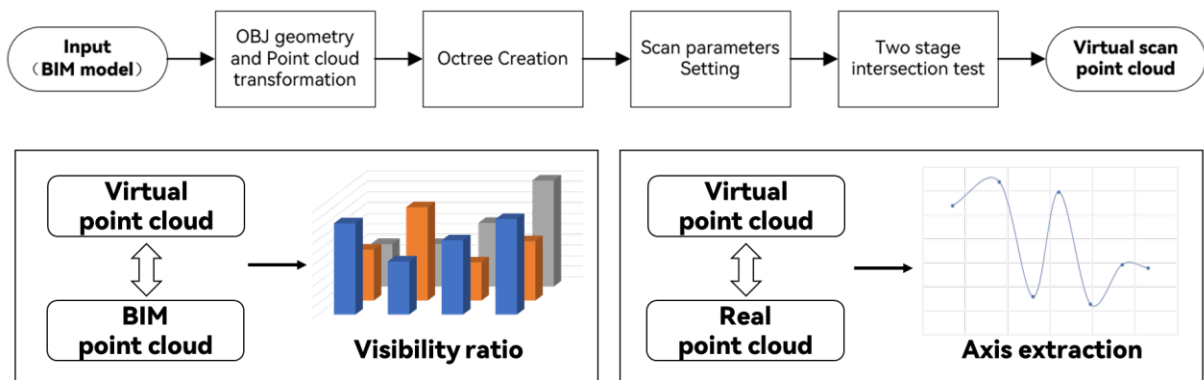


Figure 1. Virtual scan process and applications

The contributions of this method are as follows:

- (1) An efficient virtual scanning method based on BIM models. The result of the virtual scan is a point cloud model with component labels and occlusion considerations, which can be used for the creation of deep learning datasets.
- (2) An evaluation method for the occlusion degree based on the virtual scanning results.
- (3) A more general method for applying occlusion-aware virtual scan cross-sections to the extraction of actual component geometric axis lines.

The rest of this paper is organized as follows. The method of virtual scanning is introduced in Section 2. In Section 3, the proposed approach is validated to compare the virtual scan with the real scan of a steel structure. In Section 4, some noteworthy remarks and limitations of the method are presented. The conclusions are summarized in section 5.

2. METHOD

2.1 Data Transformation and Encoding

The first step of the virtual scanning process is data transformation and encoding. First, the geometric information from the BIM model is extracted and converted into the OBJ triangular mesh format. Then, the triangular mesh geometry is transformed into point clouds. The purpose of this conversion is to: (1) enable the construction of the octree data structure, and (2) determine the intersection relationship between the triangular mesh patches and the octree leaf nodes (Park et al. 2021).

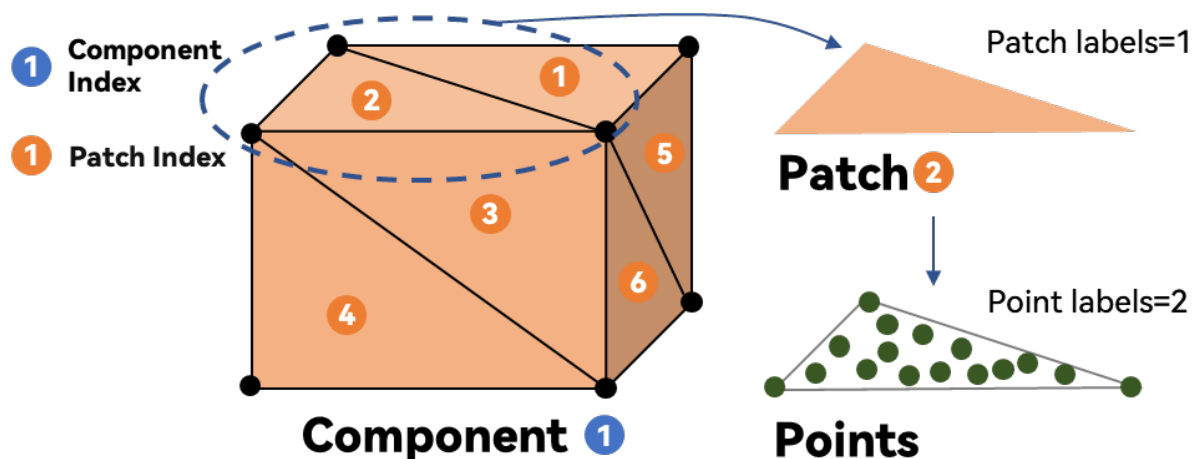


Figure 2. Multilevel Data Encoding

The density of the sparse point cloud generated from the triangular mesh should be set to at least half of the octree resolution. The encoding of each point cloud is the index of the triangular patches it belongs to, and the encoding of each triangular patches corresponds to the index of the components to which the patch belongs. The multilevel data encoding is shown in Figure 2.

2.2 Octree Creation and Ray Intersection Tests

Due to the large-scale and extensive nature of many scanning tasks, the number of triangular patches in a structural BIM model often reaches hundreds of thousands or even millions. As a result, performing intersection detection through brute-force traversal is not practical. The octree, as a spatial partitioning data structure, recursively divides three-dimensional space into eight subspaces, which helps optimize both data storage and spatial query efficiency. The conversion of the sparse point cloud into the octree data structure is illustrated in Figure 3.

Simulated laser scanning primarily involves two types of intersection detection: one is the intersection detection between the ray and the octree leaf nodes (AABB bounding boxes) as shown in Figure 4, and the other is the intersection detection between the ray and the triangles. The intersection detection between the ray and the octree leaf nodes is performed using a top-bottom approach (Revelles et al. 2000). Once the intersection between the laser ray and the octree leaf node is detected, the point cloud contained in that leaf node can be queried. Based on the triangular patch encoding carried by the point cloud, further intersection detection between the ray and the triangular patches can be performed. The intersection detection between the ray and the triangular patches employs the Möller-Trumbore algorithm (Möller and Trumbore 1997). When $t > 0$, $u \geq 0$, $v \geq 0$ and $u + v \leq 1$, the ray is determined to intersect with the triangular patch. The smallest t among all intersecting patches is selected, representing the point where the ray first strikes the BIM model.

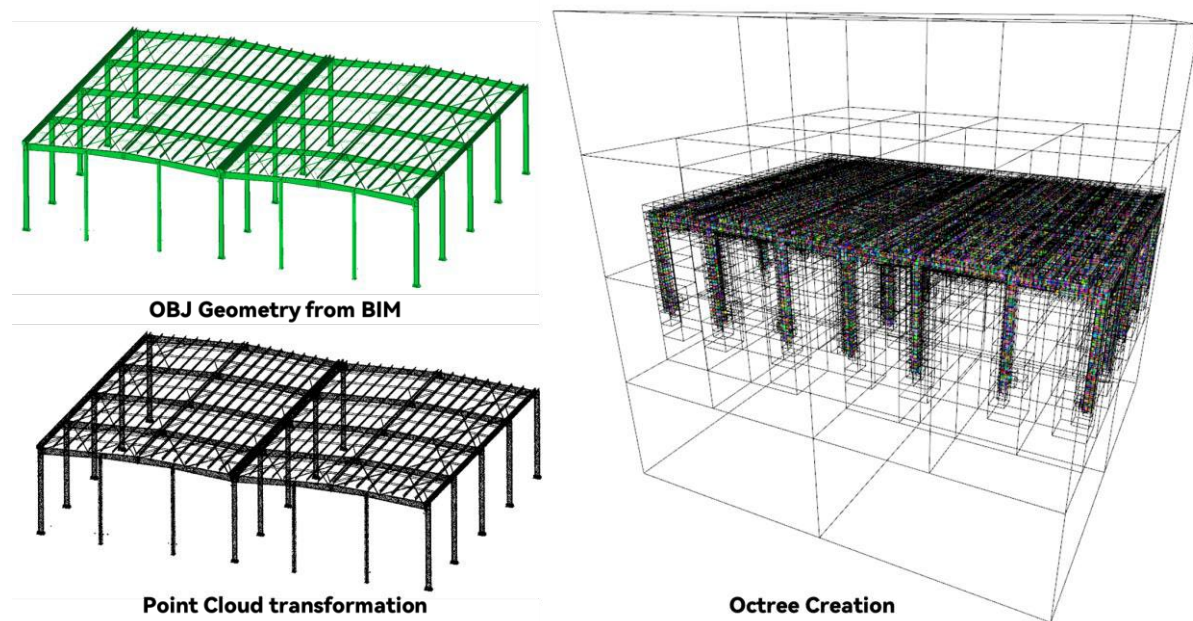


Figure 3. Data transformation and Octree creation

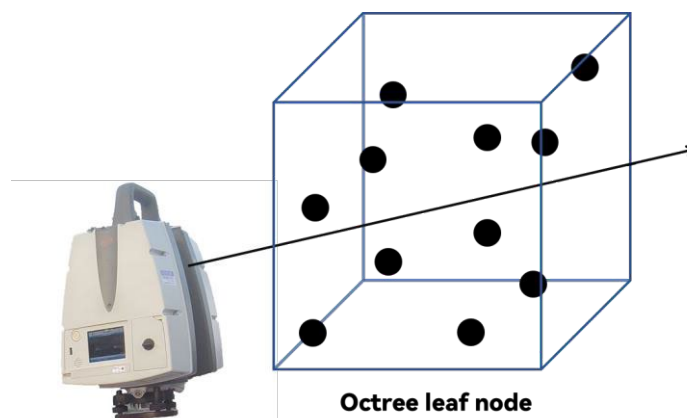


Figure 4. Ray and Octree LeafNode Intersection Diagram

2.3 Virtual Scanning Simulation

The scanning process of a terrestrial laser scanner starts from a coordinate point in space, rotating 360° in the horizontal direction while rotating 360° in the vertical direction. Taking the Leica P40 as an example, the

resolution needs to be set before scanning, such as 3.1mm@10m, which means the point cloud density is 3.1mm at a distance of 10 meters. Therefore, the angular change for each laser is determined by this resolution.

$$\alpha = \arctan\left(\frac{0.0031}{10}\right)$$

Laser direction vector

$$\begin{cases} x = \cos(\theta)\cos(\varphi) \\ y = \sin(\theta)\cos(\varphi) \\ z = \sin(\varphi) \end{cases}$$

After calculating the intersection value t of the intersecting triangle, the scan point can be obtained by $O + t \cdot D$, where O is the origin of the ray and D is the direction vector. Additionally, based on the encoding of the component where the intersecting triangle is located, the corresponding label can be assigned to the scan point cloud.

3. RESULTS

3.1 Virtual Scanning Results

The method proposed in this paper is validated on a two-span steel structure. The geometric model extracted from the BIM is shown in Figure 3, where the component, patch, and vertex count statistics can be found in Table 1. During the real scanning process, four scan stations were conducted and registered for the structure, with the real scan results and scanner positions shown in Figure 5. The specific scanner coordinates are provided in Table 2.

Num	Number
Components	3714
Patches	837146
Vertices	440200

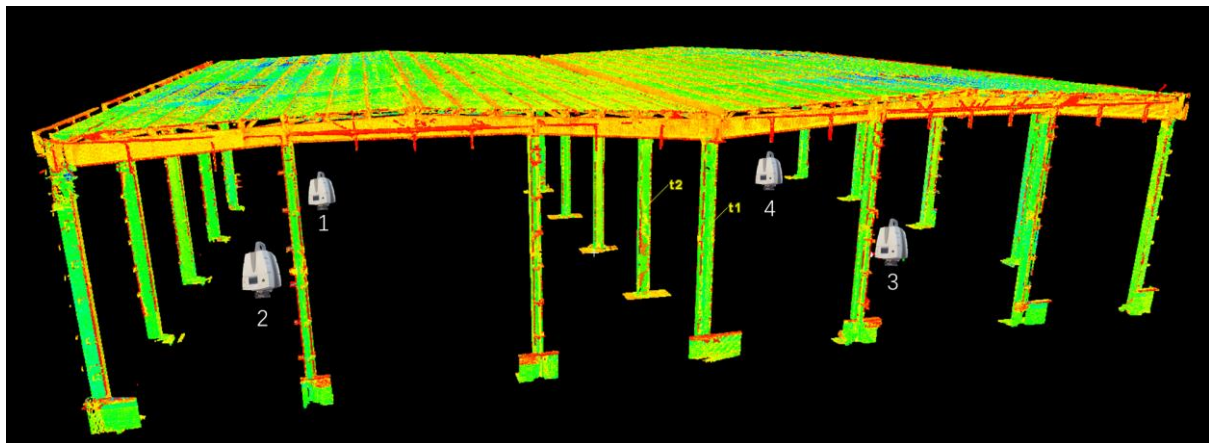


Figure 5. Real scan results and scanner positions

Virtual scanning was performed based on the coordinates of the scan stations using the method proposed in this paper. The calculations were implemented in C++ under serial computation and with a scanning resolution of 3.1mm@10m. The edge length of the octree leaf nodes is approximately 0.19m, and each scan station calculated over 100 million rays. The scanning time and point cloud quantity for each station are shown in Table 3. The scanning time for all stations is less than 10 minutes, and the point cloud of each station contained over 18 million points. The scan results are shown in Figure 6, where different colors represent different component labels. It can be observed that the scanning results vary across different stations, with significant differences in the visibility of columns between scan1 and scan4, where some columns are clearly occluded.

Origin	X/m	Y/m	Z/m
1	30.550564	41.752708	1.676277
2	13.477213	43.319534	1.571071
3	8.175106	10.082193	1.642238
4	29.850962	6.519911	1.704301

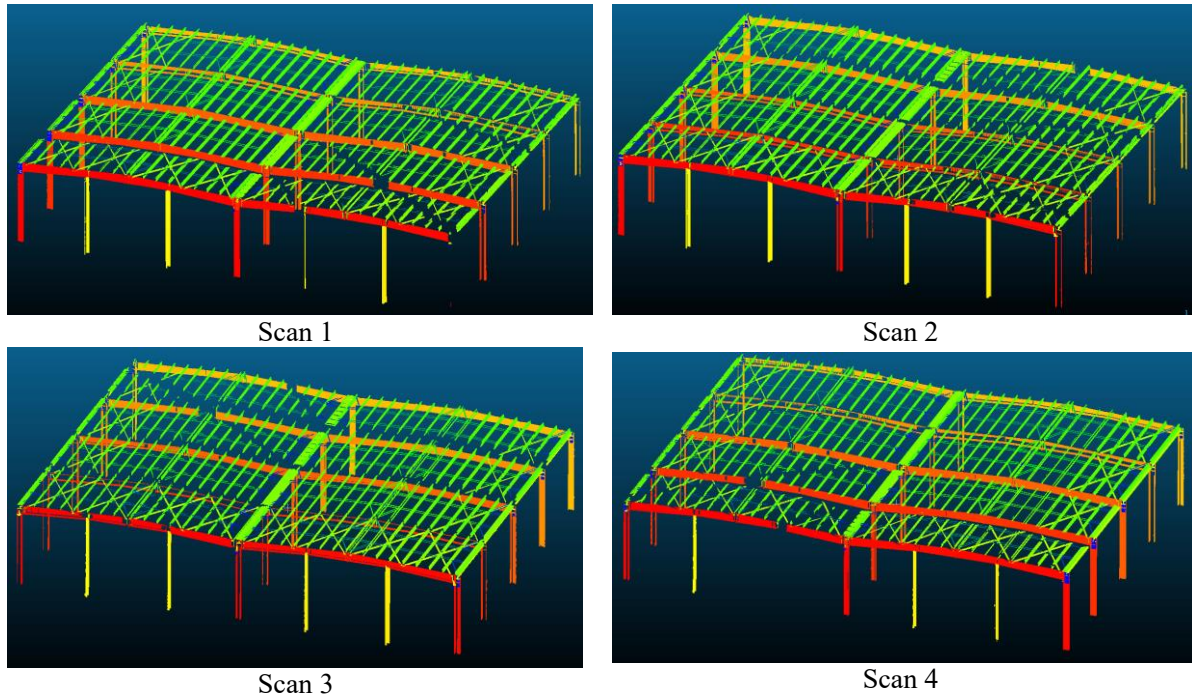


Figure 6. Virtual scanning results of different scan stations

Table 3. Coordinate of scan stations

Scan Station	Points Number	Time/s
1	18466676	414.303
2	21112038	475.774
3	23222200	526.679
4	18874496	430.148

The virtual scanning results from the four scan stations can be directly merged without the need for registration, as the point clouds from all four scans are in a unified coordinate system. After performing the point cloud merging, the point cloud was downsampled to a 5mm spacing. Based on the point cloud labels and the corresponding GUID numbers, the entire point cloud is segmented into different components. By comparing the number of virtual scanning component point clouds with the complete component point clouds at the same point cloud spacing, the visibility ratio of each component in the virtual scan can be assessed. The figure 7 shows the visibility ratio of the main components of the steel structure. Some areas with relatively low visibility are highlighted in the image, and the reasons for their poor visibility vary.

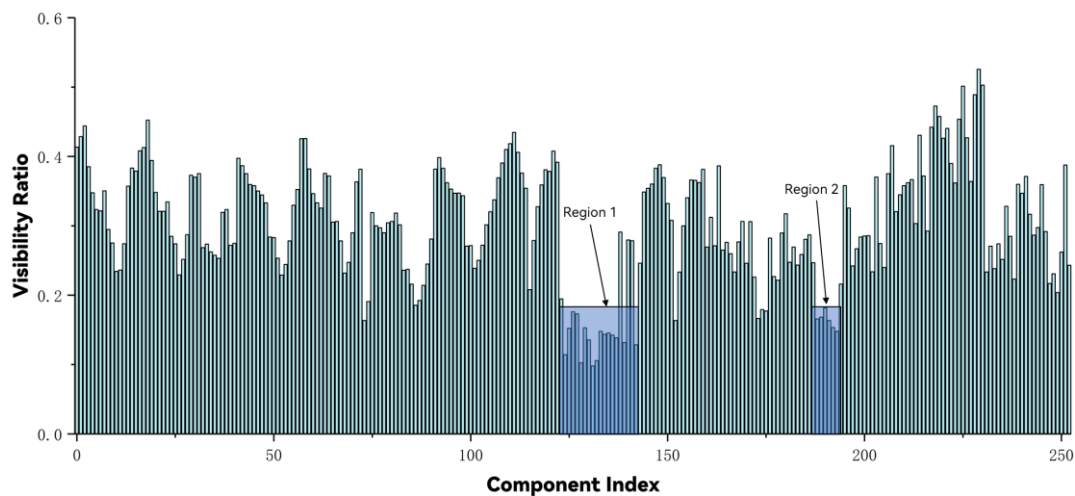


Figure 7. Visibility ratio of different components

Region 1 mainly consists of hat-shaped section components, whose low visibility is primarily due to their close proximity to adjacent channel steel components, resulting in significant occlusion, as shown in Figure 8. Region 2 mainly consists of circular tube components, and their poor visibility is due to the fact that the interior of the tubes is inherently not visible, leading to a lower overall visibility ratio.

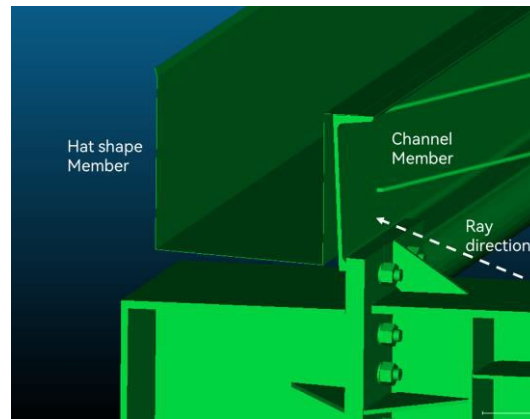


Figure 8. The layout of hat-shaped section components

3.2 Section Occlusion Comparison and Axis Extraction

The paper selects components with different cross-sections and compares the deviations between virtual scanning results and real scanning results, as shown in Figure 9. The figure shows the virtual and real scanning results of components with circular, channel, and I-beam cross-sections. It can be observed that the visibility of the virtual and real scanned cross-sections is generally consistent, further proving the accuracy of the virtual scanning method. Additionally, the registration between the virtual and real scanned cross-sections can facilitate the extraction of the component's axis.

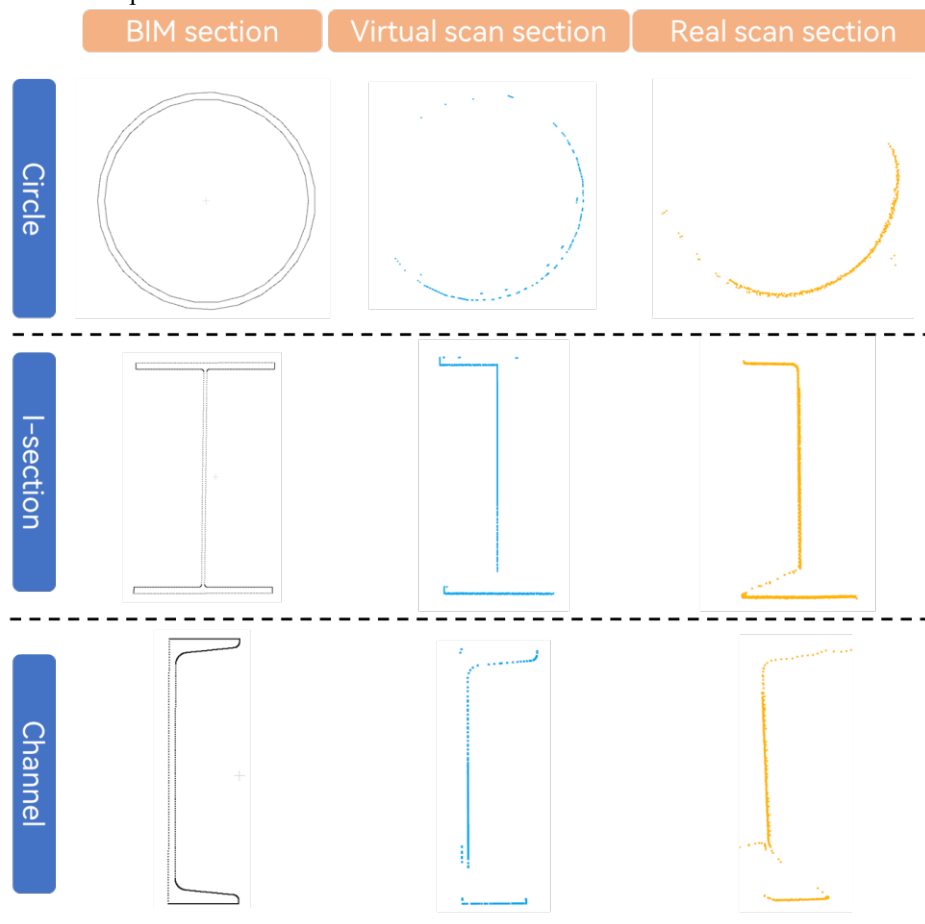


Figure 9. The comparison between virtual scan section and real scan section

Taking the I-beam component as an example, previous methods (Lin et al. 2024) involved slicing both the complete component and the real scanned component, and then aligning the full cross-section and the real scanned cross-section at corresponding positions to extract the axis. However, the extracted axis exhibits fluctuations in the plate thickness direction, as shown in Figure 10. This issue has been explained and addressed in previous work. In contrast, this paper presents a more general approach that applies to components with any cross-section by eliminating occluded point clouds through virtual scanning before performing the registration. By aligning the virtual scanned cross-section with the real scanned cross-section, the fluctuations in the extracted axis are significantly reduced, and the accuracy of the component axis extraction is improved.

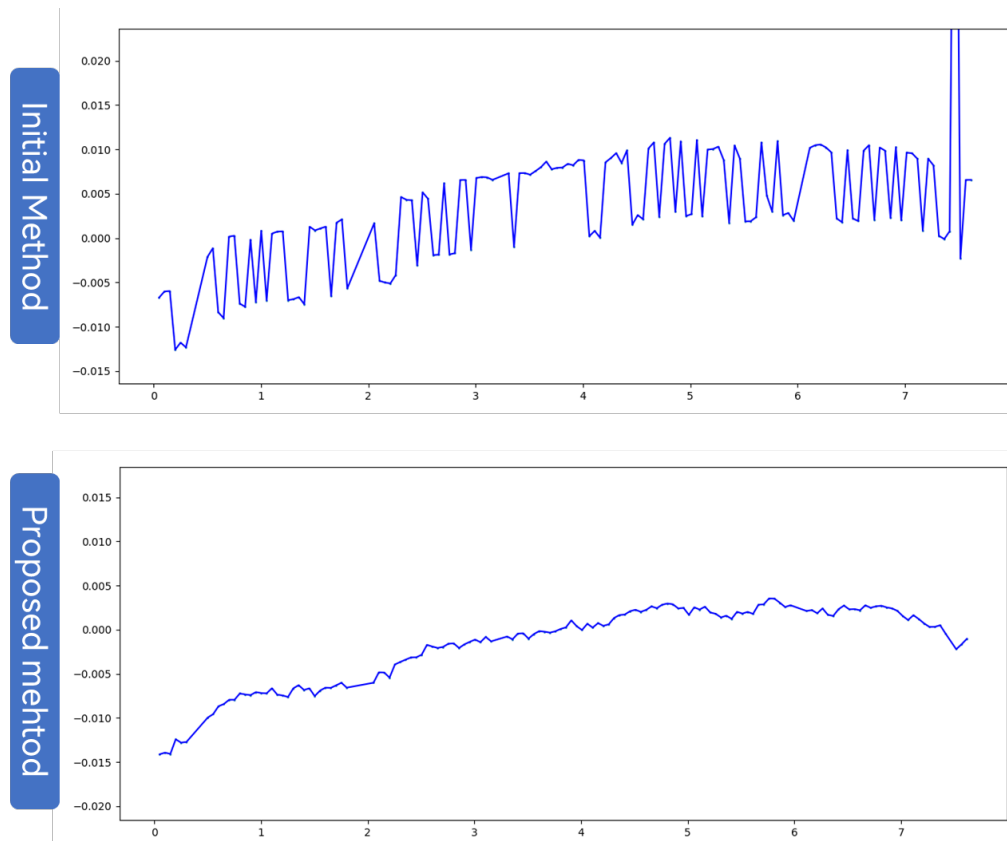


Figure 10. The comparison between initial method and proposed method considering occlusion

4. DISCUSSION

The virtual scanning method proposed in this paper has the potential for broader applications based on the discussions presented. For instance, the point cloud model generated through virtual scanning is labeled and can be directly used as a dataset for deep learning. Additionally, the point cloud considering occlusion is more aligned with the actual situation compared to the complete BIM model point cloud. The deep learning model trained based on virtual scan point cloud is expected to perform better in practical applications.

Furthermore, since virtual scanning essentially simulates the actual scanning process, its visibility is also largely consistent with real-world scanning, making it a more effective evaluation metric for scan planning. However, the current computational efficiency remains one of the obstacles to using virtual scanning for optimizing scan planning. Future work could focus on parallel computing to further assess the simulation efficiency of the proposed virtual scanning method in large-scale BIM models.

5. CONCLUSIONS

This paper introduces a virtual scanning method based on BIM and ray tracing, designed to simulate the real-world scanning process and account for occlusion effects. The method first converts the BIM geometry into triangular meshes and point clouds and applies multilevel data encoding to the triangular meshes and point clouds. An octree structure is generated from the point cloud to accelerate query performance. To reduce computation cost, two-stage intersection tests are employed: ray-octree leaf node and ray-triangle intersection tests. The scanning process is parameterized to return the nearest point with a corresponding component label along the emitted ray. The effectiveness of the virtual scanning method is validated using a steel structure containing over 800,000 triangular meshes, with calculations for more than 100 million rays completed in under 10 minutes using serial

computation. The visibility ratio of each component is determined by comparing the number of points in the virtual scan component to those in the complete component, both with the same point cloud spacing. Additionally, the method's correctness is further validated by comparing virtual scan cross-sections with real scan cross-sections. The real component axis, obtained through registration between virtual and real scan cross-sections, shows better stability.

The limitation of the method is still time-consuming for scan planning. The effectiveness of the virtual scanning will be further evaluated through parallel computing. The intersection tests between triangle patches and octree leaf nodes can be introduced to further improve the virtual scanning accuracy instead of converting the sparse point cloud. In the future, the component axis extracted can be applied to the BIM model update considering the cases of curves which can reflect a more realistic physical scenario.

ACKNOWLEDGMENTS

The authors would like to appreciate the financial support of the work addressed in this paper from the Science and Technology Commission of Shanghai Municipality (Grant No.: 21DZ1204600).

REFERENCES

- Agapaki, E., and I. Brilakis. 2020. "CLOI-NET: Class segmentation of industrial facilities' point cloud datasets." *Advanced Engineering Informatics*, 45: 101121. <https://doi.org/10.1016/j.aei.2020.101121>.
- Chen, Z., W. Zhang, R. Huang, Z. Dong, C. Chen, L. Jiang, and H. Wang. 2022. "3D model-based terrestrial laser scanning (TLS) observation network planning for large-scale building facades." *Automation in Construction*, 144: 104594. <https://doi.org/10.1016/j.autcon.2022.104594>.
- Jiang, F., L. Ma, T. Broyd, and K. Chen. 2021. "Digital twin and its implementations in the civil engineering sector." *Automation in Construction*, 130. <https://doi.org/10.1016/j.autcon.2021.103838>.
- Li, D., J. Liu, Y. Zeng, G. Cheng, B. Dong, and Y. F. Chen. 2022. "3D model-based scan planning for space frame structures considering site conditions." *Automation in Construction*, 140: 104363. <https://doi.org/10.1016/j.autcon.2022.104363>.
- Lin, S., L. Duan, B. Jiang, J. Liu, H. Guo, and J. Zhao. "Scan vs. BIM: Automated Geometry Detection and BIM Updating of Steel Framing through Laser Scanning." (Under review)
- Liu, J., L. Duan, S. Lin, J. Miao, and J. Zhao. 2024. "Concept, Creation, Services and Future Directions of Digital Twins in the Construction Industry: A Systematic Literature Review." *Arch Computat Methods Eng.* <https://doi.org/10.1007/s11831-024-10140-4>.
- Möller, T., and B. Trumbore. 1997. "Fast, Minimum Storage Ray-Triangle Intersection." *Journal of Graphics Tools*, 2 (1): 21–28. <https://doi.org/10.1080/10867651.1997.10487468>.
- Noichl, F., D. D. Lichti, and A. Borrmann. 2024. "Automating adaptive scan planning for static laser scanning in complex 3D environments." *Automation in Construction*, 165: 105511. <https://doi.org/10.1016/j.autcon.2024.105511>.
- Park, S., S. Ju, S. Yoon, M. H. Nguyen, and J. Heo. 2021. "An efficient data structure approach for BIM-to-point-cloud change detection using modifiable nested octree." *Automation in Construction*, 132: 103922. <https://doi.org/10.1016/j.autcon.2021.103922>.
- Revelles, J., C. Ureña, and M. Lastra. 2000. "An Efficient Parametric Algorithm for Octree Traversal."
- Wang, Q., and M.-K. Kim. 2019. "Applications of 3D point cloud data in the construction industry: A fifteen-year review from 2004 to 2018." *Advanced Engineering Informatics*, 39: 306–319. <https://doi.org/10.1016/j.aei.2019.02.007>.
- Yan, Y., and J. F. Hajjar. 2022. "Geometric models from laser scanning data for superstructure components of steel girder bridges." *Automation in Construction*, 142: 104484. <https://doi.org/10.1016/j.autcon.2022.104484>.



Part5

**Advancing Construction
through Robotics and Human Ma-
chine Collaboration**

AN EXPERIMENTAL STUDY FOR MULTI-ROBOT COORDINATION IN MULTI-STORY BUILDING CONSTRUCTION SITES

Haojun Luo¹ and Yantao Yu *²

1) Ph.D. Student, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: hluoaw@connect.ust.hk

2) Ph.D., Asst. Prof., Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: ceyantao@ust.hk

Abstract: Robots offer a promising solution to relieve workers from physical demanding tasks and improve safety and productivity in construction. It is critical that the robots on construction sites are coordinated effectively. However, most multi-robot coordination algorithms are designed for planar areas, neglecting the multi-story nature of building construction sites. It is still unclear how construction robots should be coordinated given the constraints of elevators while adhering to construction schedules. To fill the gap, this paper introduced the deployment of commonly used elevator algorithms and robot target allocation strategies in a multi-story construction simulation environment. Through a series of group experiments conducted in a simulated multi-story construction environment, we evaluated the performance of these algorithms and examined the characteristics of robot-elevator coordination. The results reveal that while existing algorithms with strong generalization capabilities are useful, they may be less effective in specialized scenarios like multi-story construction. This research contributes valuable insights into the future of automation in construction, paving the way for enhanced integration of robotic systems and elevator operations.

Keywords: Construction robot, Multi-robot coordination, LOOK Algorithm, Nearest neighbor, Pickup and delivery problems.

1. INTRODUCTION

Robots are increasingly recognized as a transformative force in the construction industry, offering solutions that alleviate physically demanding tasks while enhancing safety and productivity (Aghimien et al. 2020). Various construction robots have been developed for executing specific construction tasks, such as tiling (Le et al. 2020), plastering (Wang et al. 2024), painting (Seriani et al. 2015) and so on. The coordination of this heterogeneous robotic system is essential for enhancing efficiency and effectiveness in construction projects. However, most multi-robot coordination algorithms are designed for planar areas, neglecting the multi-story nature of building construction sites. It is still unclear how construction robots should be coordinated given the constraints of elevators while adhering to construction schedules.

There are two main characteristics of multi-story construction sites. In a multi-story construction project, there are several apartments on each floor, and each apartment has multiple processes that need to be completed, requiring careful management of numerous tasks to ensure that the entire project progresses smoothly. It is a dynamic and gradual process. Additionally, elevators play a critical role in transporting robots between floors in multi-story construction projects. Ineffective use of elevators can lead to delays in robot deployment across different construction sites, negatively impacting overall project timelines. Therefore, it is essential to develop a multi-robot coordination algorithm that considers elevator constraints and can complete all multi-story building construction tasks efficiently.

In multi-story construction sites, coordinating the movement of elevators and robots presents unique challenges that are not adequately addressed by existing algorithms designed for other scenarios, such as warehouses (Nielsen et al. 2016) or assembly lines (Kousi et al. 2019; Kousi et al. 2016). These algorithms typically do not account for the vertical movement requirements of robots, which is crucial in a multi-story environment. The complexity of determining where robots should go and how elevators should move complicates the coordination process, highlighting a significant gap in current research on multi-agent systems.

Therefore, we conducted an experimental study on multi-robot coordination in a multi-story building construction site. In our exploration of coordinating elevators and robots in multi-story construction sites, we try to deploy commonly used elevator algorithms, specifically the **LOOK algorithm** and the **nearest neighbor method**. These algorithms are selected for their strong applicability to our scenario due to their effectiveness in managing requests and optimizing movement. For the robot target allocation, we investigated two strategies: the **first-come-first-served method (FCFS)** and the **nearest-served method (NS)**. By combining these robot allocation strategies with the elevator algorithms, we form multiple experimental groups to evaluate their performance in the proposed multi-story environment. All algorithms are tested through a simulation of interior finishing work in a ten-story residential building. The experimental results provide valuable insights for future research on multi-agent coordination including elevators and robots, and the practical deployment of robots in multi-story construction scenarios.

2. LITERATURE REVIEW

The problem of finding the best elevator planning and robot target allocation in construction projects is similar to the Pickup and Delivery Problems (PDPs). In PDPs, goods or passengers must be transported from different departure locations to different destinations (Toth and Vigo 2014). In the proposed scenario, robots need to take the elevator from different departure floors to different destination floors depending on the task. According to the decision framework being considered and the availability of information, existing studies on these problems can be divided into two categories: static approaches and dynamic approaches. This section will first review existing related research, and then identify research gaps.

Static approaches to PDPs operate under the assumption that all relevant information, such as demand, vehicle availability, and travel times, is known in advance (Berbeglia et al. 2007). Existing research can be summarized into mathematical methods and heuristic methods.

Mathematical methods for solving static PDPs typically involve formulating the problem using optimization models that aim to minimize costs or travel times. (Berbeglia et al. 2007) presents a comprehensive classification of static PDPs and discusses various mathematical formulations, including mixed-integer linear programming models. These models are often used to derive optimal routes while considering constraints such as time windows and vehicle capacities. Once accurate mathematical models are established, solving these models can yield the optimal solution. Techniques (Cordeau 2006; Garaix et al. 2011; Qu and Bard 2015) for solving these models are developed mainly based on upon the concept of branch-and-bound (B&B). B&B is designed for general discrete and combinatorial optimization problems. However, in construction scenarios, the robot's demand for elevators depends on actual construction efficiency and environmental factors, which is non-deterministic. It is difficult to establish an accurate mathematical model to describe such a dynamic scenario.

Heuristic methods have been shown to be effective and efficient compared to mathematical methods (Ho et al. 2018). In order to find the optimal solution, B&B usually enumerates a large number of solutions. In the worst case, the computation time of the B&B program may increase exponentially. On the contrary, although the heuristic method cannot guarantee optimality, it can always find an acceptable solution in a shorter time. Common methods include tabu search (Cordeau and Laporte 2003), simulated annealing algorithm (Mauri, Antonio, and Lorena 2009), genetic algorithm (Jorgensen, Larsen, and Bergvinsdottir 2007), and hybrid heuristic algorithm (Berbeglia, Cordeau, and Laporte 2012). In the heuristic method, instead of building an accurate mathematical model, feasible solutions are constructed based on the constraints of the environment, and then the existing solutions are improved through heuristic rules to make the new solutions better and better. However, in order to calculate the fitness function for the heuristic algorithm to improve current solutions, the distribution of all demands and the benefits of completing the demands must be known in advance. In the proposed scenario, the demands for the robots to cross floors is dynamically updated over time.

Dynamic approaches typically provide a solution strategy rather than a deterministic static plan. The strategy uses the information revealed to specify what actions must be performed over time. Currently there are few studies on dynamic methods. A basic and commonly used strategy is to adapt an algorithm that solves a static version of the problem (Berbeglia, Cordeau, and Laporte 2010). There are two ways to implement this strategy. One is to treat all the information at the current time as a static problem to be solved whenever new information appears (Berbeglia, Cordeau, and Laporte 2010). The disadvantage of this strategy is that it takes too much time to solve a complete static problem every time. The other is to solve the static problem once at the beginning using the known information to get a feasible solution, and then use heuristic methods to update the feasible solution as time goes by and new information is obtained (Berbeglia, Cordeau, and Laporte 2010). The second method is more common, but still requires time and memory to maintain and update existing solutions. Most importantly, existing solutions focus on improving the path of the vehicle. Construction is an irreversible process, and constantly changing strategies during construction is not a good option.

Coordinating robot teams and elevators in multi-story construction projects is challenging due to the nature of multi-story buildings. The literature review to date shows that there is a lack of an effective method to coordinate robot teams and elevators in multi-story construction sites. Therefore, we explored the combination of commonly used elevator planning algorithms and robot target allocation algorithms to see how they perform on multi-story construction sites and further explore the characteristics of multi-agent coordination on multi-story construction sites.

3. METHOD

This study focuses on the coordination of elevators and robots in multi-story construction sites, aiming to optimize their interactions to improve operational efficiency. In the proposed scenario, we deployed two types of elevator algorithms - LOOK and Nearest Neighbor. LOOK is a classic disk scanning algorithm. For the nearest neighbor method, in addition to the traditional distance-based nearest neighbor algorithm (DNN), we also improved the metric for evaluating "nearest" and proposed the improved nearest neighbor algorithm (INN). For the robot target allocation algorithm, the FCFS method and the NS method were applied. By combining these

robot allocation strategies with the elevator algorithm, we formed multiple experimental groups to evaluate their performance in the proposed multi-story environment. This section is structured around the design of the simulation environment, the implementation of elevator and robot algorithms.

3.1 Simulation Environment

We created a simulated multi-story construction site based on C++. The environment consists of ten floors, with two elevators and multiple robots responsible for construction and delivery. The simulation framework is designed to generate requests for elevator service and robot allocation in a construction scenario to fully evaluate the algorithm under different conditions. Details are shown in Figure 1.

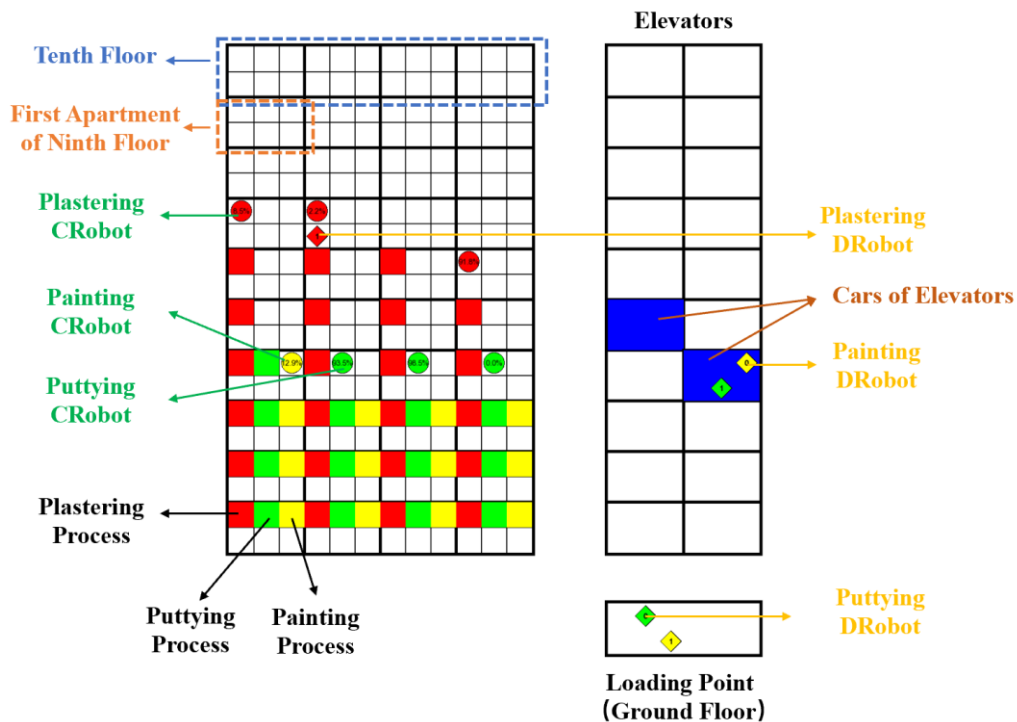


Figure 1. The composition of the simulation environment. (CRobot means construction robot and DRobot means delivery robot.)

Here comes the workflow of the environment. We consider three construction processes, plastering, puttying and painting. They follow a flow-through construction method from bottom to top. The construction robot of the next process will enter the site after the previous process is completed. Each process is handled by a construction robot. When the construction robot runs out of materials, it will suspend the operation and issue a material demand. After the material demand is received by the idle delivery robot, it will deliver the materials to the construction robot. After receiving the materials, the construction robot continues to work until the process is completed, and then goes to the next construction. The delivery robot should return to the loading point to replenish material after completing the delivery. It is worth noting that the construction robots may work on different floors at the same time. Material requirements may come from multiple floors. All robots rely on elevators to go to other floors. As for elevators, considering the size of construction robots in reality, each elevator is limited to carrying two robots. The real-time location of robots and elevators is known in the simulation, which can be achieved by existing technology i.e., RFID systems (Motroni, Buffi, and Nepa 2021).

In the above environment, there are three main cross-floor activities that rely on elevators. One is that the construction robot goes from the current floor to another floor to work, the second is that the delivery robot delivers materials from the loading point to the construction robot, and the last is that the delivery robot returns to the loading point after delivering materials on the current floor. The entities involved are the construction robot, the delivery robot, and the elevator. Figure 2 shows the states of these three entities. For an elevator, it can be in the state of going up, going down, and stopping. For the delivery robot, there are four places it can be located, namely the loading point, the apartment, inside the elevator, and the place waiting for the elevator. It may also be on the ways of loading point – elevator and apartment – elevator. A construction robot is similar. It could be in an elevator, in an apartment, or waiting for an elevator. Also, it could be on the way between the elevator-apartment.




Elevators 	<ul style="list-style-type: none"> • Stop • Going up • Going down
Delivery robot 	<ul style="list-style-type: none"> • AL: at loading point • L2E: on the way from loading point to elevator • WE: waiting elevator • IE: in the elevator • E2A: on the way from elevator to apartment • AT: at task apartment • A2E: on the way from apartment to elevator • E2L: on the way from elevator to loading point
Construction robot 	<ul style="list-style-type: none"> • Flow construction, red→green→yellow, Bottom-up • WE: waiting elevator • IE: in the elevator • E2A: on the way from elevator to apartment • AT: at task apartment • A2E: on the way from apartment to elevator

Figure 2. The states of Elevators, Delivery Robot, and Construction Robot

3.2 Algorithms

After the simulation environment is established, the algorithms are applied to it. The goal of the study is to coordinate multiple agents including robots and elevators to finish the whole construction project. The coordination algorithm consists of two parts, one is the elevator scheduling algorithm, and the other is the robot target allocation algorithm. They will be explained in detail below.

3.2.1 Robot Target Allocation Algorithms

This part is to assign targets to available delivery robots. Specifically, when a delivery robot has multiple potential targets, which one should be assigned to it to determine the target floor it needs to go to. Available construction robots will cross floors from bottom to top in a flow-line construction manner according to the settings of the simulation environment, without the need to design an algorithm for allocation. Two strategies were employed:

First-Come-First-Served: This method allocates tasks to robots based on the order in which requests are received, ensuring a straightforward allocation process. Figure 6 shows an example of FCFS method. There is an idle robot 1 that receives a task from the 7th floor. It accepts the task and takes the elevator to the 7th floor. During this process, there is a task from the 3rd floor, but according to the FCFS principle, the robot will ignore this task and go to the 7th floor first to complete the task.

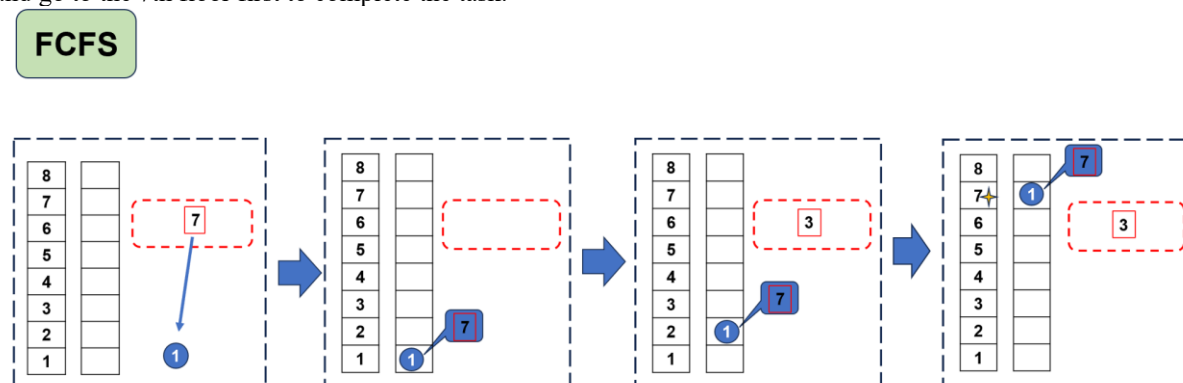


Figure 3. An example of FCFS method.

Nearest-Served Method: In this method, delivery robots of the same type share all potential targets. The delivery robot in the elevator will put all potential targets in a target set in chronological order. Once the elevator reaches a floor in the target floor set, it will assign the delivery robot that has that floor in its target floor set to serve that floor. Then the target floor sets of all delivery robots are updated. This approach prioritizes allocating the nearest target to the delivery robot, not necessarily the earliest target. Figure 7 shows an example of NS method. The same task situation as FCFS is used here. The difference is that when a task comes from the 3rd

floor, the task will be added to the robot's target floor set. The robot passes the 3rd floor first with the updated target floor set, so it will complete the task on the 3rd floor first.

NS

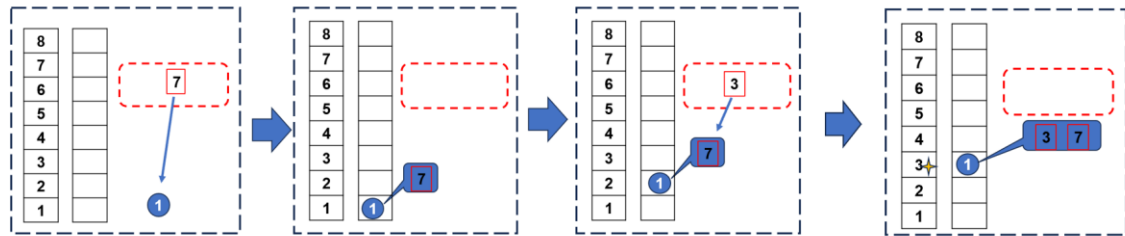


Figure 4. An example of NS method.

The difference between FCFS and NS is that the former focus on solving the material needs that are called early, while the latter solves the most recent material needs.

3.2.2 Elevator Scheduling Algorithms

The elevator scheduling algorithm is responsible for managing elevator requests efficiently. In this study, we implemented two primary algorithms:

LOOK Algorithm: This classic disk scanning algorithm operates by moving the elevator in one direction, serving requests along its path before reversing direction. It effectively minimizes wait times by ensuring that all requests in one direction are addressed before changing course. Since the real-time positions of the robot and the elevator are known, two situations may occur. One is that the robot sends an elevator request when it is waiting for the elevator. In this way, the robot will get on the elevator as soon as the elevator arrives, but the robot needs to wait for the elevator to arrive. The other is that the robot sends an elevator request when it is on the way to the elevator. When the elevator arrives but the robot has not arrived yet, the elevator needs to wait for the robot to arrive and get on the elevator before leaving. Figure 3 shows the diagram of the LOOK algorithm with these two situations.

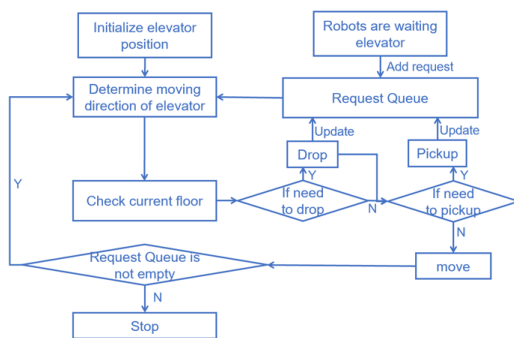


Figure 5-1. The diagram of the LOOK algorithm when the robot is waiting for the elevator

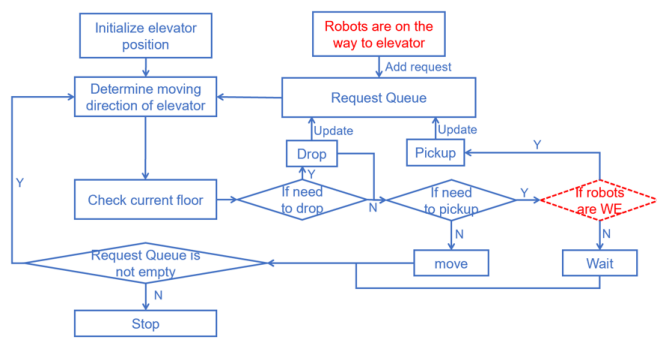


Figure 5-2. The diagram of the LOOK algorithm when the elevator is waiting for the robot

Figure 5. The diagram of the LOOK algorithm

Nearest Neighbor Method: This method focuses on fulfilling requests based on proximity. We utilized two variations:

Distance-Based Nearest Neighbor (DNN): This traditional approach allocates requests based on the shortest distance from the elevator's current position. Figure 4 shows the diagram of DNN.

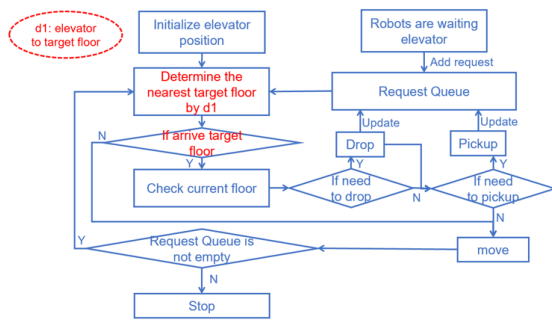


Figure 6-1. The diagram of DNN when the robot is waiting for the elevator

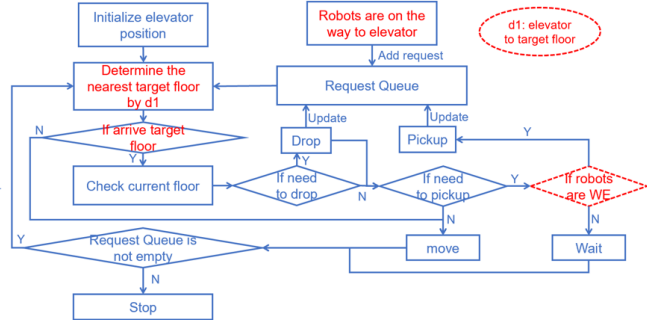


Figure 6-2. The diagram of DNN when the elevator is waiting for the robot

Figure 6. The diagram of DNN

Improved Nearest Neighbor (INN): This enhanced version refines the evaluation metric for "nearest" requests by incorporating additional factors such as current wait times and the elevator's position relative to pending requests. This adjustment aims to improve responsiveness and overall service efficiency. Figure 5 shows the diagram of INN.

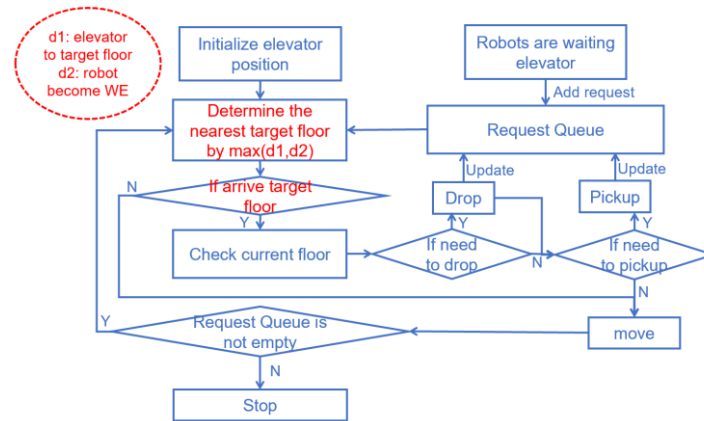


Figure 7. The diagram of INN

As can be seen from Figures 4 and 5, the main difference between DNN and INN lies in the definition of "nearest". In DNN, the nearest request is selected as the target floor based on the distance ($d1$) between the request and the current floor of the elevator. In INN, the nearest target is evaluated based on the maximum value of the distance ($d1$) and the actual waiting time ($d2$). It is worth noting that in the case of the robot waiting for the elevator, this INN will collapse into a DNN, because the actual waiting time is equal to the distance at this time.

3.2.3 Multi-agent Coordination Algorithms

To evaluate the performance of the combined algorithms, multiple experimental groups were formed by pairing each robot assignment strategy with each elevator algorithm. This resulted in four distinct experimental configurations:

1. LOOK Algorithm + FCFS
2. LOOK Algorithm + NS
3. Nearest Neighbor (DNN) + FCFS
4. Nearest Neighbor (DNN) + NS
5. Nearest Neighbor (INN) + FCFS
6. Nearest Neighbor (INN) + NS

The total construction duration (TCD) is used to estimate the performance, as shown in Equation (1).

$$\pi^* = \underset{\pi}{\operatorname{argmin}}(\text{TCD}) \quad (1)$$

π is the coordination strategy generated by the above configurations. The objective is to find the best

strategy π^* to minimize the total construction duration.

4. RESULTS AND DISCUSSION

The performance of each coordination strategy was assessed using key metrics such as total construction duration. Data collected during simulation runs were analyzed to determine which combinations of algorithms yielded optimal results in coordinating elevators and robots within a multi-story environment. Taking into account the cost of the robots and the information obtained from visits to robot construction companies (Guangdong Bright Dream Robotic Co. Ltd (BDR)), we set the parameters of construction robot as shown in Table 1 and Table 2.

Table 1. The parameters of construction robot

	Plastering	Puttying	Painting
Work speed (Percentage/unit time)	0.48	1.16	0.76
Call material frequency (unit time/1 times)	20	5	20

Table 2. The parameters of delivery robot

	Plastering	Puttying	Painting
Loading time (unit time /1 times)	3	3	3
Supply time (unit time/1 times)	5	5	5

In particular, the above settings are only for the simulation environment to make it run smoothly. These parameters are not considered as known information in all the proposed coordination algorithms.

In order to fully explore the performance of the algorithm, we set the number of robots in various ways to describe different situations, as shown in Table 3. It can be divided into three categories. The first category is groups 1-3, which is to explore the impact of increasing the number of construction robots when the number of delivery robots is very small. The second category is groups 4-6, where the number of construction robots is set inversely proportional to their construction efficiency. On this basis, we explored setting the number of delivery robots inversely proportional to the frequency of calling materials, the same number of delivery robots, and the same number of delivery robots as construction robots. The third category is groups 7-9. Here we keep the number of construction robots and delivery robots the same, and gradually increase the number of robots.

Table 3. Number settings of construction robot and delivery robot

No.		Plastering	Puttying	Painting
1	Construction robot	1	1	1
	Delivery robot	1	1	1
2	Construction robot	2	2	2
	Delivery robot	1	1	1
3	Construction robot	5	5	5
	Delivery robot	1	1	1
4	Construction robot	5	2	3
	Delivery robot	1	2	1
5	Construction robot	5	2	3
	Delivery robot	2	2	2
6	Construction robot	5	2	3
	Delivery robot	5	2	3
7	Construction robot	2	2	2
	Delivery robot	2	2	2
8	Construction robot	3	3	3
	Delivery robot	3	3	3
9	Construction robot	5	5	5
	Delivery robot	5	5	5

Table 4 and Table 5 show the results of the combination of LOOK with different robot target allocation algorithms. Table 6 and Table 7 show the total construction duration by applying DNN with different robot target allocation algorithms. Table 8 and Table 9 show the total construction duration by applying INN with different robot target allocation algorithms. All the serial numbers of results are consistent with Table 3.

Table 4. Total Construction Duration by applying LOOK Algorithm + FCFS

No.	Total Construction Duration (unit time)	
	Robots wait elevators	Elevators wait robots
1	23727	24214
2	24330	24393
3	25147	24952
4	18001	17639
5	14204	14212
6	13728	13631
7	14113	14176
8	10939	10999
9	8084	7982

Table 5. Total Construction Duration by applying LOOK Algorithm + NS

No.	Total Construction Duration (unit time)	
	Robots wait elevators	Elevators wait robots
1	23535	23468
2	23664	23881
3	24511	24383
4	16471	16165
5	13326	13246
6	13021	12827
7	13266	13273
8	9715	9954
9	6732	6629

Table 6. Total Construction Duration by applying DNN + FCFS

No.	Total Construction Duration (unit time)	
	Robots wait elevators	Elevators wait robots
1	24639	24420
2	25298	24816
3	26023	25812
4	17972	17408
5	14398	14041
6	13643	13304
7	14048	13835
8	10578	10390
9	7419	7240

Table 7. Total Construction Duration by applying DNN + NS

No.	Total Construction Duration (unit time)	
	Robots wait elevators	Elevators wait robots
1	24639	24420
2	25285	24975
3	26400	25674
4	17266	17344
5	14259	13961
6	13647	13252
7	14024	13760
8	10692	10359
9	7494	7344

Table 8. Total Construction Duration by applying INN + FCFS

No.	Total Construction Duration (unit time)
	Elevators wait robots
1	24375

2	24784
3	25735
4	17356
5	13986
6	13324
7	13823
8	10183
9	7205

Table 9. Total Construction Duration by applying INN + NS	
No.	Total Construction Duration (unit time)
	Elevators wait robots
1	24375
2	24756
3	25596
4	17222
5	13945
6	13302
7	13718
8	10070
9	7337

From the above results, we can draw the following inferences:

- The combination of LOOK algorithm + NS performed best in almost all experimental groups.
- When there is only one delivery robot for each process, increasing the number of construction robots will not improve construction efficiency but will increase the total construction duration. (Refer to the experimental results of groups 1-3.)
- From the perspective of elevators waiting for robots or robots waiting for elevators, most results are that the former is better than the latter, no matter which combination it is.
- For the nearest neighbor method, INN always performs better than DNN, whether combined with FCFS or NS.

The detailed analysis will be discussed in the following section.

5. DISCUSSION

As mentioned in Section 3.1, there are three main types of activities that rely on elevators to cross floors in the proposed scenario. Among them, the most frequent is the delivery robot delivering materials and returning to the loading point after delivering materials. In the simulation environment, the loading point is fixed on the ground floor. Therefore, the movement of the delivery robot is a scanning movement from the ground floor to the high floor, which explains why the LOOK algorithm performs best, because the LOOK algorithm is an efficient scanning algorithm.

When the number of delivery robots is limited, blindly increasing the number of construction robots is not a viable choice, because this will increase the proportion of construction robots moving across floors. Due to the limited capacity of elevators, more construction robots will compete with delivery robots for elevators, thereby reducing the delivery efficiency of delivery robots, prolonging the waiting time for construction robots in need of materials, and reducing the overall construction efficiency.

The robot waiting for the elevator means that the robot will not work for a long time, and the robot's working time is directly related to the overall construction progress. The mode of waiting for the robot for the elevator can shorten the time the robot waits for the elevator, which means that the robot will be put into work faster.

In DNN, the distance between the request and the current floor of the elevator is the only factor that determines the movement of the elevator. But in reality, it is possible that when the elevator reaches the nearest requested floor, the robot is still on the way. In this case, the elevator needs to wait for the robot to arrive and pick it up before it can move. In INN, the actual time when the robot reaches the elevator and the distance from the elevator to the requested floor are taken into account, and the "closest" is evaluated more accurately in time and space. This reduces the time the elevator waits for the robot.

We also make some speculations as to why the nearest neighbor method does not perform as expected in the proposed scenario. Previous studies (Sheridan et al. 2013) have shown that the nearest neighbor is a promising method for solving the traveling salesman problem. However, in order to increase the generalization ability of the

method, they relax the constraints and assumptions of the problem, such as the random generation of vehicle demand (Ghiani et al. 2003). In this study, the scenario is a specific multi-story construction site, and the construction robots follow the top-down flow construction, which greatly reduces the randomness of the demand for elevators. In this case, some special algorithms, such as the scanning algorithm LOOK, may produce better results because they match the movement mode of the majority of delivery robots. This observation suggests that while existing algorithms with strong generalization capabilities are valuable in broader contexts, they may have limited effectiveness in specialized scenarios such as robot-elevator coordination in construction. The findings indicate a significant research potential for developing multi-agent scheduling algorithms tailored specifically to unique operational environments like multi-story construction sites.

6. CONCLUSIONS

In conclusion, this study experimentally investigated multi-robot coordination within a multi-story construction site. We created a simulated environment consisting of ten floors, equipped with two elevators and teams of robots tasked with various cross-floor requirements. We approached the situation by implementing established elevator algorithms and robot target allocation strategies that were adapted to fit this particular context. Through group experiments, we explored the performance of these algorithms in multi-story construction settings and examined the characteristics of robot-elevator coordination. The results indicated that the combination of LOOK algorithm and NS, which operates by systematically scanning floors, is particularly well-suited for this environment. Its design aligns with the predominant motion patterns of delivery robots, facilitating more efficient elevator scheduling. By comparing our findings with conclusions from previous studies, we observed that algorithms that excel in general scenarios often perform significantly worse than those specifically tailored to the unique characteristics of multi-story construction sites. This disparity underscores the potential for further research in developing multi-agent coordination algorithms that cater to specialized operating environments. Currently, the total construction duration is the only metric considered in this study. But minimizing task completion time may increase total travel distance, leading to higher energy usage. It would provide a more comprehensive evaluation by adding additional metrics, such as resource utilization and energy consumption. Looking ahead, we plan to conduct more in-depth explorations of elevator-robot coordination in multi-story construction sites. Our future research will focus on improving these algorithms and evaluating their applicability in more complex construction scenarios, ultimately contributing to improving operational efficiency and automation in the construction industry.

ACKNOWLEDGMENTS

This work was supported by Collaborative Research Fund [grant number C6044-23GF] from University Grants Committee of The Government of the Hong Kong Special Administrative Region; and 30 for 30 Scheme [grant number 3030007] funded by The Hong Kong University of Science and Technology.

REFERENCES

- Aghimien, Douglas Omoregie, Clinton Ohis Aigbavboa, Ayodeji Emmanuel Oke, and Wellington Didibhuku Thwala. (2020). 'Mapping out research focus for robotics and automation research in construction-related studies: A bibliometric approach', *Journal of Engineering, Design and Technology*, 18: 1063-79.
- Berbeglia, Gerardo, Jean-François Cordeau, Irina Gribkovskaia, and Gilbert Laporte. (2007). 'Static pickup and delivery problems: a classification scheme and survey', *Top*, 15: 1-31.
- Berbeglia, Gerardo, Jean-François Cordeau, and Gilbert Laporte. (2010). 'Dynamic pickup and delivery problems', *European journal of operational research*, 202: 8-15.
- Berbeglia, Gerardo, Jean-François Cordeau, and Gilbert Laporte. (2012). 'A hybrid tabu search and constraint programming algorithm for the dynamic dial-a-ride problem', *INFORMS Journal on Computing*, 24: 343-55.
- Cordeau, Jean-François. (2006). 'A branch-and-cut algorithm for the dial-a-ride problem', *Operations research*, 54: 573-86.
- Cordeau, Jean-François, and Gilbert Laporte. (2003). 'A tabu search heuristic for the static multi-vehicle dial-a-ride problem', *Transportation Research Part B: Methodological*, 37: 579-94.
- Garaix, Thierry, Christian Artigues, Dominique Feillet, and Didier Josselin. (2011). 'Optimization of occupancy rate in dial-a-ride problems via linear fractional column generation', *Computers & Operations Research*, 38: 1435-42.
- Ghiani, Gianpaolo, Francesca Guerriero, Gilbert Laporte, and Roberto Musmanno. (2003). 'Real-time vehicle routing: Solution concepts, algorithms and parallel computing strategies', *European journal of operational research*, 151: 1-11.
- Ho, Sin C, Wai Yuen Szeto, Yong-Hong Kuo, Janny MY Leung, Matthew Petering, and Terence WH Tou. (2018). 'A survey of dial-a-ride problems: Literature review and recent developments', *Transportation Research*

Part B: Methodological, 111: 395-421.

- Jorgensen, Rene Munk, Jesper Larsen, and Kristin Berg Bergvinsdottir. (2007). 'Solving the dial-a-ride problem using genetic algorithms', *Journal of the operational research society*, 58: 1321-31.
- Kousi, Niki, Spyridon Koukas, George Michalos, and Sotiris Makris. (2019). 'Scheduling of smart intra-factory material supply operations using mobile robots', *International Journal of Production Research*, 57: 801-14.
- Kousi, Niki, Spyridon Koukas, George Michalos, Sotiris Makris, and George Chrysosolouris. (2016). 'Service oriented architecture for dynamic scheduling of mobile robots for material supply', *Procedia CIRP*, 55: 18-22.
- Le, Anh Vu, Rizuwana Parween, Phone Thiha Kyaw, Rajesh Elara Mohan, Tran Hoang Quang Minh, and Charan Satya Chandra Sairam Borusu. (2020). 'Reinforcement learning-based energy-aware area coverage for reconfigurable hRombo tiling robot', *IEEE Access*, 8: 209750-61.
- Mauri, Geraldo Regis, Luiz Antonio, and Nogueira Lorena. (2009). 'Customers' satisfaction in a dial-a-ride problem', *IEEE Intelligent Transportation Systems Magazine*, 1: 6-14.
- Motroni, Andrea, Alice Buffi, and Paolo Nepa. (2021). 'A survey on indoor vehicle localization through RFID technology', *IEEE Access*, 9: 17921-42.
- Nielsen, Izabela, Ngoc Anh Dung Do, Peter Nielsen, and Yohanes Khosiawan. (2016). "Material Supply Scheduling for a Mobile Robot with Supply Quantity Consideration—A GA-based Approach." In, 41-52. Cham: Springer International Publishing.
- Qu, Yuan, and Jonathan F Bard. (2015). 'A branch-and-price-and-cut algorithm for heterogeneous pickup and delivery problems with configurable vehicle capacity', *Transportation Science*, 49: 254-70.
- Seriani, Stefano, Alessio Cortellessa, Sandro Belfio, Marco Sortino, Giovanni Totis, and Paolo Gallina. (2015). 'Automatic path-planning algorithm for realistic decorative robotic painting', *Automation in Construction*, 56: 67-75.
- Sheridan, Patricia Kristine, Erich Gluck, Qi Guan, Thomas Pickles, Barış Balciog, and Beno Benhabib. (2013). 'The dynamic nearest neighbor policy for the multi-vehicle pick-up and delivery problem', *Transportation Research Part A: Policy and Practice*, 49: 178-94.
- Toth, Paolo, and Daniele Vigo. 2014. *Vehicle routing: problems, methods, and applications* (2nd ed.). Society for Industrial and Applied Mathematics Philadelphia.
- Wang, Yue, Liangxi Xie, Jin Chen, Mengmeng Chen, Teng Hu, Hongyu Liao, Shibin Sun, and Jian Chen. (2024). 'Mortar spraying and plastering integrated robot for wall construction', *Automation in Construction*, 165: 105533.

UNDERSTANDING WORKERS' ACCEPTANCE OF ROBOTICS IN CONSTRUCTION

Zhida Zhang¹, Fanfan Meng², Mun On Wong³ and Mi Pan^{4*}

1) Research Assistant, Department of Civil and Environmental Engineering, University of Macau, Macau, SAR. Email: zhidazzd@um.edu.mo

2) Ph.D. Student, Department of Civil and Environmental Engineering, University of Macau, Macau, SAR. Email: yc47936@um.edu.mo

3) Ph.D., Research Assistant Professor, Department of Civil and Environmental Engineering, University of Macau, Macau, SAR. Email: mowong@um.edu.mo

4) Ph.D., Ass. Prof., Department of Civil and Environmental Engineering, University of Macau, Macau, SAR. Email: mipan@um.edu.mo (Corresponding Author)

Abstract: Robotics is expected to enhance productivity and safety in the construction industry, but the real-world application remains limited. Introducing robotics in construction may require humans and robots to work together for the same tasks or in close proximity. While significant attention has been paid to organizational-level robot adoption, little exploration has been done from the perspective of construction workers. This paper aims to provide a comprehensive understanding of the many factors that may influence workers' attitudinal acceptance of robotics in construction. A case study including observations and interviews with 40 construction workers of a project in the Guangdong-Hong Kong-Macao Greater Bay Area was conducted, coupled with semi-structured interviews with 13 site managers. Various factors influencing workers' acceptance were identified, including individual differences of workers, technological performance, and output quality of robots. Additionally, external factors including organizational support and social influences can affect workers' attitudes toward robots. The findings reveal that most workers will passively accept construction robots when their organization mandates their utilization, although changes in income remain a major concern. Strategies are recommended for future research and practice of robots for various stakeholders, such as guaranteeing workers' income, strategizing practice-based technology, improving multi-level robot interface management, and enhancing government support. This study should encourage different stakeholders to design and adopt construction robots guided by human-centered design principles.

Keywords: construction robotics, worker acceptance, human-centered robotics, technology acceptance model, robot adoption.

1. INTRODUCTION

Construction represents a significant component of the global economy. In 2023, the gross annual output of global construction projects accounts for about 7% of the world's gross output (McKinsey, 2024). Despite its significance, the construction industry faces severe challenges such as labor shortages and tight workforce markets. Construction projects heavily depend on the manual tasks provided by skilled construction workers (Durdyev et al., 2017). However, retirement, shorter job cycles, and talent competition have led to a notable decline in the skill and experience levels of construction labor (World Economic Forum, 2023). For example, in Hong Kong, the Construction Industry Council (2023) states that there is a significant manpower shortage of skilled construction workers, with a supply-demand mismatch of 5-15% in 2023 expected to increase to 15-20% by 2027. Moreover, the gradual decline in the number of young individuals entering the industry has resulted in the trend of an aging workforce, exacerbating the talent shortage (Pan et al., 2020b). Therefore, there is an urgent need to consider and introduce new methods to alleviate labor shortage and boost productivity across the construction sector.

Robotics has been widely recognized as an important technological innovation in construction (Pan & Pan, 2020). However, robot adoption is still limited. One major challenge facing the use of robots, not only in construction, is related to humans and robots working together for same tasks or in close proximity. There is also increasing concern that robots may displace human workers, leading to negative attitudes of humans toward working with robots in many industries (Nomura et al., 2006; Takayama et al., 2008). Given the great potential of robotics in future construction, a comprehensive understanding of workers' acceptance of robotics in construction is needed (You & Robert, 2017). Scholars have also explored the determinants of construction robot adoption from an organizational perspective (Pan et al., 2020a; Vora et al., 2024). However, little is known from the workers' perspective in terms of which factors could influence workers' attitudinal acceptance of robotics, especially for those working closely with robots.

To address the gap, this paper aims to provide a comprehensive understanding of potential factors that may influence workers' attitudinal acceptance of robotics in construction. Guided by the technology acceptance model (TAM), the research was carried out in the combination of a critical literature review, a case study of a residential project involving interviews with construction workers, and supplementary semi-structured interviews

with site managers, which provide a comprehensive understanding of influencing factors on workers' attitudinal acceptance of construction robots. The findings should provide practical implications and policy recommendations to facilitate the development and adoption of robotics in the construction industry.

2. RELATED WORK

2.1 Innovation Adoption Theory

The technology acceptance model (TAM) is a generic innovation theory for predicting individual adoption, which is widely used for studying behavior intention toward technological innovations and examining factors affecting user acceptance (Zhang et al., 2008; Charness & Boot, 2016). Specifically, TAM explores why users accept or reject an innovation and which factors influence their willingness to accept it (Davis 1986). As shown in Figure 1, TAM focuses on exploring the prospective overall attitude of corresponding users toward using a given system. The effect of external variables on attitudinal acceptance is mediated by two fundamental determinants, perceived usefulness (refers to the degree to which a person believes that using a particular system would enhance his or her job performance) and perceived ease of use (refers to the degree to which a person believes that using a particular system would be free of effort). Davis's work (1986) laid the foundation for many innovation studies on individual acceptance, and has been extended to considering additional external variables, such as social influence and cognitive instrumental processes (Venkatesh & Davis, 2000), and the application to understand innovation in different fields, such as the acceptance of information technologies (Venkatesh & Bala, 2008). TAM has also been effectively used in construction research, in combination with other innovation theories, to understand the acceptance of different technologies (Katebi et al., 2022; Park et al., 2023). Previous studies have demonstrated the feasibility of applying and extending TAM by adding further variables to understand user acceptance of innovative technologies on different applications. This study employed TAM as the theoretical basis considering its solid theoretical foundation and applicability in exploring potential external variables influencing workers' attitudinal acceptance of robotics in construction.

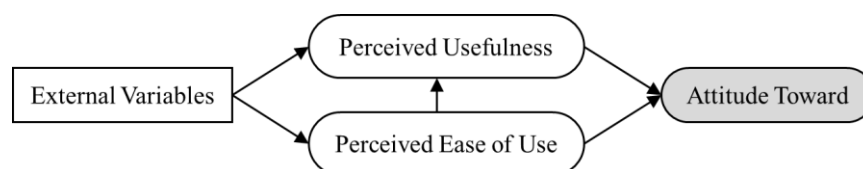


Figure 1. Technology acceptance model (Davis 1986)

2.2 Construction Robot Adoption

Research has been carried out to examine determinants for the adoption of construction robots. Pan & Pan (2020) investigated the determinants of construction robot adoption from the perspectives of building contractors based on the technology organization environment (TOE) framework and identified top management support as the predominant factor in robot adoption. Kim et al. (2022) collected the perceptions from trade workers and managers and illustrated that job complexity and safety risks had the most significant influence on construction personnel's desired capabilities and perceived usefulness of robots. Park et al. (2023) investigated construction personnel's behavioral intention to accept human-robot collaboration, including craftsmen, supervisors, managers, engineers, and directors in the US construction industry, and indicated that perceived usefulness and perceived ease of use had a positive and significant impact on the intention. Vora et al. (2024) found that saving time by reducing rework and facilitating quality control are the top driving factors of various stakeholders to adopt construction robots. The interest in bringing robots onto construction sites will force robots and human workers into closer proximity. However, little has been done to explore factors influencing construction worker acceptance of robotics from the individual and practical perspective. Although robot adoption decisions are typically made at management or organizational level, the actual adoption performance would be affected by workers' attitudinal acceptance.

3. RESEARCH METHOD

The study was conducted by combining a critical literature review, a case study including observations and semi-structured interviews with 40 construction workers of a residential building project in the Guangdong-Hong Kong-Macao Greater Bay Area, and supplementary semi-structured interviews with 13 site managers. This mixed-method research design enables a comprehensive and in-depth understanding of attitudes and perceptions of workers toward the adoption of construction robots.

3.1 Case Study with a Residential Building Project

The case study method was employed to enable an empirical investigation (Taylor et al., 2006), using multiple sources of evidence: (1) document analysis based on factual data; (2) observations including participant

observations by one researcher engaged as a project engineer from project inception to closeout, with a particular focus on challenges and pain points encountered by workers; (3) semi-structured interviews with 40 workers from eight major subcontractors.

The case study was based on a residential project with a floor area of approximately 200,000 square meters. The project duration is approximately 2 years, including 1,017 residential units, 136 commercial units, 3 level basement, and a kindergarten. Many construction robots have been deployed on a trial basis in this project, including a construction measuring robot, a building cleaning robot, an indoor concrete wall grinding robot, a concrete floor smoothing/screeding robot, and an intelligent follow-up distribution machine.

Semi-structured interviews were conducted with 40 workers selected through purposive sampling, a method used to select respondents most likely to provide appropriate and useful information (Campbell et al., 2020). Selection criteria include employment on the site with first-hand experience, holding certain degree of knowledge and awareness of construction robots, and trades related to applicable fields of robots, including electrician, steeplejack, handyman, bricklayer, plaster, painter, concreter, and tiler (Table 1). The majority of interviewees (52.5%, n=21) were over 40 years old, reflecting the aging trend in the construction industry. Each interview lasted approximately 20-30 minutes and was audio-recorded with permission. This includes an introductory explanation of existing construction robots with the aid of photos and videos, which allow workers to better understand the functions and current usages of construction robots, followed by in-depth discussions on robot acceptance. The transcripts and notes taken were translated, and analyzed through seven steps formulated by Easterby et al. (2002) and NVivo version 15 software.

Table 1. Details of semi-structured interviewees with construction workers (n=40)

Trade	n	Percentage	Work experience	n	Percentage	Age	n	Percentage
Plaster	3	7.5%	0-9 years	16	40.0%	20-29 years	7	17.5%
Electrician	3	15%	10-19 years	11	27.5%	30-39 years	12	30.0%
Steeplejack	6	12.5%	20-29 years	7	17.5%	40-49 years	7	17.5%
Handyman	5	7.5%	> 30 years	6	15.0%	50-59 years	14	35.0%
Bricklayer	3	7.5%						
Painter	10	25%						
Concretor	4	10%						
Tiler	6	15%						

3.2 Semi-structured Interviews with Site Managers

Supplementary semi-structured interviews were conducted with 13 site managers, which provided insights from the management on the workers' acceptance of construction robots, as well as suggestions for future human-centered construction robot development. The purposive sampling method was used for interviewee selection. Participants held key positions from key stakeholder groups in the construction industry, including clients, contractors, subcontractors, and supervisors (Table 2). Notes taken during site manager interviews were analyzed similarly to those taken during case study interviews.

Table 2. Details of semi-structured interviewees with site managers (n=13)

Company	Position	n	Company	Position	n
Client	Engineering manager	1	Contractor	Project manager	2
	Electrical engineer	1		Quality manager	1
	Civil engineer	1		Technical manager	1
Supervisor	Supervising engineer	1	Subcontractor	Decoration manager	2
	Chief supervisory engineer	1		Landscape project manager	2

4. RESULTS AND ANALYSES

The data obtained from the case study and semi-structured interviews was carefully analyzed to tease out pertinent themes and categories. As a result, five dimensions and their associated factors that could influence workers' acceptance of construction robots were identified (Table 3). The study also revealed considerable variation in workers' attitudes toward robots. Specifically, 43% (n=17) of workers expressed a positive attitude toward robots, while 30% (n=12) and 28% (n=11) of workers exhibited neutral and negative attitudes, respectively.

Table 3. Influencing factors of workers' acceptance on construction robots

Dimension	Influencing factor	Description
Individual differences	Years of work experience	Duration of workers' work experience
	Trade	Type of work for which the worker is responsible
	Age	Age of workers
Technological performance	Adaptability	The robot's capacity to adapt to complex working environments and its user-friendly interface
	Technology maturity	Technology readiness level of robots
Output quality	Auxiliary capability	Performance of construction robots assisting workers in completing a task
	Construction quality	Robot's performance in construction results
Organizational support	Work interface management	Division of tasks and responsibilities between construction robots and workers
	Income assurance	Remuneration of workers for their labor
Social influence	Social processes	Developments and changes in society and industry
	Job competition	Employment pressure on workers from construction robots

4.1 Individual Differences

Individual differences refer to personality and/or demographics that can influence an individual's perception of the usefulness and ease of use of a particular product or service (Venkatesh & Bala, 2008). The findings indicate that years of work experience and trades significantly impact workers' attitudinal acceptance of robotics, while the age of workers has a relatively minor effect.

Years of work experience: The length of work experience has a clear impact on workers' acceptance of robotics. Specifically, 62.5% (n=25) of workers with 0-9 years of experience have a positive attitude toward robots, while only 16.7% of workers with more than 30 years of experience hold a similar view (Figure 2). Some workers indicated that they were reluctant to embrace changes in their roles and were uncertain about their ability to adapt to new challenges posed by the introduction of robots. As workers accumulate experience, they become more familiar with their tasks and responsibilities, leading to greater resistance to change and new challenges.

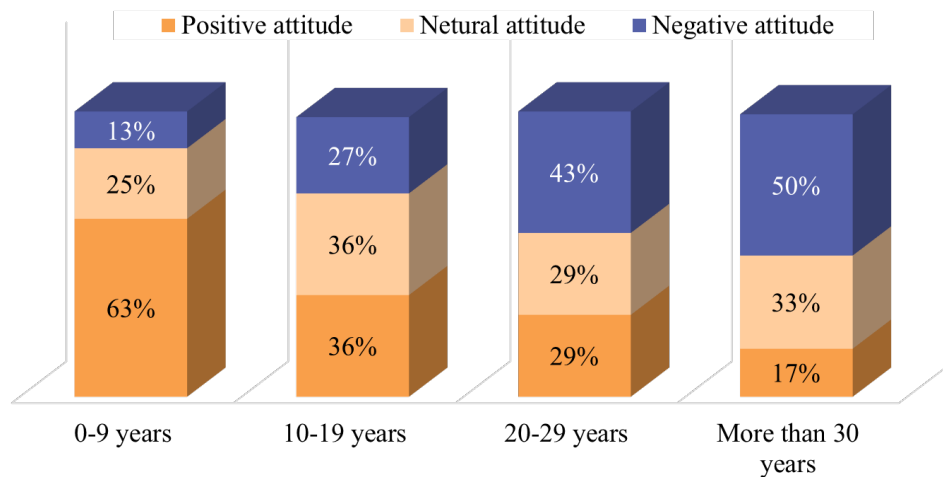


Figure 2. Relationship between workers' attitudes and years of work experience

Trade: There is a notable discrepancy in the acceptance of robots among workers with different trades (Figure 3). Researcher's observations and interviews with site managers identified the following reasons for different workers.

- Electricians (100% positive attitude) have high acceptance of robots because the skills required for electrical services are highly specialized and require a high level of knowledge and expertise, making them more open to new technologies.
- Handymen and bricklayers (80% and 67% positive attitude, respectively) have a high acceptance of robots due to the relatively simple nature of their tasks, which often involve cleaning and lifting. Robots can help them save physical strength to a large extent and are technically easier to implement.
- Steeplejacks (67% positive attitude) believe that robotics can assist in dangerous tasks, resulting in a more

favorable outlook towards robots.

- Plasterers, painters, and concretors (33%, 20%, and 25% positive attitude, respectively) are less receptive to robots as their jobs usually require more manual experience to ensure good output quality.
- Tilers (17% positive attitude) have the lowest level of acceptance. Tiling requires flexible adjustments based on site conditions, and workers believe that robots lack the necessary intelligence for the task.

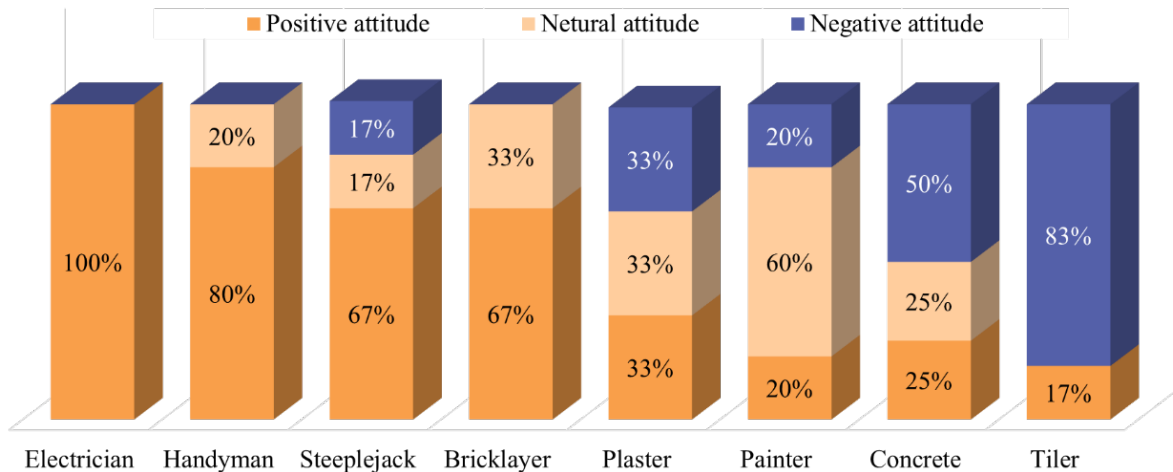


Figure 3. Relationship between workers' attitudes and trades

Age of workers: In terms of age, both the 20-29 and 50-59 age groups expressed positive attitudes toward robots, with 71.4% and 50% acceptance, respectively (Figure 4). The former (younger) group is more willing to accept new things, while the latter (older) is usually curious or indifferent to new things as they approach retirement. For others, results from the case study and site manager interviews indicated that workers aged 30-49 generally face greater financial pressure, view robots as competitors, and are concerned that robots may replace them, leading to reluctance to accept robots.

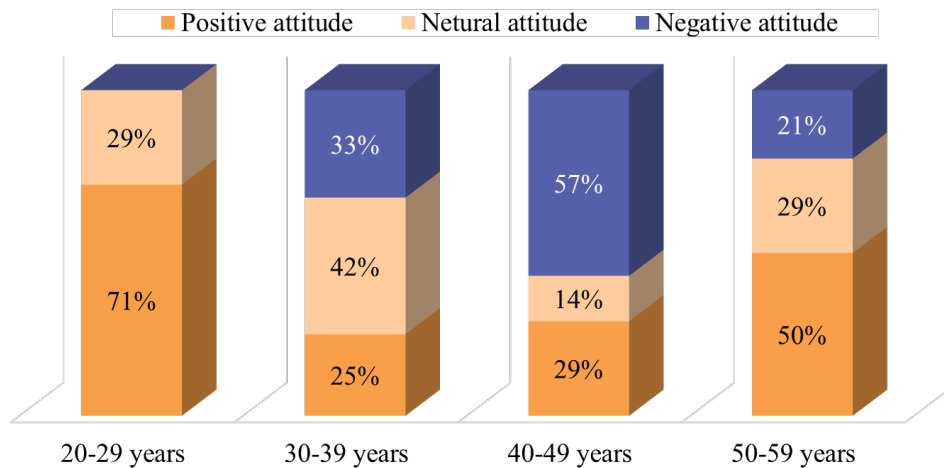


Figure 4. Relationship between workers' attitudes and their age

4.2 Technological Performance

Technological performance considers the performance of construction robotics at the technological level, which was identified as an important factor influencing workers' acceptance, particularly in terms of adaptability and maturity. Key viewpoints obtained during the case study and interviews are summarized as follows.

Adaptability: Nearly half of the workers (48%, n=19) believed that the adaptability of robots significantly affects their acceptance. Generally, robots are not as flexible as workers in dealing with problems encountered in the construction process and are largely confined to work according to established procedures and standards, limiting their adaptability to dynamic construction environments. Particularly, the complex environment of construction sites was perceived by 30% (n=12) of the workers as a significant challenge for the application of construction robots. Primary challenges include: 1) presence of obstacles or restricted walking paths within the

designated working environment would prevent the robot from reaching the work area and carrying out its intended work; 2) substandard construction quality of preceding process would adversely affect the normal operation of the robot, and it is challenging for the robot to adaptively address issues such as ‘uneven ground’ and ‘misaligned walls’; and 3) a wide range of disparate tasks, processes, and cross-works make it difficult to provide a conducive environment for robots to operate independently. Consequently, the performance of robots operating in complex construction environments is often suboptimal. Furthermore, the complex operation and commissioning of robots present challenges for workers, providing additional workload for workers and affecting workers’ efficiency and willingness to collaborate with robots.

Technology maturity: During the interviews, most of the workers who were neutral about robots stated that they would accept robotic technology if it reached a high maturity level. Conversely, workers are reluctant to collaborate with robots if the technology maturity is low. Some workers even expressed their willingness to purchase highly useful robots themselves if existed, indicating that technology maturity is a key factor affecting workers’ attitudes toward robots. Construction site managers also believed that a high readiness level of robots would encourage companies to promote the application vigorously.

4.3 Output Quality

Output quality in the context of robot acceptance could be interpreted as the extent to which an individual perceives the robot to be effective in fulfilling the requirements of their job tasks (Venkatesh & Davis, 2000). The identified factors influencing workers’ acceptance and the primary perspectives collected from the case study and interviews are summarized below.

Auxiliary capability was identified as the uppermost factor influencing workers’ acceptance of robotics in construction, raised by 35 out of the 40 workers (88%) and 10 of the 13 site managers (77%). The strenuous physical nature of construction works means that if robots can assist workers in their tasks with greater ease, this would positively influence worker perception and acceptance. In addition, robots that enable time saving and productivity enhancement could help workers generate more income within a certain period, fostering a positive perception from workers on robots. Moreover, robots can help workers complete hazardous tasks (such as working at heights) and enter harsh environment, reducing workers’ physical harm and improving their acceptance of robots.

Construction quality: More than half of workers (53%, n=21) identified the quality of the robot’s work as a significant factor. This is because when robots do not complete tasks to the required standard, responsibility for rectifying construction defects falls on the workers. The case study also revealed that the majority of workers preferred to complete work in one go and were reluctant to assist robots with rework tasks. However, as noted by some workers, if the robot’s work quality surpassed that of humans, they would be more open to accepting the robot and learning to operate it.

Furthermore, many workers (62.5%, n=25) and site managers (over half) perceived limited coverage as a technological disadvantage of construction robots. However, it was found to have little impact on the workers’ acceptance of robots, since complete coverage by robots would challenge workers’ employment opportunities. The limited coverage is reflected in two main aspects: 1) the robot’s size, movement ability, and other factors limit its capability to complete tasks in corners and narrow areas, which restricts its work coverage; 2) immature technology makes it difficult for robots to complete tasks for certain construction processes and special positions that heavily reliant on human experience.

4.4 Organizational Support

Organizational support refers to the influence from the organization on workers’ acceptance of robots, mainly in the areas of work interface management and income assurance, as elaborated below.

Work interface management: As highlighted by many workers (40%), the division of work between robots and workers is a significant issue that requires attention at the organizational level. The main purpose of work interface management is to clarify tasks and define the responsibilities between robots and workers, thereby preventing mutual interference during construction. Proper work interface management is crucial during the early stages of robot technology to avoid disrupting workers’ normal activities, which could adversely affect the acceptance of robots by workers.

Income assurance: Almost half of the employees (48%, n=19) believe that the company must guarantee their income regardless of robot performance. This can be interpreted in two ways: (1) when the introduction of robots affects the efficiency of workers, they are happy to accept them if guaranteed a reasonable income (e.g. time-based pay); (2) even though robots somehow save workers’ time and effort, workers are unwilling to reduce their income as a result. It is evident that income is the primary concern for workers and significantly affects their acceptance of robots.

4.5 Social Influence

Social influence refers to various social processes and mechanisms that influence workers’ perceptions of construction robotics (Venkatesh & Bala, 2008), which was also found to influence the acceptance of robots by

workers, grouped into social processes and job competition, as mentioned below.

Social processes: The development of society and advances in technology across various industries have shown workers the benefits new technology has brought, encouraging a positive attitude toward construction robotics. Furthermore, the increasingly aging population suggests that robots are likely to be widely used in the future to supplement the workforce. It was perceived that accepting and learning robotics could enhance workers' competitiveness. However, the development trend of the construction industry may influence workers' perceptions of construction robotics. For example, the current downturn of the construction industry in China negated interviewed workers' acceptance of this new technology, as they believed the advancement of construction robotics was contingent upon industry growth.

Job competition: During the case study interviews, a majority of workers (58%, n=23) perceived robots as competitors and potential threats to their employment. There is a growing concern that robots are taking job opportunities away from human workers. The introduction of robotics will reduce labor demand, potentially leading to job losses or income decline for construction workers. This trend may engender negative attitudes toward accepting construction robotic in their work sites.

5. DISCUSSION

5.1 Factors Influencing Workers' Acceptance of Robotics

The study examined the attitudinal acceptance of construction robots from the perspective of workers, who are the key group that works most closely with robots in practice. The study identified five major dimensions that influence workers' acceptance of construction robots, which are related to workers, robots, and external influences. Specifically, a new framework of workers' acceptance of construction robots (WACR) was derived, as shown in Figure 5. Key findings are elaborated below.

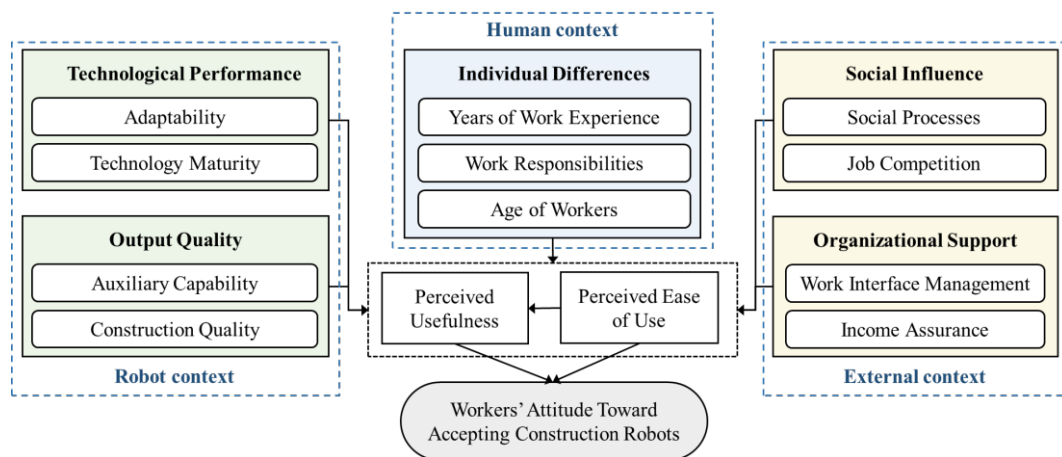


Figure 5. Framework of workers' acceptance of construction robots (WACR)

Firstly, the findings of the case study and semi-structured interviews indicated that differences in workers' age, years of work experience, and trades affect their acceptance of construction robots. The findings echo with those reported by Esterwood, et al. (2022), who highlighted that factors such as personality traits, age, gender diversity, and tasks impact the acceptance of general robots. Besides, younger workers with less experience were more likely to accept the use of construction robots. Additionally, workers engaged in different types of work exhibited varying levels of acceptance.

Secondly, previous studies revealed that both technological performance and output quality are key determinants of the acceptance of new technological innovations (Marangoni & Andrina., 2015; Silberer et al., 2023), which are consistent with the results of this study. The ability of construction robots to perform complex tasks in site environments is crucial for workers' acceptance of robots, particularly for workers with a neutral attitude toward robots. Besides, if robots can assist workers to achieve better construction results while saving time and labor, it encourages workers to form a more positive attitude.

Thirdly, external influences from the organizational and societal levels primarily impact workers' employment opportunities and income, which indirectly influence their acceptance of construction robots, echoing earlier research (e.g. Meissner et al., 2020).

In summary, the extent to which robots can assist workers in achieving high returns with minimal effort was considered to significantly influence workers' acceptance of construction robots. However, this finding differs from Park et al. (2023)'s study, which identified that the perceived effect of life—considering that robots could bring better working conditions, increase productivity, and reduce occupational risks—has no significant impact

on construction personnel acceptance of robotic assistants in construction works. This might be due to several possible reasons: 1) the different focused regions, considering China and US; 2) the different surveyed personnel, as this study focused on frontline workers while Park et al. (2023) considered five groups and targeted not just worker level; and 3) the different survey distribution mode considering whether the survey is allocated as a top-down task. This distinction further reveals that the impact of construction robotics on the quality of life could be a significant concern for frontline workers, whereas other construction personnel who engage in non-heavy labor and do not directly participate in construction activities tend to exhibit different attitudes.

5.2 Strategies for Future Development of Construction Robotics

Based on the research findings, the paper proposed several strategies to facilitate the workers' acceptance of construction robots at three levels: organizational, technological, and governmental.

Firstly, the study reveals that most workers could passively accept the introduction of robots as long as their reasonable income is guaranteed. To facilitate workers' acceptance of robots at the development stage of robotics, more attention should be paid to managing work interfaces for optimal human-robot collaboration. Besides, construction organizations should accept potential risks associated with the introduction of robots, such as schedule delays and quality defects. Furthermore, it is essential to ensure that workers' incomes will not be diminished by robots, which could be achieved by improving contractual agreements or providing subsidies to workers.

Secondly, site manager interviews indicate that companies typically compel workers to use robots without considering worker ideas and feedback. Together with the findings of the case study, several strategies are proposed for the design and advancement of construction robotics: firstly, the design needs to be fully integrated with practical experience and the needs of workers, with ongoing iterative updates; secondly, there is no necessity to pursue the complete automation of robots unless they reach high maturity and intelligence, and it is important to combine the respective advantages of humans and robots to develop intelligent auxiliary equipment to assist workers in improving efficiency; thirdly, priority should be given to the promotion of robots with high technological readiness and real-world performance to enhance acceptance.

Thirdly, enhanced government support could potentially increase workers' acceptance of construction robots. The government should vigorously promote the development of construction robotics. This may be achieved by enhancing government-industry-academia collaboration, transferring useful research findings to industry practices (Faisal et al., 2017), and implementing other measures. As previously mentioned in this paper, the technological performance of robots greatly affects workers' acceptance of them. Besides, the government should provide economic and policy support to construction companies that are implementing construction robots. It would be beneficial for the government to supervise these companies to ensure the rights and income of workers. Additionally, it is advised that the government should provide workers with advanced technical training to enhance their capacity to adapt to the transformation of automation and robotics in the construction industry.

6. CONCLUSION

This paper investigates the workers' attitudinal acceptance of construction robots. The research combines a critical literature review, a participatory case study with a residential project, and semi-structured interviews with site managers. The study reveals five major dimensions influencing the acceptance of robots by workers and proposes a new acceptance framework of workers' acceptance of construction robots. In addition to individual differences, technological performance and output quality of robots are significant concerns for workers. Furthermore, external influences such as organizational support and social influences have a significant impact on workers' acceptance of construction robots. A comprehensive analysis of these factors concludes that workers' acceptance of robots fundamentally depends on whether the use of robots can bring benefits to workers, such as saving time, reducing labor, and increasing income. Strategic recommendations are also derived for the future development of construction robots.

The study enriches the theory of robot acceptance in the construction industry from the perspective of construction workers by proposing a new framework for workers' acceptance of construction robots. This research contributes to providing a novel perspective on understanding construction robot adoption from end-users, which should assist robotic technology suppliers in promoting human-centered design, building contractors in introducing robotic systems, and related institutions in publishing favorable guidance for the development and application of construction robots. However, the study has the limitation that workers in only one project in the Greater Bay Area were involved and the findings may not be generalizable to other regions and contexts. Therefore, future work will be twofold. Firstly, the study will engage practitioners from different regions and projects with various demographic profiles to increase sample scale and diversity to obtain more representative and generalizable findings. Secondly, a quantitative follow-up study will be conducted to validate the proposed framework and explore the interrelationships between the influencing factors.

ACKNOWLEDGEMENTS

We acknowledge funding support from University of Macau and University of Macau Development Foundation (File no. SRG2023-00006-FST, MYRG-GRG2024-00209-FST-UMDF). We would also like to express our sincere gratitude to the participants in this study for sharing their experiences and insights.

REFERENCES

- Construction Industry Council, Construction Manpower Forecast. (2023) [https://www.cic.hk/files/page/56/ConstructionManpowerForecastResults\(Eng\)_20230207.pdf](https://www.cic.hk/files/page/56/ConstructionManpowerForecastResults(Eng)_20230207.pdf) (Accessed October 2024).
- Davis, F. D. (1986). A technology acceptance model for empirically testing new end-user information systems: theory and results. Doctoral dissertation. *MIT Sloan School of Management*, Cambridge, MA. 1986.
- Durdyev, S., Omarov, M., Ismail, S. (2017). Causes of delay in residential construction projects in Cambodia. *Cogent Engineering*, 4(1), 1-12.
- Easterby-Smith M., Thorpe R., & Lowe A. (2002). Management Research: An Introduction. in SAGE series in Management Research. *SAGE Publications*.
- Esterwood, C., Essenmacher, K., & Yang, H. et al. (2022) A Personable Robot: Meta-Analysis of Robot Personality and Human Acceptance. *IEEE Robotics and Automation Letters*, 7(3), 6918-6925.
- Faisal, R., Chong, A. L., & Yee, A. S. V. (2017). Systematic review of sustainable knowledge transfer process in government-industry-academia consortium. *Asian Journal of Innovation and Policy*, 6(3), 295-312.
- Katebi, A., Homami, P., & Najmeddin, M. (2022). Acceptance model of precast concrete components in building construction based on Technology Acceptance Model (TAM) and Technology, Organization, and Environment (TOE) framework. *Journal of Building Engineering*, 45, 103518.
- Kim, Y., Kim, H., Murphy, R., Lee, S., & Ahn, C. R. (2022). Delegation or collaboration: Understanding different construction stakeholders' perceptions of robotization. *Journal of Management in Engineering*, 38(1), 04021084.
- McKinsey. (2024). Delivering on construction productivity is no longer optional, Retrieved from McKinsey website: <https://www.mckinsey.com/capabilities/operations/our-insights/delivering-on-construction-productivity-is-no-longer-optional>
- Meissner, A., Trübswette A., & Antonia S., et al. (2020) Friend or foe? understanding assembly workers' acceptance of human-robot collaboration. *ACM Transactions on Human-Robot Interaction (THRI)*, 10(1), 1-30.
- Nomura, T., Suzuki, T., Kanda, T., & Kato, K. (2006). Measurement of negative attitudes toward robots. *Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems*, 7(3), 437-454.
- Pan, M., & Pan, W. (2020). Understanding the determinants of construction robot adoption: Perspective of building contractors. *Journal of Construction Engineering and Management*, 146(5), 04020040.
- Pan, M., Linner, T., Pan, W., Cheng, H. M., & Bock, T. (2020a). Influencing factors of the future utilisation of construction robots for buildings: A Hong Kong perspective. *Journal of Building Engineering*, 30, 101220.
- Pan, W., Chen, L., & Zhan, W. (2020b). Implications of construction vocational education and training for regional competitiveness: Case study of Singapore and Hong Kong. *Journal of Management in Engineering*, 36(2), 05019010.
- Park, S., Yu, H., Menassae, C. (2023) A comprehensive evaluation of factors influencing acceptance of robotic assistants in field construction work. *Journal of Management in Engineering* 39(3), 04023010.
- Silberer, J., Astfalk, S., & Planing, P. et al. (2023) User needs over time: the market and technology maturity model (MTMM). *Journal of Innovation and Entrepreneurship*, 12(1), 39.
- Takayama, L., Ju, W., & Nass, C. (2008). Beyond dirty, dangerous and dull: what everyday people think robots should do. In *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, 25-32.
- Taylor, J. E. (2006). Exploring the antecedents of boundary object technological change in interorganizational networks. In *Academy of Management Proceedings*, 2006(1), 1-6. Briarcliff Manor, NY 10510: Academy of Management.
- Venkatesh, V., & Bala, H. (2008). Technology acceptance model 3 and a research agenda on interventions. *Decision sciences*, 39(2), 273-315.
- Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management science*, 46(2), 186-204.
- Vora, C., Aryal, A., Willoughby, S., & Wang, C. (2024). Investigating Stakeholder Perception and Developing a Decision Framework for Robot Adoption in Construction. *Journal of Construction Engineering and Management*, 150(4), 04024012.
- World Economic Forum. (2023). Future of jobs report, Retrieved from World Economic Forum website:

<https://www.weforum.org/publications/the-future-of-jobs-report-2023/>

You, S., & Robert, L. (2018). Teaming up with robots: An IMOI (inputs-mediators-outputs-inputs) framework of human-robot teamwork. *International Journal of Robotic Engineering*, 2(3).

ASSESSMENT OF AUTOMATION POTENTIAL FOR ON-SITE CONSTRUCTION TASKS

Bing Sun¹, Chao Mao², Wen Yi³, and Tingpeng Wang⁴

1) Master's Student, School of Management Science and Real Estate, Chongqing University, Chongqing, China. Email: sunbing@stu.cqu.edu.cn

2) Professor, School of Management Science and Real Estate, Chongqing University, Chongqing, China. Email: maochao1201@126.com

3) Assistant Professor, Department of Building and Real Estate, The Hong Kong Polytechnic University, Hong Kong, China. Email: wen.yi@polyu.edu.hk

4) Ph.D. Candidate, School of Management Science and Real Estate, Chongqing University, Chongqing, China. Email: wangtingpeng@stu.cqu.edu.cn

Abstract: In the context of “machine substitution”, advanced automation technologies, such as construction robots, are anticipated to mitigate future labor shortages in the construction industry. However, the feasibility and urgency of automating various construction tasks vary significantly, indicating that not all tasks are suitable for automation. This study focuses on construction task attributes and proposes a method for assessing the automation potential of construction tasks. By comparing and analyzing 45 key construction tasks in terms of safety and technology dimensions, it is found that: decoration and furnishing construction tasks have a higher potential for automation than the main structure construction tasks; the rebar cage welding task has the lowest potential for automation and the floor skimming task has the highest automation potential; curtain wall installer has the highest automation potential, followed by painter. This study would provide valuable insights for investment decisions in construction robots and the future strategic allocation of construction tasks between human workers and robots.

Keywords: Automation potential, Construction Robot, Construction task

1. INTRODUCTION

The construction industry is facing severe low productivity and skilled labor shortages (McKinsey Global Institute, 2017; AGC, 2020). Automation technologies such as construction robotics are expected to be an effective means of increasing productivity and improving safety and health in the construction industry (Warszawski and Rosenfeld, 1994; Laborde and Sanvido, 1994; ABB, 2021). However, the adoption of robotics in the construction industry is not optimistic, especially in on-site construction (Delgado et al., 2019). ABB's global survey results show that only 55% of construction companies use robots, much lower than the 84% in the automotive industry and 79% in the manufacturing industry (ABB, 2021). One of the main reasons is that construction activities are complex, decentralized, highly nonlinear, and unpredictable, compared with the streamlined, standardized, and forward-lagging production of other industries (Ma et al., 2022). Therefore, not all construction tasks can be and are suitable for automation. Only with a full understanding of the construction task and a determination of whether construction robots are of sufficient value will cost-sensitive construction companies be willing to invest in and adopt construction robots (Melenbrink et al., 2020). Similarly, it is more profitable for construction robot manufacturers to analyze the needs of construction companies and develop construction robots with more market potential.

There are existing studies that assess the risk of automation and the degree of replacement of workers from a competency perspective. Paolillo et al. (2022) calculated the risk of automation for nearly 1,000 existing occupations by quantitatively assessing the extent to which robotics and AI capabilities could replace the human capabilities required for these jobs. Ma et al. (2022) determined the degree to which construction robots can replace construction workers by analyzing four competency dimensions: perceptual, analytical, decision-making, and executive skills. However, the comparison of robot and human capabilities ignores the fact that robots may far outperform workers in specific tasks in the future. Additionally, using occupation as a level of assessment is not suitable for construction robots. Because a type of workers may perform multiple tasks, whereas the currently prevalent construction robots are often single-task. Malik and Bilberg (2019) identified the automation potential of a gear assembly task based on the physical characteristics of the part and its task description. However, their method is clearly not applicable to construction tasks.

On this basis, the aim of this study is to propose a model for evaluating the automation potential of construction tasks based on the construction task characterization perspective and to explore the automation potential of key construction tasks with a view to providing useful insights for the development of construction robots.

2. METHODOLOGY

2.1 Research Framework

The purpose of this study is to analyze construction task characteristics, quantify construction task automation potential, and rank automation priorities for construction tasks. In this study, automation potential refers to the extent to which a task or an activity can be performed or replaced by automated technologies, mainly construction robots. First, this study constructs a model for assessing the automation potential of construction tasks from the safety dimension and the technology dimension. The safety dimension reflects the urgency of these tasks to be robotized in terms of two metrics: ergonomic risk and safety risk. The technology dimension reflects the feasibility of construction tasks being robotized from two perspectives: complexity and environmental difficulty. Then, the proposed model is utilized to assess the automation potential of 45 key on-site construction tasks. Based on the evaluation results, recommendations are provided for research and development decisions of construction robot manufacturers and investment decisions of construction contractors.

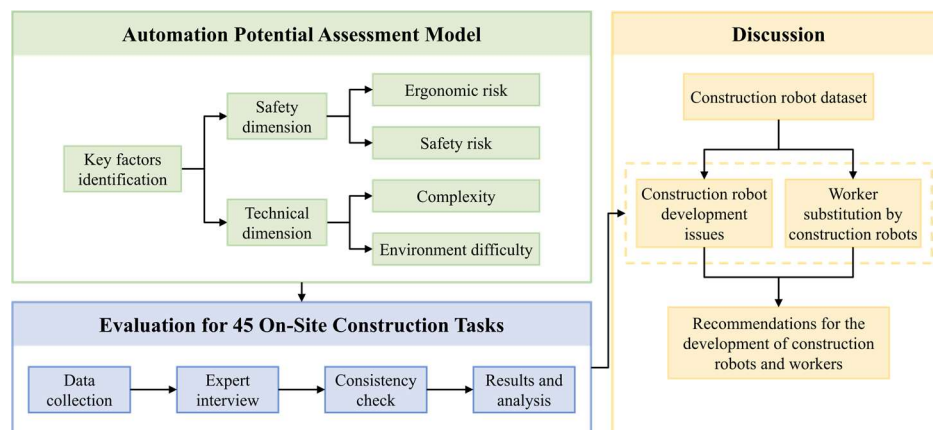


Figure 1. Research Framework

2.2 Automation Potential Assessment Model

(1) Ergonomic Risk

Construction workers, who often perform long, repetitive, and physically demanding manual tasks in the workplace, are exposed to significant ergonomic risks (Wang et al., 2023). The current ergonomic assessment methods can be categorized into manual assessment (e.g., self-reporting and systematic observation) and automated assessment using intelligent devices (e.g., computer vision, wearable sensors, and multimodal fusion-based methods) (Li et al., 2024; Tian et al., 2022). Expert observation and quantification using ergonomic scales is the most widely recognized and proven technique for assessing ergonomic risk. This study utilizes the Rapid Entire Body Assessment (REBA) to assess the ergonomic risk of construction workers performing various construction tasks. The assessment should contain at least one complete action cycle when videoed or observed on-site. After fully understanding the REBA method, each expert selects multiple representative postures from the videos of the construction tasks for ergonomic risk assessment. The maximum of all posture scores is selected as the result of one expert's assessment for a construction task.

(2) Safety Risk

According to ABB's report, the most important anticipated advantages that construction robots will provide are bettering the working environment for workers to prevent them from engaging in risky tasks and enhancing their safety and health (ABB, 2021). This study uses the British Standards Institution's risk matrix, which divides construction tasks into three risk levels by combining the likelihood of injury with its potential severity (British Standards Institution, 2004). Based on the difference in research objectives, different risk types have different criteria. Tables 1 and 2 present the assessment criteria for this study, which focuses on the safety risk that workers face while completing each construction task (Gul and Guneri, 2016).

Table 1. Assessment criteria for severity of consequences (Gul and Guneri, 2016)

Assignment	Rating category	Description
1	Insignificant	No loss of working hours and requiring first aid

2	Minor	No loss of working days, requiring outpatient treatment without a lasting impact and requiring first aid
3	Moderate	Minor injury, requiring inpatient treatment
4	Major	Major injury, requiring long-term treatment and therapy, occupational disease
5	Catastrophic	Death, permanent total disability

Table 2. Assessment criteria for the possibility (Gul and Guneri, 2016)

Assignment	Rating category	Description
1	Rare	Hardly ever
2	Unlikely	Remote (once a year), only in abnormal conditions
3	Possible	Occasional (a few events in a year)
4	Likely	Frequent (monthly)
5	Almost certain	Very frequent (once a week, every day), under normal working conditions

Table 3. Safety risk scoring criteria

Likelihood severity	multiplied by	Risk level	Score
1-6		Low	0
8-15		Medium	0.5
16-25		High	1

(3) Complexity

Highly complex tasks have high uncertainty and failure probability (Gutman et al., 2021). More advanced perceptual and manipulative control are needed for intricate coordinated movements, precise grasping, or hand control, which may not be compatible with the current state of robotic autonomy. (Lamon et al., 2019). In addition, increased task complexity requires robots to have more skills (George et al., 2023). Whereas, current popular single-task robots have limited skills, which makes it difficult to solve high-complexity tasks. Although multi-robot systems can simplify a task overall, the programming effort required can be greatly increased due to their requirement for precise coordination and teamwork (George et al., 2023). The complexity assessment equations for construction tasks oriented to human-robot collaboration are suggested based on the information entropy theory and complexity models in previous studies.

An action primitive library for construction robots is constructed by analyzing the action primitives, motion primitives, and therbligs proposed in the recent literature related to construction and manufacturing tasks and robots. Based on the action primitive, this study measures the action scale of construction tasks by analyzing the number of action primitives that can add value to each construction task (Gao, 2018). Equation (1) shows the method for calculating the scale of construction tasks.

$$H(x) = \ln A \quad (1)$$

Where $H(x)$ is the action scale of the construction activity, and x is the construction activity analyzed, A is the number of action primitives.

Due to their specialized end-effectors, construction robots are less able to adapt to various situations and objects (Qiao et al., 2022). But it is easier for them to manage standardized materials with pre-programming (Strategic Research Agenda, 2016). For fragile or highly sensitive materials, robots require precise force control or specialized end-effectors capable of evenly distributing pressure. In addition, Furthermore, different construction tasks demand varying levels of precision, with complexity influenced by factors such as configuration, accuracy, and other specific task requirements (George et al., 2023). Therefore, task difficulty is assessed from the degree of standardization and sensitivity of construction materials, and placement or connection requirements of the process (as shown in Table 4).

Table 4. Task difficulty scoring criteria

Category	Indicator	Attribute	Score
Material	Degree of standardization	Standardized	0
		Semi-standardized	0.5
		Non-standardized	1
	Sensitivity	Robust	0
		Damage at high force	0.25

Process	Placement connection requirements	or	Damage in light force	0.75
			Highly sensitive	1
			Not required	0
			Release of an object at a known predefined position, but the final orientation of the object is not required	0.25
			Placed in a predefined position with a fixed orientation	0.75
			Placed in a predefined position with a fixed orientation and required to be aligned with existing objects	1

The equation for calculating the material difficulty of a construction activity is:

$$O(x) = \frac{\sum_1^M \sum_1^N O_{m,n}}{MN} \quad (2)$$

Where $O(x)$ is the material difficulty of the construction activity x , and $O_{m,n}$ denotes the value of the n -th material attribute of the m -th material in the construction task, M is the total number of materials involved in the construction task, and N is the total number of material attributes evaluated.

The process difficulty is calculated as:

$$P(x) = \frac{\sum_1^L \sum_1^K P_{l,k}}{LK} \quad (3)$$

Where $P(x)$ is the process difficulty of the construction activity, $P_{l,k}$ denotes the value of the k -th process attribute at the l -th step of the construction task, L is the number of processes included in the construction task, and K is the total number of process attributes evaluated.

The difficulty of the construction task x is then a weighted average of the material difficulty and the process difficulty:

$$K(x) = \frac{O(x) \sum_1^M \sum_1^N O_{m,n} + P(x) \sum_1^L \sum_1^K P_{l,k}}{\sum_1^M \sum_1^N O_{m,n} + \sum_1^L \sum_1^K P_{l,k}} \quad (4)$$

Construction activities are quasi-repetitive. For example, each brick on the same brick wall is in a different location but in the same orientation, whereas each concrete molding installed may be in a different orientation and location. This necessitates the parametrization of motions which increases the difficulties to explore solutions through methods like reinforcement learning (Wang et al., 2023). Therefore, the diversity is assessed by counting the number of features that can lead to internal differences in construction activities, such as orientation, location, and material types (as shown in Equation 5).

$$D(x) = -p \cdot \ln p \quad (5)$$

Among them $p = \frac{1}{q}$, q is the number of difference characteristics of the construction task.

Then the complexity is calculated by the formula:

$$C(x) = K(x)(1 + D(x))H(x) \quad (6)$$

(4) Environment Difficulty

Construction robots frequently interact with their surrounding environment. As the complexity of the environment increases, the robot's level of autonomy decreases, which can consequently impact its overall performance (Lampe and Chatila, 2006). In this study, the construction environment difficulty is evaluated in terms of two dimensions: lighting and ground conditions. The environmental difficulty score was the average of the lighting score and the ground condition score.

Table 5. Environment difficulty scoring criteria

Category	Attribute	Score
Lighting	Poor	0
	Good	1
Ground condition	Poor	0
	Good	1

(5) Automation Potential Score

The four metrics were normalized using the Min-Max Normalization method and summed to obtain the automation potential score for each construction task.

3. EVALUATION OF AUTOMATION POTENTIAL FOR ON-SITE CONSTRUCTION TASKS

3.1 Data Collection

Through conducting interviews with experts with rich working experience in on-site construction, this study identifies 45 key construction tasks based on importance, repetition, duration, and frequency of occurrence. The 45 construction tasks extracted were relatively comprehensive,

covering two scenarios and eight types of work. Two experts in the field of construction were invited to evaluate the ergonomic risks, and two mid-level engineers from construction companies with many years of experience were invited to evaluate the environmental difficulties of each construction task, respectively. The complexity is evaluated based on the proposed action primitive library and information entropy model. The results of the two experts and two engineers were each tested for consistency and their average scores were taken as the final score.

Table 6. 45 key construction tasks in on-site construction

Scenario	Construction task		Abbreviation	Construction task		Abbreviation
Main Structure	Concrete Pouring	Floor	CFP	ALC Wall Panel Installation		AWPI
	Concrete Leveling	Floor	CFL	ALC Wall Panel Fixing		AWPF
	Concrete Smoothing	Floor	CFS	ALC Wall Panel Interconnection		AWPIT
	Concrete Troweling	Floor	CFT	Vertical Rebar Installation		VRI
	Concrete Grinding	Floor	CFG	Vertical Rebar Welding		VRW
	Concrete Cutting	Floor	CFCT	Rebar Cage Welding		RCW
	Concrete Grinding	Wall	CWG	Rebar Cage Tying		RCT
	Concrete Cutting	Wall	CWC	Slab Rebar Tying		SRT
	Concrete Grinding	Ceiling	CCG	Concrete Formwork Installation		CFIS
	Brick Masonry	Wall	BWM	Concrete Formwork Connection		CFCN
	Wall Spraying		WSP	Wall Tiling		WT
	Wall Rolling		WR	Roof Tile Laying		RTL
	Wall Grinding		WSD	Glass Curtain Wall Installation		GCWI
	Wall Wallpapering		WW	Glass Curtain Wall Connection		GCWC
	Ceiling Spraying		CSP	Ceiling Installation	Frame	CFI
	Ceiling Rolling		CR	Ceiling Connection	Frame	CFC
	Ceiling Grinding		CS	Ceiling Gypsum Board Installation		CGBI
	Floor Spraying		FSP	Ceiling Gypsum Board Connection		CGBC
	Floor Skimming		FSK	Wall Frame Installation		WFI
	Floor Rolling		FR	Wall Frame Connection		WFC
Decoration and Finishing	Wall Plastering		WP	Wall Gypsum Board Installation		WGBI
	Ceiling Plastering		CP	Wall Gypsum Board Connection		WGBC
	Floor Tiling		FT			

3.2 Results and Analysis

(1) Comparison of Automation Potential for Decoration and Furnishing and Main Structure Construction Tasks

The average automation potential for the decoration and furnishing tasks was 2.3080, while the average value for the main structure construction was below 1.8857, which is lower than the average automation potential of 2.1203 for the 45 construction tasks. The main reason may be that the decoration and furnishing tasks are usually indoor, with the floors having been leveled and hardened. Moreover, the main structure construction tasks are usually outdoors at night, relying on searchlights on tower cranes that are far away from the workers and have poor lighting conditions. Decoration and furnishing tasks,

on the other hand, can utilize temporary lighting systems indoors to ensure adequate lighting of the work area.

(2) Differences in Automation Potential Across Construction Tasks

The construction task with the highest potential for automation is floor skimming, at 2.8954. It is technically feasible because the construction steps and objects are simple, single, and performed indoors. It also has a high ergonomic risk due to the fact that the worker has to crouch on the floor and swing his arms considerably, with an unstable center of gravity of the body. Next is the connection and installation of glass curtain walls. Because they need the workers to work outdoors or even at height, and glass curtain walls are usually heavy leading to workers' high ergonomic and safety risks. Robots, on the other hand, are very good at solving such problems. A wide range of curtain wall installation robots such as the Hephaestus are already available. The lowest potential for automation is for rebar cage welding at 0.9964, followed by vertical rebar installation at 1.0199. Although welding robotic arms have been very well developed in industry, when fabricating steel cages at construction sites, the ground is usually not hardened and is still a dirt floor that is not easy for robots to walk around. Floor tiling has the third lowest automation potential. Tiling is a complex task that involves not only two easily damaged materials, ceramic tile, and mortar, but also a number of actions, including picking up a tile, taking adhesive, applying adhesive, positioning the tile, apply pressure to the tiles.

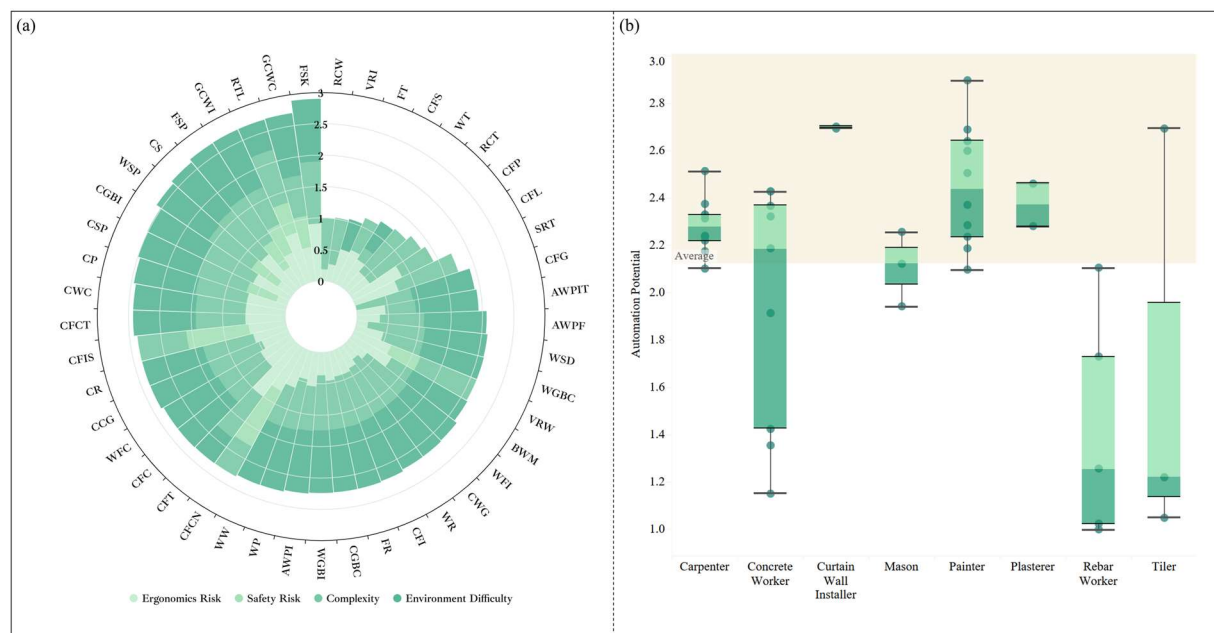


Figure 2. Results of automation potential assessment (a) automation potential for each construction task (b) automation potential for each construction worker type

(3) Differences in Automation Potential by Construction Worker Types

The type of construction worker most likely to be replaced by robots and with the highest potential for automation is the curtain wall installer with a mean of 2.6964, followed by the painter with a mean of 2.4353. Most of the construction tasks undertaken by the carpenter, painter, and plasterer have higher than average automation potential. Concrete workers, masons, rebar workers, and tilers are responsible for construction tasks that vary widely in automation potential. Especially for tilers, the difference between tile laying and floor tiling is 1.6475.

4. DISCUSSION

4.1 Construction Robotics Companies Prefer to Develop Robots with High Technical Feasibility

101 construction robots that can be applied to on-site construction in China currently available on the market are collected and they are matched with 45 construction tasks and 8 construction worker types. Scatter plots are drawn based on the safety index (sum of ergonomics score and safety risk score) and technology index (sum of complexity score and environmental difficulty score) of the construction tasks (as shown in Figure 3a). The automation potential was categorized into four quadrants using the mean values of the technology index and safety index as boundaries: quadrant I (high urgency, high feasibility), quadrant II (low urgency, high feasibility), quadrant III (low urgency, low feasibility) and

quadrant IV (high urgency, low feasibility).

Overall, the development of construction robots is on a reasonable path, taking into account both urgency and feasibility. 58.6207% of highly feasible construction tasks and 55.0000% of highly urgent construction tasks are already being progressively automated with robots. Quadrant II has the highest number of robotic products at 53, accounting for 52.4752%. On the contrary, quadrant III has the lowest number of robots, at only 11.8812%. Only half of the construction tasks in quadrant IV are starting to be substituted by robots. This is consistent with the results of our enterprise survey, i.e., enterprises are more inclined to develop construction robots whose necessary hardware and software have been mature in order to capture the market for construction robots and obtain policy incentives.

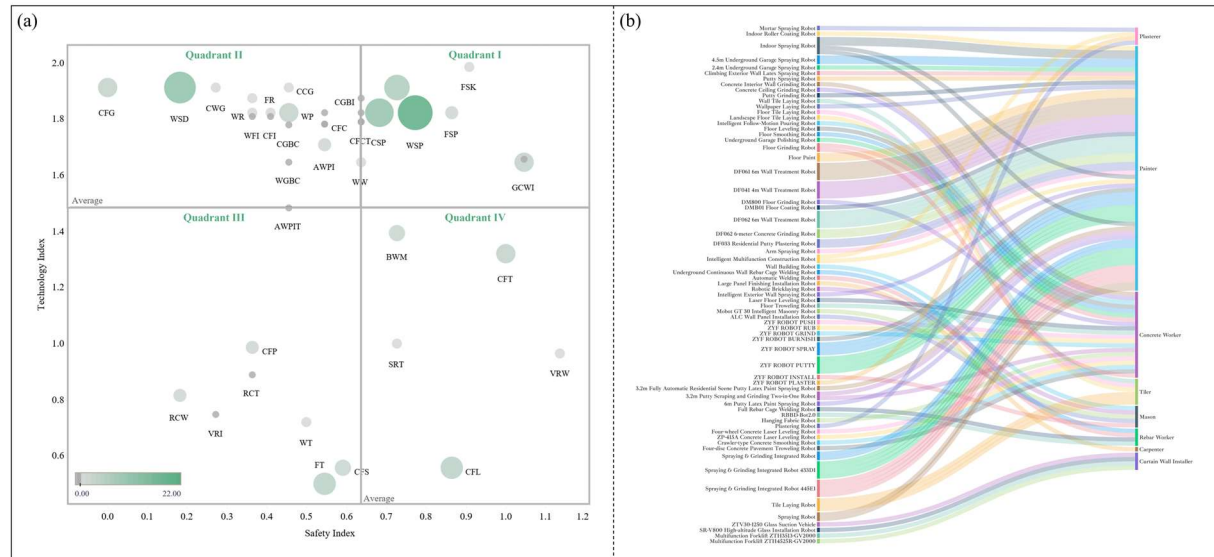


Figure 3 Current development of construction robots in China (a) distribution of construction robot products in 4 quadrants (b) mapping of construction robot products to construction worker types

4.2 Painters Are the Jobs Most Likely to Be Replaced by Robots

Figure 3b shows the current status of “machine substitution” for various types of construction workers. The highest potential for replacement by construction robots is for painters. Construction robots can currently spray paint and putty and other coatings on columns, walls, floors, and ceilings. The concrete worker ranks second. Nowadays, construction workers and robots can collaborate easily when constructing concrete floors by using a group of robots that can pour concrete, level floors, smooth floors, and polish underground garages. The carpenter is the hardest of the eight worker types to replace. Only a small number of companies are currently developing large panel finishing installation robots. Therefore, it is recommended that when training new workers, construction companies should, not only gradually develop workers’ skills in using automated tools such as robots, but purposefully train for worker types that are more difficult to replace by automation and have a larger labor gap in the future.

5. CONCLUSIONS

There is a trend towards the gradual automation of construction tasks. Ranking the automation potential of construction tasks can, to a certain extent, guide the development of automation technologies such as construction robots and predict labor changes on construction sites. To fill the gap in current research, this study proposes a model for assessing the automation potential of construction tasks and evaluates the automation potential of 45 key construction tasks. This study analyzes from three dimensions: construction scenario, task type, and worker type, and draws the following conclusions: the automation potential of decoration and finishing tasks is higher than that of main structure construction tasks; the automation potential of rebar cage welding is the lowest, and the automation potential of floor skimming is the highest; and the automation potential of curtain wall installers is the highest, followed by painters. Combining the results with the construction robot products currently available in the Chinese market, the analysis found that: the current development of robots in the construction field takes both urgency and feasibility into account, but robot manufacturers prefer to develop products with high technological feasibility; painters are the worker type most likely to be replaced by robots. This study not only provides suggestions for the development of construction robots but also serves as a basis for human-robot task allocation in the construction field.

ACKNOWLEDGMENTS

This study was supported by the National Natural Science Foundation of China (Grant No. 72471039 and 72201229), the Humanities and Social Science Fund of Ministry of Education of China (Grant No. 24YJA630064), the Natural Science Foundation of Chongqing (Grant No. CSTB2022NSCQ-MSX1622), and the Research Grants Council of the Hong Kong Special Administrative Region, China (Grant No. 25211722).

REFERENCES

- ABB (2021). ABB Robotics 2021 Construction Survey <https://express.adobe.com/page/5QaTFLrupXbYh/>.
- AGC (2020). 2020 Construction Outlook Survey Results National Results. https://www.agc.org/sites/default/files/Files/Communications/2020_Outlook_Survey_National.pdf.
- British Standards Institution (2004). Occupational health and safety management systems. *BSI Business Information*.
- Delgado, J. M. D., Oyedele, L., Ajayi, A., Akanbi, L., Akinade, O., Bilal, M., and Owolabi, H. (2019). Robotics and automated systems in construction: Understanding industry-specific challenges for adoption. *Journal of Building Engineering*. 26: 11.
- Gao, Y. (2018). The Task Allocation of Human-robot Collaboration (HRC) Based on Complexity. Master Thesis.
- George, P., Cheng, C. T., Pang, T. Y., and Neville, K. (2023). Task Complexity and the Skills Dilemma in the Programming and Control of Collaborative Robots for Manufacturing. *Appl. Sci.-Basel*. 13(7): 26.
- Gul, M., and Guneri, A. F. (2016). A fuzzy multi criteria risk assessment based on decision matrix technique: A case study for aluminum industry. *Journal of Loss Prevention in the Process Industries*. 40: 89-100.
- Gutman, D., Olatunji, S., and Edan, Y. (2021). Evaluating Levels of Automation in Human-Robot Collaboration at Different Workload Levels. *Appl. Sci.-Basel*. 11(16): 15.
- Laborde, M., and Sanvido, V. (1994). Introducing New Process Technologies into Construction Companies. *Journal of Construction Engineering and Management*. 120(3): 488-508.
- Lamon, E., Franco, A. D., Peternel, L., and Ajoudani, A. (2019). A Capability-Aware Role Allocation Approach to Industrial Assembly Tasks. *IEEE Robotics and Automation Letters*. 4(4): 3378-3385.
- Lampe, A., and Chatila, R. (Year) Published. Performance measure for the evaluation of mobile robot autonomy. *Proc., Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006.*, 4057-4062.
- Li, Z., Yu, Y., Xia, J., Chen, X., Lu, X., and Li, Q. (2024). Data-driven ergonomic assessment of construction workers. *Automation in Construction*. 165: 105561.
- Ma, X., Mao, C., and Liu, G. (2022). Can robots replace human beings? —Assessment on the developmental potential of construction robot. *Journal of Building Engineering*. 56: 104727.
- Malik, A. A., and Bilberg, A. (2019). Collaborative robots in assembly: A practical approach for tasks distribution. *Procedia CIRP*. 81: 665-670.
- McKinsey Global Institute (2017). Reinventing construction through a productivity revolution. McKinsey.
- Melenbrink, N., Werfel, J., and Menges, A. (2020). On-site autonomous construction robots: Towards unsupervised building. *Automation in Construction*. 119: 21.
- Paolillo, A., Colella, F., Nosengo, N., Schiano, F., Stewart, W., Zambrano, D., Chappuis, I., Lalive, R., and Floreano, D. (2022). How to compete with robots by assessing job automation risks and resilient alternatives. *Science Robotics*. 7(65): eabg5561.
- Qiao, H., Zhong, S., Chen, Z., and Wang, H. (2022). Improving performance of robots using human-inspired approaches: a survey. *Science China Information Sciences*. 65(12): 221201.
- Strategic Research Agenda (2016). Robotics 2020 Multi-Annual Roadmap. https://eu-robotics.net/wp-content/uploads/Multi-Annual_Roadmap2020_ICT-24_Rev_B_full.pdf.
- Tian, Y., Li, H., Cui, H., and Chen, J. (2022). Construction motion data library: an integrated motion dataset for on-site activity recognition. *Scientific Data*. 9(1): 726.
- Wang, J., Li, X., Han, S., and Al-Hussein, M. (2023). 3D standard motion time-based ergonomic risk analysis for workplace design in modular construction. *Automation in Construction*. 147: 104738.
- Wang, X., Wang, S., Menassa, C. C., Kamat, V. R., and McGee, W. (2023). Automatic high-level motion sequencing methods for enabling multi-tasking construction robots. *Automation in Construction*. 155: 105071.
- Warszawski, A., and Rosenfeld, Y. (1994). Robot for Interior - Finishing Works in Building: Feasibility Analysis. *Journal of Construction Engineering and Management*. 120(1): 132-151.

A Simulation-based Genetic Algorithm Schedule Optimization Method for Bridge Construction

Yuan Yao¹, Vivian WY Tam^{2,*}, Jun Wang³, Khoa N Le⁴, Anthony Butera⁵, and Wenchi Shou⁶

1) Ph.D. Candidate, School of Engineering, Design and Built Environment, Western Sydney University, Australia. Email: 22050001@student.westernsydney.edu.au

2, *) Ph.D., Prof., School of Engineering, Design and Built Environment, Western Sydney University, Australia. Email: vivianwytam@gmail.com

3) Ph.D., Lecturer, School of Engineering, Design and Built Environment, Western Sydney University, Australia. Email: Jun.wang@westernsydney.edu.au

4) Ph.D., A/Prof, School of Engineering, Design and Built Environment, Western Sydney University, Australia. Email: K.Le@westernsydney.edu.au

5) Ph.D., Lecturer, School of Engineering, Design and Built Environment, Western Sydney University, Australia. Email: A.Butera@westernsydney.edu.au

6) Ph.D., Lecturer, School of Engineering, Design and Built Environment, Western Sydney University, Australia. Email: W.Shou@westernsydney.edu.au

Abstract: Off-site construction has become widely acknowledged for its advantages, such as saving time, enabling faster assembly, and being cost-efficient. The sector's rapid growth has driven the demand for more advanced and effective methods of construction scheduling. Construction scheduling is naturally complicated due to the numerous constraints it involves, including those connected to workforce and resource availability. Conventional approaches, like the Critical Path Method (CPM), fail to account for multiple constraints, which limits their effectiveness in practical project scenarios. This research presents a simulation-based Genetic Algorithm (S-GA) approach to develop optimal construction schedules while accounting for constraints in labour and resources. Reducing the total project duration is the objective of proposed method. The proposed S-GA framework enhances the ability to manage scheduling across all construction phases. A real-world case which contains a prefabricated bridge with 6 spans was conducted to assess the method. For comparison, traditional methods and the evolution algorithm (EA) were adopted, and the findings indicated that S-GA not only produced superior construction schedules but also operated with less computational time. Proposed S-GA generated the best construction schedule with shortest project duration within least computational time. As a result, the proposed approach offers an advanced scheduling method that is applicable to real-world construction projects. Project managers could use proposed method to make plans for their construction projects.

Keywords: Simulation; Genetic Algorithm; Automated Scheduling; Scheduling Optimization

1. INTRODUCTION

Automated construction scheduling offers project managers support in effectively handling labor, equipment, timelines, costs, and other key project aspects. This approach serves as a decision-making tool, pinpointing necessary tasks and establishing both the timing and method for their execution. According to Ding et al. (2023), automated scheduling in construction aligns with the Resource Constrained Project Scheduling Problem (RCPSP) framework. Typically, a construction schedule divides the entire workload into a hierarchy of work breakdown structures and specific tasks, which are then assigned to corresponding subcontractors (Bai et al., 2009). Construction planning not only includes scheduling but also integrates various planning functions, such as material handling, site layout arrangement, equipment mobilization, and general site logistics. A skilled scheduler with substantial prior experience is required to carefully manage all on-site information to complete a construction schedule, although this task can demand considerable time and effort (Amer et al., 2021).

A range of methods and theories have been formulated to support construction scheduling. Construction project scheduling, as a decision-making process, seeks to determine the best sequence of activities and allocate diverse resources to tasks while navigating a range of constraints (Zhou et al., 2013). The Critical Path Method (CPM) is one of the most widely applied techniques (Koskela, 2000), breaking down projects into distinct work breakdown structures (WBS). It requires manual work to reorder WBS for better scheduling. Alamode and Plaza (1994) introduced case-based reasoning (CBR) techniques, where past cases are used as references to solve new scheduling challenges. Although these techniques rely heavily on prior experiences, they often lack the flexibility necessary for managing a range of construction projects. Heuristic methods and Genetic Algorithms (GA) offer a different approach to scheduling by framing it as an optimization problem, where specific conditions limit one or more objectives to be achieved (Ahmed et al., 2021; Bettemir & Sonmez, 2015; Erdal & Kanit, 2021; Lin et al., 2022; Yuan et al., 2021). Christodoulou (2010) introduced an agent-based ant colony optimization (ACO) approach to address resource limitations in construction projects. Cheng et al. (2016) applied a symbiotic organisms search (SOS) method for tackling multiple project scheduling challenges. El-Rayes and Jun (2009) used a genetic algorithm to minimize unwanted resource fluctuations and reduce idle time. Ma et al. (2021) examined

the construction order for various PC and CiS components and introduced a multi-objective discrete symbiotic organisms search method aimed at reducing the project makespan. Liu et al. (2021) created a heuristic-driven GA designed to assist construction managers in reaching target profits and enhancing project oversight. Despite their strengths, these methods generally struggle with adaptability to various construction. Organizing tasks in real-world construction projects presents challenges, as multiple constraints—such as labor, resources, and construction techniques—determine the sequence in which activities can proceed.

Combining simulation methods with GA (S-GA) to solve schedule optimization problems offers several advantages. First, simulation provides a realistic model of the complex, dynamic interactions within scheduling systems, capturing variability in factors such as resource availability, processing times, and demand fluctuations. This realism enhances the quality of solutions, as they are tested in conditions that closely mimic actual operations. Meanwhile, GAs offer a robust search mechanism capable of exploring a wide solution space by using evolutionary principles of selection, crossover, and mutation. The combination allows for effective exploration and exploitation of scheduling options, as GAs can efficiently search through potential solutions while the simulation evaluates each solution's performance in a realistic environment. This hybrid approach is especially beneficial for complex scheduling problems, where traditional optimization methods may struggle due to non-linearity or high-dimensionality. In summary, integrating simulation with GA enhances solution robustness, adaptability, and practical applicability in real-world scheduling environments.

2. METHOD

2.1 Problem Definition

The construction process for a prefabricated bridge is organized into different structural categories, including the superstructure and the substructure. Figure 1 presents a simplified WBS of a bridge, divided into two main sections: the superstructure and the substructure. The superstructure includes precast girders and joints, with 3 and 4 sub-activities, respectively. The substructure comprises precast pier caps, precast piers, pile caps, and piles, with each component involving 4, 4, 6, and 7 sub-activities, respectively. This hierarchical structure helps to organize and manage the construction tasks involved in building the bridge.

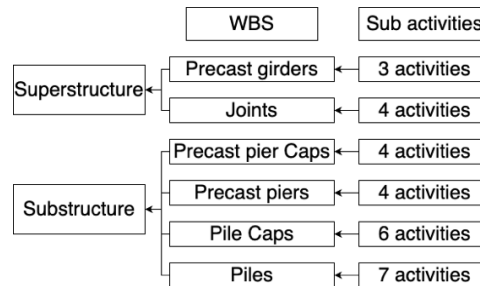


Figure 1: Construction activities representation

The proposed optimization focuses on minimizing the total duration of the construction project to achieve earlier completion than scheduled (Equation 1). This approach operates within the boundaries of resource limitations and task order dependencies (Equations 2 to 4).

Objective:

$$\text{minimize } D_{\text{project}} \quad (1)$$

Subject to:

$$\sum R_t(a_{i,j,k}) \leq R \quad (2)$$

$$t_{k+1} - t_k \geq d_k \quad (3)$$

$$N_c = 0 \quad (4)$$

Where D_{project} denotes the total duration of the project; $R_t(a)$ denotes the resources required for activity a ; R represents the resources for each operation; t_k denotes the starting time of the k^{th} activity; d_k represents the duration of k^{th} activity; N_c denotes the number of constraint violations.

2.2 Environment Development

This flowchart (Figure 2) represents a simulation process for generating and evaluating a genetic algorithm chromosome in a scheduling optimization context. The process begins with the initialization of a new chromosome, which represents a potential solution by encoding a sequence of activities to be executed. Structure groups within the chromosome are identified, and activities are selected and looped through, where each activity's crew and equipment requirements are allocated. Activities are then added to an activity buffer and executed, with

their duration monitored and incremented. For each activity, the system checks if its duration has been completed; once an activity ends, the allocated resources (crew and equipment) are released. Completed activities are recorded along with essential information such as structure group ID and segment ID. After each activity completion, the simulation updates three main states: the resource usage state (R_t), which tracks resource availability and utilization; the structure state (G_t), which monitors the progress of each structure group; and the ongoing activity state (P_t), which reflects the progression of activities. This looping process continues until all activities within the chromosome have been executed. By simulating the execution and resource allocation for each chromosome, this process allows the GA to evaluate and optimize scheduling performance, promoting efficient resource usage and timely completion of tasks.

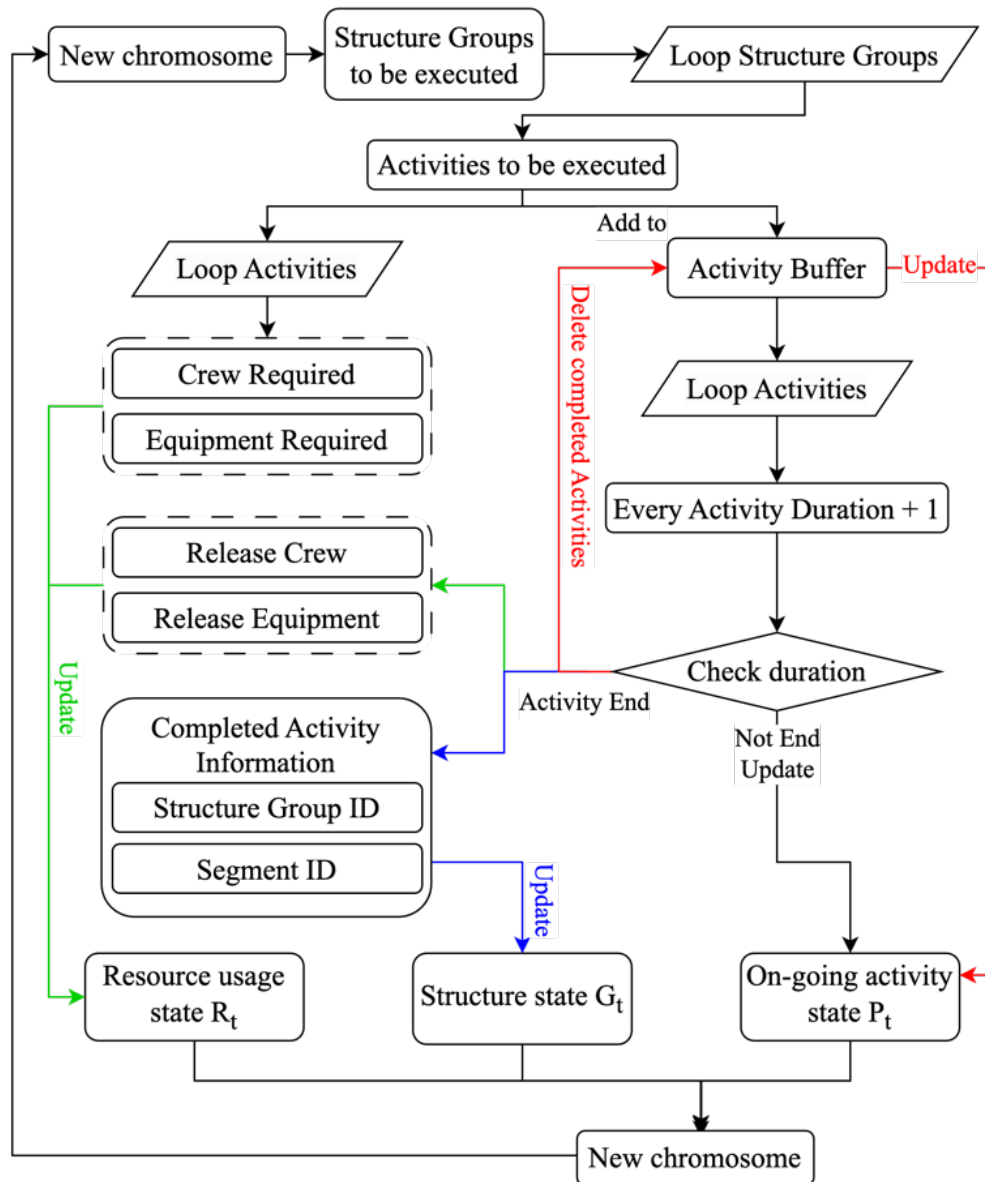


Figure 2: Simulation environment

2.3 GA Design

Due to the diverse cases and assumptions in existing studies, current scheduling algorithms are not directly applicable to this research. To address this, a new approach is developed that integrates a simulation-based GA. This approach merges a modified S-GA with a simulation model, where each gene represents a specific construction activities, and the gene sequence determines the order in which these activities are arranged. The model tackles construction scheduling challenges by using the simulation environment as a decoder for GA chromosomes, ensuring optimal utilization of resources and labours. For generating the initial population, the algorithm employs regret-based biased random sampling (RBRS) alongside a serial schedule generation scheme

(SSGS). The evolutionary process involves selecting the best individuals, performing two-point crossover, swapping genes (swap mutation), and applying a mutation method with a probability that gradually changes over time (parameters detailed in Table 1). The study's results are compared with a conventional scheduling method, namely Critical Path Method (CPM) and an artificial intelligence method, namely Q-Learning. The GA configurations are displayed in Table 1.

Table 1: S-GA parameters

Population Size	200
Crossover Rate	0.8
Mutation Rate	0.2
TOP	0.15

2.4 Case Study

This study focuses on a prefabricated construction project located on Longdong Avenue in Shanghai, China. In prefabricated bridge construction, the majority of above-ground components are cast off-site. The on-site construction work breakdown structure (WBS) includes tasks such as pile construction and platform casting. Following these steps, precast structures—such as columns, cover beams, and girders—are assembled sequentially. Multiple crews will handle various tasks with different equipment types. For instance, a dedicated lifting crew, equipped with a walking crane, will manage activities like lifting piles, piers, pier caps, and girders. In this study, the construction process is broken down into 28 distinct activities, involving 12 crews and 4 types of equipment. Segment classifications are based on bridge structures and construction methods, including: 1) piles, 2) pile caps, 3) piers, 4) pier caps, 5) girders, and 6) seams. The 9-span bridge consists of 10 sets each of piles, pile caps, piers, and pier caps, along with 9 sets of girders and seams. Each segment encompasses various activities, totalling 273 sub-activities. There are 4 lifting and hammer & piling crews, while other crews number 2 each. The equipment includes 4 units of mobile cranes and hammers, with 2 units each of other equipment.

3. RESULTS

The result of proposed method is compared with other algorithms and displayed in Table 2. The manual schedule is derived from the initial master plan, which sets a 45-day completion target for the project. With a standard workday of 8 hours, this timeframe equates to a total construction period of 360 hours. The manual outcome is then modified with CPM and the total duration is shortened to 336h. Q-learning performs better than traditional methods with better schedule duration. GA generates the best outcome with 310h. From the perspective of algorithm running time, Q-learning runs for over 500s and GA only costs 135.1s.

Table 2: Scheduling results

Method	Manual	CPM + manual	Q-Learning	S-GA (150 iters)
Duration (h)	360	336	321	310
CPU times	/	/	518.7s	135.1s

4. DISCUSSION

The results summarized in Table 2 highlight the effectiveness of various scheduling methods, including a S-GA, Q-learning, and conventional manual and CPM-modified schedules. Each method's performance was assessed based on both the project completion time and the computational efficiency of the algorithm. The manual schedule, set against a 45-day completion target with an equivalent work period of 360 hours, serves as a baseline. When adjusted with CPM, the project duration is reduced to 336 hours. This improvement shows that the CPM method can enhance scheduling efficiency to some degree by identifying and optimizing the critical path, though its capability remains limited by the static nature of the initial master plan.

In comparison, Q-learning demonstrates a further improvement over CPM, achieving a project duration of 321 hours. As a reinforcement learning technique, Q-learning offers the advantage of learning optimal actions over time through interaction with the environment. However, its prolonged runtime of 518.7 seconds indicates a significant computational cost, which could hinder its applicability in scenarios requiring rapid scheduling adjustments or where computational resources are limited. The S-GA method outperforms all other approaches, reducing the project duration to 310 hours and achieving the best balance of efficiency and effectiveness among the tested methods. S-GA's capacity to explore a wide range of possible solutions through iterative optimization allows it to find a near-optimal solution with only 150 iterations. Notably, S-GA achieves this result in a significantly shorter time (135.1 seconds) than Q-learning, suggesting that S-GA is more computationally efficient

while also producing the most favorable project schedule.

This study exists some limitations. First, the proposed method only considers shortening the duration as the only objective. All the algorithm features are designed for single objective purpose. Secondly, proposed S-GA architecture is very simple, larger case studies is still required to test its effectiveness. Thirdly, the method considers the construction process as statistic process without unexpected disruptions.

5. CONCLUSIONS

The comparison of scheduling methods reveals that the proposed S-GA significantly outperforms traditional scheduling techniques and Q-learning in both project duration and computational efficiency. The manual schedule and CPM-modified plan provide a foundational timeline, but they lack the adaptability and optimization capacity necessary to achieve the best outcomes. Q-learning, while effective in reducing project time compared to traditional methods, comes with a high computational cost, making it less practical for environments requiring quick adjustments. In contrast, S-GA achieves the shortest project duration of 310 hours with a computation time of only 135.1 seconds, showcasing its ability to optimize effectively within a limited runtime. This combination of fast processing and superior schedule reduction underscores the suitability of S-GA for complex scheduling tasks where both time efficiency and resource allocation are crucial. Consequently, S-GA emerges as the preferred method, providing a balanced solution for scheduling in dynamic and time-sensitive project environments.

Building on the promising results of the S-GA in optimizing project scheduling, future research could explore hybrid approaches that combine S-GA with other machine learning techniques, such as Q-learning or neural networks, to enhance both scheduling accuracy and adaptability. Hybrid models may harness the exploration capacity of reinforcement learning to guide S-GA toward more diverse solutions, potentially improving performance in highly dynamic environments where project parameters frequently change. Additionally, future work could focus on optimizing the S-GA parameters and exploring real-time scheduling adjustments in response to project changes. Integrating real-time data inputs and developing adaptive scheduling mechanisms could enable a more responsive model that recalibrates the schedule in real-time as project conditions evolve. Lastly, the application of these methods to larger, multi-phase projects could test scalability and reveal insights into S-GA's efficiency across complex, resource-intensive scheduling scenarios.

ACKNOWLEDGMENTS

The authors wish to acknowledge the financial support from the Australian Research Council (ARC), Australian Government (No: DP200100057, IH200100010; FT220100017).

REFERENCES

- Aamodt, A., & Plaza, E. (1994). Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI Communications*, 7, 39-59. <https://doi.org/10.3233/AIC-1994-7104>
- Ahmed, T., Hossain, S. M., & Hossain, M. A. (2021). Reducing completion time and optimizing resource use of resource-constrained construction operation by means of simulation modeling. *International Journal of Construction Management*, 21(4), 404-415. <https://doi.org/10.1080/15623599.2018.1543109>
- Amer, F., Koh, H. Y., & Golparvar-Fard, M. (2021). Automated methods and systems for construction planning and scheduling: Critical review of three decades of research. *Journal of Construction Engineering and Management*, 147(7), Article 03121002. [https://doi.org/10.1061/\(asce\)co.1943-7862.0002093](https://doi.org/10.1061/(asce)co.1943-7862.0002093)
- Bai, Y., Zhao, Y., Chen, Y., & Chen, L. (2009). Designing domain work breakdown structure (DWBS) using neural networks. 5553, 1146-1153. https://doi.org/10.1007/978-3-642-01513-7_127
- Bettemir, O. H., & Sonmez, R. (2015). Hybrid genetic algorithm with simulated annealing for resource-constrained project scheduling. *Journal of Management in Engineering*, 31(5), Article 04014082. [https://doi.org/10.1061/\(asce\)me.1943-5479.0000323](https://doi.org/10.1061/(asce)me.1943-5479.0000323)
- Cheng, M.-Y., Prayogo, D., & Tran, D.-H. (2016). Optimizing multiple-resources leveling in multiple projects using discrete symbiotic organisms search. *Journal of Computing in Civil Engineering*, 30(3), 04015036.
- Christodoulou, S. (2010). Scheduling resource-constrained projects with ant colony optimization artificial agents. *Journal of computing in civil engineering*, 24(1), 45-55.
- Ding, H. Y., Zhuang, C. B., & Liu, J. H. (2023). Extensions of the resource-constrained project scheduling problem. *Automation in Construction*, 153, Article 104958. <https://doi.org/10.1016/j.autcon.2023.104958>
- El-Rayes, K., & Jun, D. H. (2009). Optimizing resource leveling in construction projects. *Journal of Construction Engineering and Management*, 135(11), 1172-1180.
- Erdal, M., & Kanit, R. (2021). Scheduling of construction projects under resource-constrained conditions with a specifically developed software using genetic algorithms. *Tehnicki Vjesnik-Technical Gazette*, 28(4), 1362-1370. <https://doi.org/10.17559/tv-20200305101811>
- Koskela, L. (2000). *An exploration towards a production theory and its application to construction*. VTT

Technical Research Centre of Finland.

- Lin, J. C. W., Lv, Q., Yu, D. H., Srivastava, G., & Chen, C. H. (2022). Optimized scheduling of resource-constraints in projects for smart construction. *Information Processing & Management*, 59(5), Article 103005. <https://doi.org/10.1016/j.ipm.2022.103005>
- Liu, W. L., Zhang, J. W., & Li, W. J. (2021). Heuristic Methods for Finance-Based and Resource-Constrained Project Scheduling Problem. *Journal of Construction Engineering and Management*, 147(11), Article 04021141. [https://doi.org/10.1061/\(asce\)co.1943-7862.0002174](https://doi.org/10.1061/(asce)co.1943-7862.0002174)
- Ma, Z. L., Li, S. Y., Wang, Y., & Yang, Z. Q. (2021). Component-level construction schedule optimization for hybrid concrete structures. *Automation in Construction*, 125, Article 103607. <https://doi.org/10.1016/j.autcon.2021.103607>
- Yuan, Y. S., Ye, S. D., Lin, L., & Gen, M. S. (2021). Multi-objective multi-mode resource-constrained project scheduling with fuzzy activity durations in prefabricated building construction. *Computers & Industrial Engineering*, 158, Article 107316. <https://doi.org/10.1016/j.cie.2021.107316>
- Zhou, J., Love, P. E. D., Wang, X., Teo, K. L., & Irani, Z. (2013). A review of methods and algorithms for optimizing construction scheduling. *Journal of the Operational Research Society*, 64(8), 1091-1105. <https://doi.org/10.1057/jors.2012.174>

TOWARDS HIGHER INFORMATION DENSITY IN IMAGE TRANSMISSION: LEARNED IMAGE COMPRESSION FOR CONSTRUCTION SITE MONITORING

Jiucui liu¹, Chengzhang Chai², Linghan Ouyang³, and Haijiang Li^{4*}

1) Ph.D. Candidate, BIM for Smart Engineering Centre, School of Engineering, Cardiff University, Cardiff, UK. Email: LiuJ151@cardiff.ac.uk

2) Ph.D. Candidate, BIM for Smart Engineering Centre, School of Engineering, Cardiff University, Cardiff, UK. Email: chaic1@cardiff.ac.uk

3) Ph.D. Candidate, BIM for Smart Engineering Centre, School of Engineering, Cardiff University, Cardiff, UK. Email: OuyangL@cardiff.ac.uk

4) Ph.D., Prof., BIM for Smart Engineering Centre, School of Engineering, Cardiff University, Cardiff, UK. Email: lih@cardiff.ac.uk

* Corresponding author

Abstract: The digital transformation in the Architecture, Engineering, and Construction (AEC) sector underscores the growing need for efficient data transmission, especially in computer vision tasks that depend on the transfer of large volumes of images. In this work, a novel method is introduced to enhance data transmission efficiency in an edge-cloud coordinated architecture using Learned Image Compression (LIC). By integrating the LIC model with multiple downstream task models (Mask R-CNN and Faster R-CNN), the proposed framework aligns their respective latent features, resulting in a task-oriented LIC model that optimises compression for specific tasks. The approach increases the proportion of task-relevant information—referred to as information density—in the transmitted bitstream. Experimental results demonstrate that this method significantly reduces data transmission load while concentrating the transmitted bits on regions essential for downstream tasks, all without a notable decrease in task accuracy.

Keywords: Learned Image Compression, Machine Vision, Data Transmission, Object Detection, Instance Segmentation.

1. INTRODUCTION

1.1 Background

The Architecture, Engineering, and Construction (AEC) industry is actively pursuing productivity improvements through advanced technologies such as robotics (Zhang et al., 2023), digital twins (Tuhaise et al., 2023), and extended reality (XR) (J. C. P. Cheng et al., 2020). These technologies provide opportunities for automating human-robot interaction, robotic-based construction, and infrastructure inspection. These technologies rely on fundamental processes like data collection, transmission, and processing. With advancements in Deep Learning (DL) and camera sensors, Computer Vision (CV) has gained significant attention in the AEC sector, raising the demand for extensive image data to support tasks such as image classification, object detection, and instance segmentation (Xu et al., 2021). However, most current research focuses on enhancing task-specific accuracy through increasingly complex methods, with limited attention to evaluating the usability and real-time performance of CV algorithms from the perspectives of system architecture and data transmission.

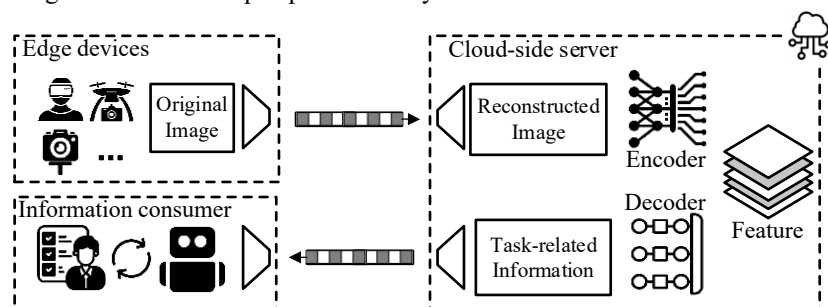


Figure 1. The Edge-Cloud System in AEC (from the Perspective of Information Flow)

Due to the high computational demands of deep learning, equipping each edge device with powerful GPUs is inefficient and costly. Consequently, many studies propose an edge-cloud architecture as shown in Figure 1, where computationally intensive tasks are handled in the cloud-side server, while edge devices focus on data collection, data transmission, and information visualization (Alizadehsalehi et al., 2020; Cheng et al., 2020). For CV-related applications, on the edge side, captured images need to be compressed and encoded as bitstream for data transmission. On the cloud side, latent feature will be extracted from the reconstructed image by the encoder

of pre-trained DL models and the task-related semantic information will be generated by the decoder of pre-trained DL models. This setup addresses the limited processing capacity of edge devices but highlights the requirements for data transmission and information extraction.

Although there is research on data transmission in the AEC sector, it mainly focuses on hardware and communication layer protocols (Tuhaise et al., 2023), without addressing how to handle transmitted data and improve transmission efficiency. On the other hand, for information extraction in CV applications, more dense DL models are increasingly used in AEC, yet they are often deployed on powerful GPUs, neglecting the interaction between edge devices and the cloud-side server. Additionally, inspired by the research of Huang and Wu, high-level features extracted by encoders can be reused across multiple downstream tasks (Huang and Wu, 2024). However, current AEC research mainly trains independent DL models for each task, leading to inefficient resource utilization and reduced real-time performance.

To address these limitations, this research focuses on two primary aspects: First, it aims to improve data transmission efficiency by enhancing the information density of transmitted image data within AEC domain. Specifically, Learned Image Compression (LIC) is introduced to prioritize task-relevant information rather than transmitting the entire image content. Second, a multi-task processing solution is proposed to integrate multiple downstream tasks within a unified framework, thereby maximizing computational reusability and improving efficiency.

1.2 Image Compression and Learned Image Compression

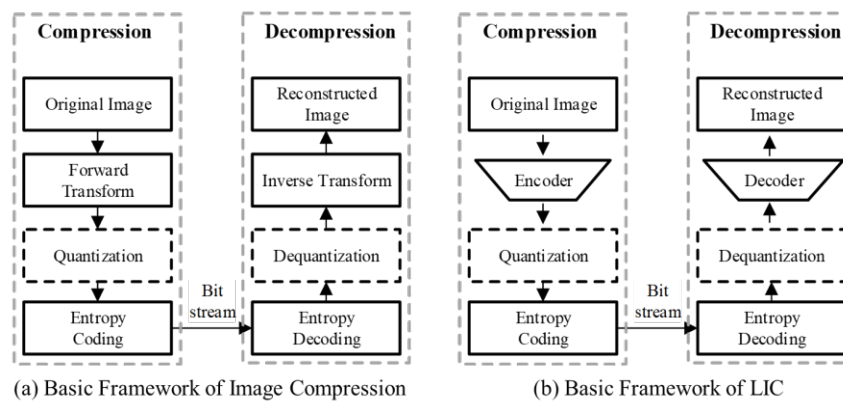


Figure 2. Basic Framework of Image Compression and Learned Image compression

The basic format of data transmission is bitstream. For computer vision tasks, digital images are encoded into bitstreams and reconstructed by codecs, a process known as image compression. As shown in Figure 2, image compression generally consists of two main phases: encoding and decoding. The encoding process involves forward transform, quantization, and encoding, while the corresponding decoding process includes inverse encoding, inverse quantization, and image reconstruction. Traditional image compression algorithms, such as JPEG (Wallace, 1992), use handcrafted operators to remove redundant information (e.g., large uniform color areas), effectively reducing storage and transmission load with minimal quality loss of image according to human perception.

Recently, with the rise of deep learning, LIC has emerged, transforming images into feature representations and removing redundancy through deep learning-based feature extractors, thereby supporting efficient downstream image reconstruction. Ballé et al. introduced the first end-to-end optimised model for image compression (Ballé et al., 2017). Later, Ballé et al. extended this work by incorporating a hyper-prior to better capture spatial dependencies in the latent representation (Ballé et al., 2018). Building on the success of autoregressive priors in probabilistic generative models, Minnen et al. further improved the entropy model by adding an auto-regressive component (Minnen et al., 2018). Z. Cheng et al. enhanced the network architecture using residual blocks and integrated a simplified attention module, replacing the commonly used Single Gaussian Model (SGM) with a Gaussian Mixture Model (GMM) (Z. Cheng et al., 2020). To reduce the need for serial processing in autoregressive context models, Minnen and Singh proposed a channel-wise auto-regressive entropy model (Minnen & Singh, 2020).

1.3 LIC-enhanced Machine Vision

With advancements in computer vision, images captured by cameras are increasingly processed by Artificial Intelligence (AI) algorithms rather than being solely consumed by humans. Consequently, LIC has evolved in two directions: human-perception-oriented LIC and machine-vision-oriented LIC. Human-perception-oriented LIC focuses on transmitting data for reconstructing high-fidelity digital images while machine-vision-

oriented LIC aims to convey task-relevant information. Figure 3 illustrates the high-level JPEG AI framework referenced in the JPEG AI white paper, featuring three distinct pipelines. In this learning-based image coding framework, digital images serve as input, and the output bitstream can be processed in two ways: it can either be reconstructed through a standard pipeline for human visualization or it can be optimised for machine-vision-oriented LIC. This dual functionality demonstrates the framework's capability to meet both human and machine processing requirements.

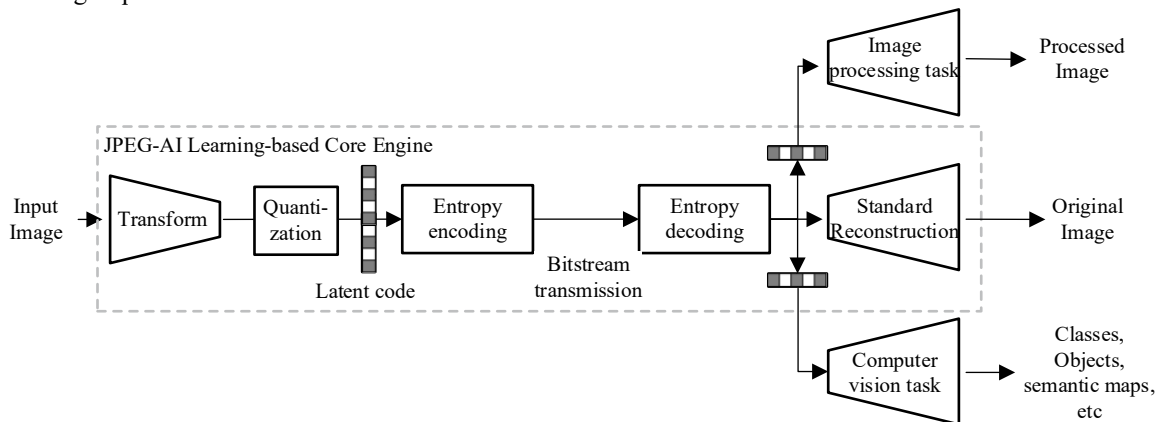


Figure 3. JPEG AI learning-based image coding framework

The concept of machine-vision-oriented LIC aligns well with the core principles of machine learning, where the objective is to extract essential features from raw data to create a latent representation for downstream decision-making or restoration tasks. This parallels the Minimum Description Length (MDL) principle in information theory, which seeks to find the most efficient model by minimizing the code length required to describe both the model and the data (Grünwald and Roos, 2019). Thus, machine-vision-oriented LIC not only enhances compression efficiency by focusing on task-specific information but also reflects the broader objective of machine learning to extract and compress the most relevant features, creating synergies between visual analysis and data compression.

The research of machine-vision-oriented LIC starts from Zhang et al.'s research, which they propose a research question of whether the bitstream of images and image features could be unified to serve both compression and retrieval simultaneously (Zhang et al., 2017). To address that, they proposed a content-based image retrieval system in which images are encoded once, and the encoded bitstream can be used for both image reconstruction and direct comparison for similar image retrieval. Several studies have focused on designing bitrate-efficient quantization for image compression while minimizing classification accuracy loss (Chamain et al., 2019; Liu et al., 2018). These joint rate–distortion–accuracy approaches aim to optimise quantization steps for JPEG and JPEG 2000 encoders to reduce both classification loss and bitrate.

With the success of LIC, interest in visual compression for machine vision has surged, focusing on maintaining machine task performance on compressed data. Le et al. introduced the first end-to-end learned system that optimises the rate-distortion trade-off, where the distortion term includes the training loss of a pre-trained neural network task (Le et al., 2021a). Le et al. introduced a content-adaptive fine-tuning method applied during inference, which aims to optimise the latent representation to enhance compression efficiency for machine-based tasks (Le et al., 2021b). Wang et al. developed an inverted bottleneck structure for the encoder and explored ways to refine the network architecture, specifically to improve compression performance for machine vision applications (Wang et al., 2021).

2 METHOD

This article presents a unified machine vision framework enhanced by LIC and applies it to the real-world scenario of construction site monitoring in the AEC domain. This chapter will introduce the proposed method from three aspects: overall structure, feature alignment, and training details.

2.1 Overall Structure

As shown in Figure 4, the proposed architecture comprises a single LIC model along with two downstream task models tailored for object detection and instance segmentation. Serving as the bottleneck of the overall framework, the LIC model is divided into an Encoder and a Decoder. The Encoder, deployed on an edge device, is primarily responsible for mapping raw images into a high-dimensional latent feature while effectively filtering out redundant information. It dynamically selects high-level features pertinent to the downstream tasks and encodes these features into a bitstream. In contrast, the Decoder operates in the cloud side alongside the downstream task models, focusing on reconstructing high-level features from the bitstream and subsequently regenerating the digital images. These reconstructed digital images are then input into the pre-trained models of

the downstream tasks, enabling the extraction of essential high-level semantic features. For object detection tasks, the extracted semantic feature includes class labels and bounding boxes; for instance segmentation tasks, it encompasses class labels and segmentation masks.

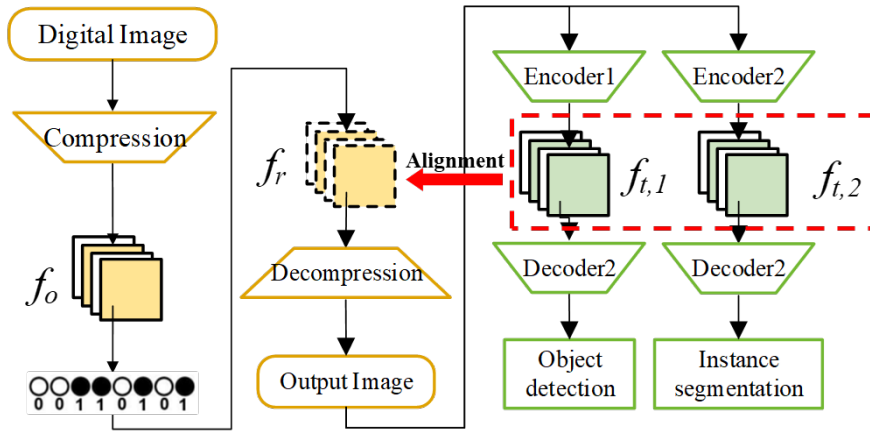


Figure 4. Overall Framework of LIC-Enhanced Multi-Task Machine Vision

2.2 Detail of LIC Model

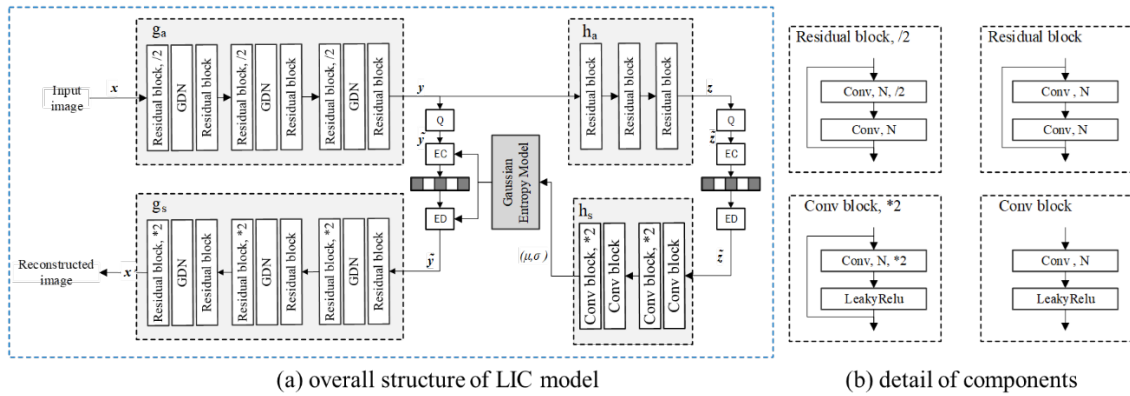


Figure 5. Detail of LIC Model

Figure 5 illustrates the fundamental architecture of the proposed LIC model, which primarily draws from the hyper-prior model presented by Ballé et al. (Ballé et al., 2018). This model is a dual-branch Variational AutoEncoder (VAE), designed to efficiently encode and compress image data while leveraging a hyperprior to enhance the modelling of latent representations. Q, EC and ED denote the quantization, entropy coding and entropy decoding, respectively.

The g_a - g_s branch is the main branch of the model, focusing on transforming the input image into its latent representation through encoding and then reconstructing it during decoding. The main branch is responsible for the actual compression and reconstruction of the image, ensuring that the salient details necessary for downstream tasks are preserved.

The h_a - h_s is the hyperprior of the model. It acts as an auxiliary model that captures the statistical dependencies among the latent representations generated by the encoder. By modelling these dependencies, the hyperprior enhances the efficiency of entropy coding, allowing for more compact representations of the image data. It essentially predicts the probability distribution of the latent variables, which is crucial for effective lossless compression. This hierarchical approach improves reconstruction quality by reducing the overall bitrate while maintaining or enhancing image fidelity.

The hyperprior and the main branch optimise the image compression process by enabling efficient representation and reconstruction of images. The hyperprior enhances the overall model's capacity to predict and encode data, while the main branch ensures that the essential features of the image are effectively captured and transmitted.

Notably, the model depicted in Figure 5 is deployed in a distributed manner, with the encoder implemented on the edge device and the decoder on the cloud side. The encoder comprises components such as Q, EC, the

Gaussian Entropy Model, along with the h_a and g_a modules, while the decoder includes Q, EC, the Gaussian Entropy Model, and the h_s and g_s modules, executing on the cloud side.

For downstream tasks, this study focuses on two classic CV applications: object detection and instance segmentation. For the object detection task, the Faster R-CNN model is selected, while the Mask R-CNN model is chosen for instance segmentation. Both models utilise the Feature Pyramid Network (FPN) with a ResNet-50 backbone.

Table 1 presents the parameter counts for the various models within this architecture. Notably, the LIC model exhibits a relatively smaller model size compared to the downstream task models. Furthermore, from the perspective of model deployment, the distribution of the computational load is facilitated by the separation of the LIC encoder and decoder across different devices, allowing for an efficient allocation of processing resources.

Table 1. Comparison of Model Sizes between LIC Encoder, LIC Decoder, and Downstream Task Models

Margin	LIC Encoder	LIC Decoder	Mask R-CNN	Faster R-CNN
Total params	6,604,337	9,237,565	43,982,622	41,356,561
FLOPs (G)	9.13	9.26	14.180	12.058

2.3 Latent Feature Alignment

In deep learning models, the carrier of information is the latent feature. The primary task of human-perception-oriented LIC is to ensure that the latent feature contains more information relevant to human perception, such as colour and texture. However, machine-vision-oriented LIC differs in that the latent feature it reconstructs should be more relevant to downstream tasks, often forming a subset of the information related to human perception. In light of this, this paper proposes a Latent Feature Alignment method, allowing the LIC model to selectively retain information pertinent to downstream tasks in an end-to-end manner, thereby improving compression efficiency without significantly impacting task accuracy.

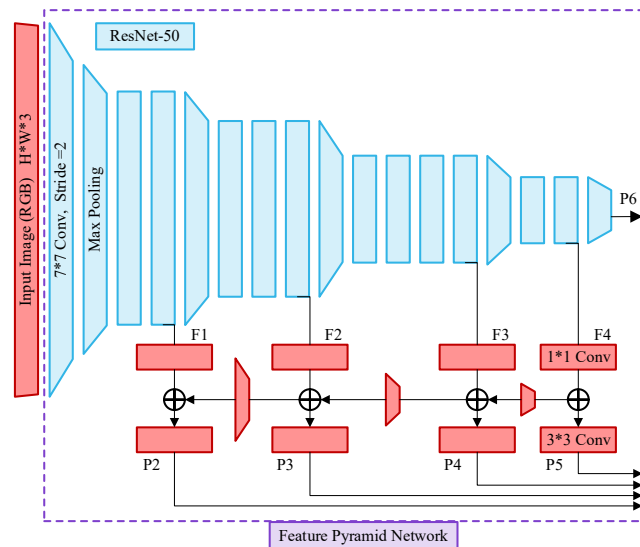


Figure 6. Feature Pyramid Network used for Faster R-CNN and Mask R-CNN

To adapt LIC model for generating a decoded image suitable for machine perception, three steps are employed.

- First, feature extraction is performed on the original image using the FPN of a pre-trained task model, selecting several latent features—P2, P3, P4, P5, and P6—as \hat{f} , as illustrated in the figure.
- Next, the LIC model is combined with the FPN, and the LIC model is trained while keeping the FPN frozen. In the same way, several latent features extracted by the FPN are denoted as f .
- Finally, the perceptual loss, defined as the difference between f and \hat{f} , is used as an additional term in the loss function, and the LIC model is trained through backpropagation.

Through this end-to-end training setup, the information transmitted by the LIC model is aligned with the information extracted by the FPN relevant to downstream tasks, allowing the LIC model to prioritize the transmission of task-relevant information instead of the complete set of information present in the image.

2.4 Training Detail

The training data utilised in this study is derived from the publicly available dataset, which is divided into

a training set (19,404 images), a validation set (4,000 images), and a testing set (18,264 images) (Xuehui et al., 2021). This dataset encompasses a total of 13 object classes, amounting to 116,380 labels. For the purposes of this research, a subset of the overall dataset was randomly sampled, with the distribution of the data illustrated in Table 2.

Notably, the original dataset exhibits a significant long-tail distribution issue, which can adversely affect the recognition accuracy of classes with fewer instances. Consequently, this study focuses exclusively on four object classes, each having a label count greater than 500, to ensure sufficient representation and improve recognition performance.

Table 2. Distribution of Instances across Different Types

Category	Worker	Excavator	Static crane	Truck
Train	12103	1431	1124	740
Val	586	116	75	55

The neural network constructed in this study has two primary optimization objectives: to reduce the bit rate (the amount of data transmitted) and to minimize the distortion (the loss of quality or accuracy). However, there generally exists a trade-off between these two objectives. To address this, the loss function was developed following the standard format of rate-distortion cost commonly used in the LIC domain, as illustrated in Equation (1).

$$L = R + \lambda D \quad (1)$$

where λ is a Lagrange multiplier that controls the trade-off between rate R and distortion D . By adjusting λ , one can prioritize minimizing either the rate or the distortion depending on the specific requirements of the application.

$$R = -\log p(\tilde{z}) - \log p(\tilde{y}|\tilde{z}) \quad (2)$$

R represents the number of bits required to represent the compressed data, which is typically expressed in bits per symbol or bits per pixel in image compression contexts. It is calculated based on information theory principles, particularly using concepts from Shannon's entropy.

$$D = d(f, \hat{f}) = \frac{1}{5} \sum_{i=2}^6 \text{MSE}(P_i(x), P_i(x')) \quad (3)$$

In LIC, D typically reflects the difference between the original and the reconstructed digital image. In this paper, the perceptual loss D measures the Mean Squared Error (MSE) between latent features extracted from the pretrained FPN as shown in Figure 6.

3 RESULTS

To evaluate the effectiveness of the proposed method, this paper discusses the results from both qualitative and quantitative perspectives.

3.1 Qualitative Result

Figure 7 contrasts the original image and the image reconstructed by the traditional LIC model, Faster-RCNN-enhanced LIC model and Mask-R-CNN-enhanced LIC model, while Figure 8 contrasts the task output based on the reconstructed image.

The traditional LIC model effectively reconstructs some details of the original image but also introduces some artifacts. However, these artifacts have little impact on the accuracy of downstream tasks.

On the contrary, the proposed machine-vision-oriented LIC model focuses specifically on regions relevant to downstream tasks during the image compression and reconstruction process, achieving high data reconstruction quality in those regions, which in turn results in certain accuracy for downstream tasks. For irrelevant regions, fewer bits are allocated, leading to lower image reconstruction quality.



Figure 7. Example of Reconstructed Images



Figure 8. Example of Task Output based on Reconstructed Images

Figure 9 illustrates the comparison of bit allocation across various regions in images reconstructed by the traditional LIC model and the Faster-R-CNN-enhanced LIC model. The results demonstrate a clear trend where the Faster-R-CNN-enhanced model concentrates a higher proportion of bits in regions crucial for downstream tasks, while allocating fewer bits to less relevant areas. This suggests that the proposed model effectively boosts the density of task-relevant information in the bitstream, thereby enhancing information efficiency.

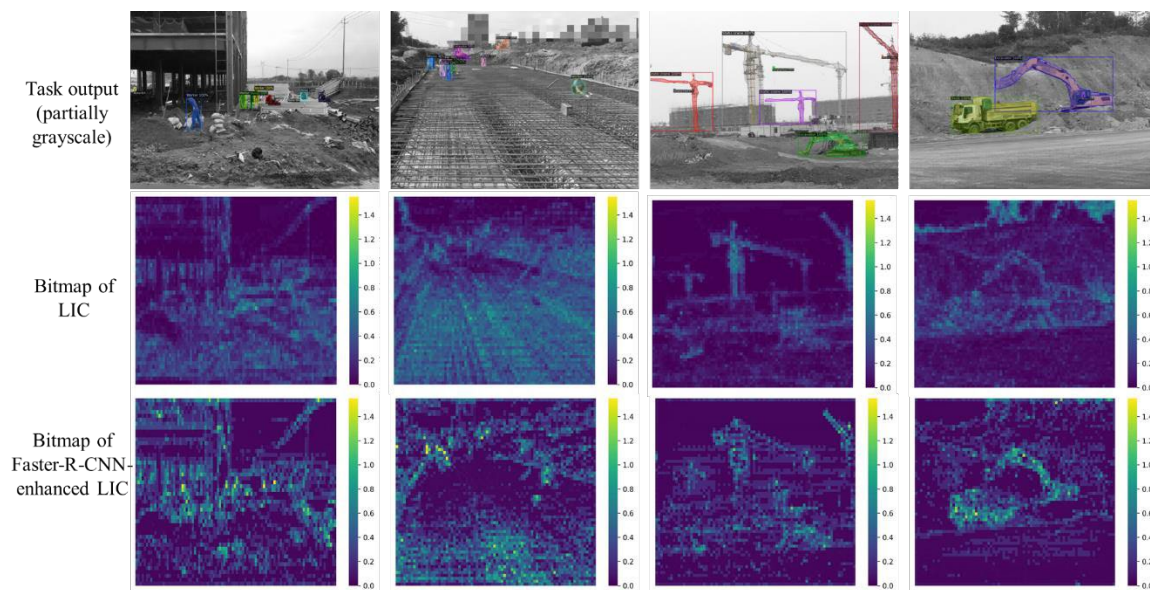


Figure 9. Example of Reconstructed Images and Corresponding Bitmaps during Image Compression

3.2 Quantitative Result

The quantitative assessment of the proposed approach primarily centers on the trade-off between compression efficiency and the accuracy of downstream tasks. The selected evaluation metrics include the following:

For image compression, the key performance indicator is bits per pixel (bpp), which indicates the average transmission load of the image. Higher bpp indicates lower compression ratio.

$$\text{bpp} = \frac{\text{File Size(in bits)}}{\text{Number of Pixels}} \quad (4)$$

For object detection, the primary metric for evaluation is:

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^N \int_0^1 P_i(R) dR = \frac{1}{N} \sum_{i=1}^N AP_i \quad (5)$$

where N is the number of object classes; $P_i(R)$ is the precision as a function of recall for class i ; $\int_0^1 P_i(R) dR$ represents the area under the precision-recall curve for class i , which gives the AP for that class.

For instance segmentation, the main metrics for evaluation are:

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{|T|} \sum_{t \in T} \int_0^1 P_i^t(R) dR \right) = \frac{1}{N} \sum_{i=1}^N AP_i \quad (6)$$

where N is the Number of object classes; T is set of IoU thresholds, typically ranging from 0.5 to 0.95 with a step of 0.05; $P_i^t(R) dR$ represents the precision at recall R for class i at threshold t .

Table 3. Comparison of bpp for LIC and mAPs for downstream task

Category	JPEG	LIC	Faster-R-CNN-enhanced LIC	Mask-R-CNN-enhanced LIC
bpp	3.0281	0.133	0.078	0.108
MSE	---	0.00116	0.00394	0.00291
mAP (object detection)	51.354	48.512	44.817	48.179
mAP (instance segmentation)	36.495	37.026	35.910	32.954

Table 4. Comparison of AP for different instance types

Task	Instance type	JPEG	LIC	Faster-R-CNN-enhanced LIC	Mask-R-CNN-enhanced LIC
object detection	Worker	52.157	50.995	47.071	45.481
	Excavator	65.875	65.724	61.961	58.231
	Static crane	48.995	46.78	43.608	37.046
	Truck	56.813	54.924	49.452	47.352
instance segmentation	Worker	44.528	43.481	39.398	36.509
	Excavator	41.557	41.207	39.773	37.643
	Static crane	26.148	26.079	22.283	20.803
	Truck	43.388	42.757	40.765	38.086

Table 3 presents a comparison of four methods—JPEG, LIC, Faster-R-CNN-enhanced LIC, and Mask-R-CNN-enhanced LIC—in terms of image compression efficiency and downstream task accuracy. The results indicate that the proposed method significantly improves data transmission efficiency compared to the original image, albeit with a minor reduction in task performance. When compared to traditional LIC models, the proposed approach substantially reduces the transmission load while maintaining similar levels of accuracy for downstream tasks, highlighting its effectiveness in increasing the density of task-relevant information in the bitstream. Table 4 further examines the detection accuracy of the four label types outlined in Table 2, reinforcing the findings and supporting the same conclusions.

When combined with the conclusions drawn from Figure 7, these results suggest that the proposed method reduces the overall data transmission load while increasing the proportion of relevant information (information density) in the transmitted data, with minimal impact on the accuracy of downstream tasks.

Moreover, the table compares the accuracy of semantic segmentation using Mask-R-CNN on images reconstructed by the Faster-R-CNN-enhanced LIC and the accuracy of object detection using Faster-R-CNN on

images reconstructed by the Mask-R-CNN-enhanced LIC. The accuracy difference between these tasks is relatively small, indicating that the latent features extracted by the FPNs in different task models exhibit a notable degree of generalizability across different downstream tasks. This finding suggests the potential to further improve the reusability of modules within the proposed multi-task processing framework. Future work could focus on exploring strategies to further enhance this reusability to optimise calculation efficiency across multiple tasks.

4. CONCLUSIONS

The AEC sector is experiencing a transformation towards automation and intelligence. With rapid advancements in computer vision-based intelligent sensing, issues such as efficient data transmission and large-scale data storage have become increasingly important. However, current research on data transmission tends to focus either on optimizing existing hardware or on developing frameworks, with limited attention given to improving the information density of transmitted data.

To address this gap, this paper introduces two perspectives: LIC and machine vision. LIC offers a research pathway for effective data compression using end-to-end methods, while the machine vision perspective integrates data transmission and multiple downstream computer vision tasks within a unified framework.

Building upon these perspectives, this paper proposes a machine-vision-oriented LIC model that unifies LIC and machine-vision models through latent feature alignment. The proposed approach was validated using a construction site monitoring dataset, and the results demonstrate that the model focuses transmission on task-relevant areas, allocating fewer resources to irrelevant image regions. Consequently, the method achieves a substantial improvement in data transmission efficiency without significantly compromising downstream task accuracy.

ACKNOWLEDGMENTS

This research is funded by the Cardiff University – China Scholarship Council joint programme.

REFERENCES

- Alizadehsalehi, S., Hadavi, A., Huang, J.C., 2020. From BIM to extended reality in AEC industry. *Automation in Construction* 116, 103254. <https://doi.org/10.1016/j.autcon.2020.103254>
- Ballé, J., Laparra, V., Simoncelli, E.P., 2017. End-to-end Optimized Image Compression.
- Ballé, J., Minnen, D., Singh, S., Hwang, S.J., Johnston, N., 2018. Variational image compression with a scale hyperprior.
- Chamain, L.D., Cheung, S.S., Ding, Z., 2019. Quannet: Joint Image Compression and Classification Over Channels with Limited Bandwidth, in: 2019 IEEE International Conference on Multimedia and Expo (ICME). Presented at the 2019 IEEE International Conference on Multimedia and Expo (ICME), IEEE, Shanghai, China, pp. 338–343. <https://doi.org/10.1109/ICME.2019.00066>
- Cheng, J.C.P., Chen, K., Chen, W., 2020. State-of-the-Art Review on Mixed Reality Applications in the AECO Industry. *J. Constr. Eng. Manage.* 146, 03119009. [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0001749](https://doi.org/10.1061/(ASCE)CO.1943-7862.0001749)
- Cheng, Z., Sun, H., Takeuchi, M., Katto, J., 2020. Learned Image Compression with Discretized Gaussian Mixture Likelihoods and Attention Modules.
- Grünwald, P., Roos, T., 2019. Minimum Description Length Revisited. *Int. J. Math. Ind.* 11, 1930001. <https://doi.org/10.1142/S2661335219300018>
- Huang, C.-H., Wu, J.-L., 2024. Unveiling the Future of Human and Machine Coding: A Survey of End-to-End Learned Image Compression. *Entropy* 26, 357. <https://doi.org/10.3390/e26050357>
- Le, N., Zhang, H., Cricri, F., Ghaznavi-Youvalari, R., Rahtu, E., 2021a. Image coding for machines: an end-to-end learned approach, in: ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 1590–1594. <https://doi.org/10.1109/ICASSP39728.2021.9414465>
- Le, N., Zhang, H., Cricri, F., Ghaznavi-Youvalari, R., Tavakoli, H.R., Rahtu, E., 2021b. Learned Image Coding for Machines: A Content-Adaptive Approach, in: 2021 IEEE International Conference on Multimedia and Expo (ICME). Presented at the 2021 IEEE International Conference on Multimedia and Expo (ICME), IEEE, Shenzhen, China, pp. 1–6. <https://doi.org/10.1109/ICME51207.2021.9428224>
- Liu, Z., Liu, T., Wen, W., Jiang, L., Xu, J., Wang, Y., Quan, G., 2018. DeepN-JPEG: A Deep Neural Network Favorable JPEG-based Image Compression Framework, in: 2018 55th ACM/ESDA/IEEE Design Automation Conference (DAC). Presented at the 2018 55th ACM/ESDA/IEEE Design Automation Conference (DAC), IEEE, San Francisco, CA, pp. 1–6. <https://doi.org/10.1109/DAC.2018.8465809>
- Minnen, D., Ballé, J., Toderici, G., 2018. Joint Autoregressive and Hierarchical Priors for Learned Image Compression.
- Minnen, D., Singh, S., 2020. Channel-wise Autoregressive Entropy Models for Learned Image Compression.
- Tuhaise, V.V., Tah, J.H.M., Abanda, F.H., 2023. Technologies for digital twin applications in construction.

- Automation in Construction 152, 104931. <https://doi.org/10.1016/j.autcon.2023.104931>
- Wallace, G.K., 1992. The JPEG still picture compression standard. IEEE Transactions on Consumer Electronics 38, xviii–xxxiv. <https://doi.org/10.1109/30.125072>
- Wang, Shurun, Wang, Z., Wang, Shiqi, Ye, Y., 2021. End-to-end Compression Towards Machine Vision: Network Architecture Design and Optimization.
- Xu, S., Wang, J., Shou, W., Ngo, T., Sadick, A.-M., Wang, X., 2021. Computer Vision Techniques in Construction: A Critical Review. Arch Computat Methods Eng 28, 3383–3397. <https://doi.org/10.1007/s11831-020-09504-3>
- Zhang, M., Xu, R., Wu, H., Pan, J., Luo, X., 2023. Human–robot collaboration for on-site construction. Automation in Construction 150, 104812. <https://doi.org/10.1016/j.autcon.2023.104812>
- Zhang, Q., Liu, D., Li, H., 2017. Deep network-based image coding for simultaneous compression and retrieval, in: 2017 IEEE International Conference on Image Processing (ICIP). Presented at the 2017 IEEE International Conference on Image Processing (ICIP), pp. 405–409. <https://doi.org/10.1109/ICIP.2017.8296312>

TEACHING ROBOT END EFFECTORS TO GRASP CONSTRUCTION TOOLS BASED ON DEEP REINFORCEMENT LEARNING

Xiaohu Yu¹, Yantao Yu²

- 1) Research Assistant, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: xiaohutruuth@ust.hk
- 2) Ph.D., Assis. Prof., Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: ceyantao@ust.hk

Abstract: There are numerous delicate tasks that require skilled workers to perform in onsite construction. A key challenge in automating these tasks is developing motor skills in robots trained through reinforcement learning (RL), as manipulating irregular and delicate objects like hammers, scaffolding, and drills remains difficult. To address this issue, this paper proposes an RL-based approach to perform delicate tasks using a robotic arm with grippers. We present a simulation-based policy learning framework utilizing the Critic-Actor algorithm in Pybullet to control the robotic arm. In experimental trials, the learned policy was used to grasp six different types of construction tools, and the results demonstrated the feasibility of training with randomly shaped objects to manipulate the construction tools with a reasonable success rate. This method provides a foundation for enhancing the manipulative skills of construction robots, potentially reducing labor costs in the industry.

Keywords: Construction robot; Reinforcement learning; Construction tools; Grasping

1. INTRODUCTION

The construction industry needs adequate workers to meet the construction demand[1]. However, construction tasks purely based on manual labor are time-consuming and inefficient[2]. Meanwhile, onsite construction is dangerous for workers, and the monotonous physical activities affect their health status[3]. Therefore, improving the automation level in construction is critical to this industry in the future. To this end, construction robots show huge potential to tackle this urgent issue across a wide range of fields, from off-site prefabrication[4] to on-site construction[5]. Although construction robots can solve these problems, the technology is not widely used in civil engineering and on-site construction due to the complexity of construction scenarios. Currently, most robots utilized in onsite construction are preprogrammed[6] or motion-planned[7], which are only suitable under a limited number of working conditions[8]. These robots perform poorly in unseen construction environments because it is impossible to plan all different possibilities in one program.

Recent review papers[9] discussing and envisioning robotics in construction synchronously share the perspective that the emergence of reinforcement learning (RL) provides a promising avenue. With the prospect of machine learning, RL is promising for developing construction robots with adaptability and flexibility, which can handle the problems mentioned above. RL can learn a specific policy (from observations to actions) for conducting different tasks by interacting with given construction environments through trial and error[10].

Next, we explain how RL components operate within a specific context—training a robotic arm with an RGB-D camera to perform object pickup tasks. The robot arm is identified as the agent, which interacts with its surroundings. The environment encompasses the entire experimental setup, including anything that the robot interacts with during its task. Observations consist of visual data collected by the RGB-D camera, providing the robot arm with the necessary information to make decisions about its actions. Actions refer to the commands sent to the robot arm to control its joints and grippers, dictating how the arm moves and interacts with objects in its environment. The reward function is crucial in RL as it provides the incentive for the agent to learn[11]. It is used to update the control policy based on the actions taken. Crafting an effective reward function is particularly challenging in complex, long-horizon tasks, which involve multiple steps or subtasks. The control policy dictates the robot arm's movements based on the input from the visual data. A well-trained policy enables the robot to successfully complete tasks like picking up objects.

In construction, researchers have conducted many attempts to train a robot using RL in various fields, such as installation[12], assembly[13], navigation[14] and so on. A more comprehensive summary of RL-based robots in construction is provided in recent literature[15]. On the one hand, existing RL-based works have concentrated on heavy construction tasks involving equipment such as robotic arms, cranes, wheel loaders, and unmanned ground vehicles. There is limited attention in delicate work, such as drilling, bolt insertion, and nut tightening, which need to be carried out by either humans or light small-scale robots. On the other hand, RL has emerged as a model-free approach that enables end-to-end robot grasping by learning through continuous trial-and-error interactions with the environment. This allows for self-supervised learning in robotic grasping tasks. However, a major challenge in learning-based robot grasping is how to realize the generalization—whether a robot can apply the knowledge it has learned in self-supervised settings to new environments and unfamiliar objects. While RL has been widely applied in video games and simple simulated robots, its application to complex robotic grasping

remains limited. For example, successful grasping of new objects in prior research required extensive training, such as 580,000 grasp attempts over several weeks[16] or 1 million grasp attempts in other studies, making the process costly and difficult to replicate[17].

To address this challenge, this paper introduces a simulation-based policy learning framework utilizing the Critic-Actor algorithm. By training for just 12 hours with randomly shaped objects, this approach achieves a 50-60% grasp success rate on a test set on unseen construction tools. Moreover, the on-policy DRL algorithm Proximal Policy Optimization (PPO) is employed, which offers stable and efficient training by iteratively updating the actor network with a small number of samples over multiple training rounds. This method addresses the issues of step-size selection and variance in traditional policy gradient algorithms.

The key contributions of this paper are:

1. We propose a model-free robotic grasping approach based on an on-policy reinforcement learning (RL) algorithm that enables generalization within a shorter training time, without requiring prior knowledge of target objects or large amounts of experience data.
2. In the experiment, six types of construction tools were used as objects for robotic manipulation, including a hammer, pipe, drill, and others.
3. The experimental results demonstrate that our approach successfully trains the robotic hand to grasp construction tools, achieving success rates of approximately 50-60%.

This article is organized as follows: we proposed a simulation-based policy learning framework utilizing the Critic-Actor algorithm in Section.2. The experiment settings using robot arm to grasp construction tools are detailed in Section 3. Experiment results and conclusion are in Section 4 and 5.

2. METHOD

2.1 Reinforcement learning algorithms

The reinforcement learning (RL) is a novel computational approach. Different from other computational approaches, the agent (Construction robot with gripper) in RL learns by directly interacting with its environment, without the need of supervision or prior knowledge of the environment. The RL process is the formal framework of a Markov decision process (MDP). which can be defined by a quadruplet (s_t, a_t, s_{t+1}, r_t) during the interaction between agents and the environment. During interaction at each episode, the agent based on its specific policy function $\pi_\theta(a|s)$ which generates an action a_t . The action next act on the environment in the state s_t , then the environment represents as a next state s_{t+1} . The interaction result can be assessed by the reward $r(a_t, s_t) = r_t$ (see Fig. 1). Normally, the agent considers a finite-horizon, which means agent explores finite steps T at each episode. The final goal of the agent is to understand the task and learn an optimal policy $\pi_\theta(a|s)$ to maximize the discounted cumulative returns G_t , which is expressed by:

$$G_t = \sum_{k=t}^T \gamma^{k-t} r_k(a_k, s_k) \quad (1)$$

Where $t \in [1, T]$, $\gamma \in [0, 1]$ is the discount factor.

In this paper, an actor-critic algorithm is introduced, which combines two different networks. The actor network is responsible for actions a_t through a specific policy function $\pi_\theta(a|s)$. The critic network generates the estimation of the value function $V(s)$ or the action value function $Q(s, a)$ to evaluate the actor network.

Firstly, the critic network will be updated by minimizing the time difference error (TD-error). The action value function $Q(s_t, a_t)$ estimated by the critic network can be given by:

$$Q(s_t, a_t) = E_{s_{t+1}, a_{t+1}, \dots} (G_t | s_t, a_t) \quad (2)$$

Which Q represents the long-term accumulated return given a specific action choice a_t at each state s_t .

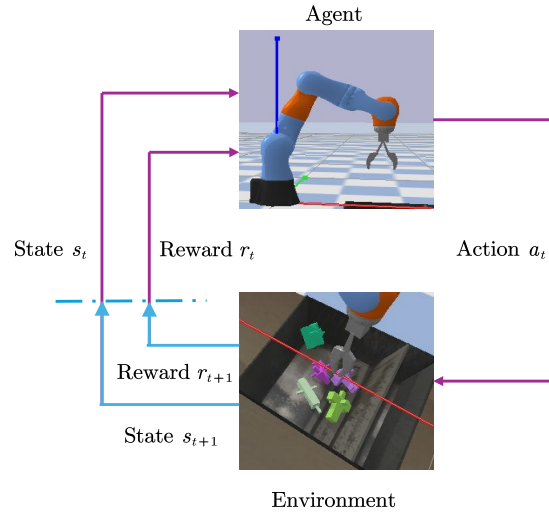


Fig. 1 Interaction between agent and environment in reinforcement learning
The state-value function $V(s_t)$ can be expressed by:

$$V(s_t) = E_{a_t, s_{t+1}, \dots}(G_t | s_t) \quad (3)$$

Unlike the action value function, the value function does not take the action into account and only considers the state.

Furthermore, the loss function for the critic network is defined by:

$$L(\omega_t) = r_t + \gamma \max_{a_{t+1}} Q(s_{t+1}, a_{t+1}; \omega_{t-1}) - Q(s_t, a_t; \omega_t) \quad (4)$$

Where $a_t \sim \pi_\theta(a_t | s_t)$, $s_{t+1} \sim P(s_{t+1} | (s_t, a_t))$. P is the transition probability distribution. The critic network updates parameters of ω_t to decrease $L(\omega_t)$ so that the real return by take the a_{t+1} approaches the expected return to take the action a_t .

Secondly, the actor network is trained by strategic gradient method, in which θ is updated according to the ascending direction of the gradient to maximize the critic output:

$$\theta \leftarrow \theta + \nabla_\theta \log \pi_\theta(a_t | s_t) V(s_t) \quad (5)$$

Where θ represents the parameters of the Actor network, $\pi_\theta(a_t | s_t)$ is the probability of taking action a_t in state s_t , and $V(s_t)$ is provided by the Critic network. The update rule indicates that the parameters of the Actor network are updated in the direction of increasing the output of the Critic network, meaning that the Actor is learning to take actions that the Critic network will lead to higher rewards.

2.2 Model structure

The actor-critic structure utilized in this paper is shown in Fig. 2. A common feature-network is shared by both actor and critic networks to extract the input s_t features, where s_t is an RGB image of size $(48, 48, 3)$. To be more specific, the actor network generates continuous actions represented by Cartesian displacements $(dx, dy, d\theta)$, where dx , dy represent the x -axis offset, y -axis offset, respectively. $d\theta$ denotes the orientation of the gripper, which rotates along the local coordinate system's z -axis.

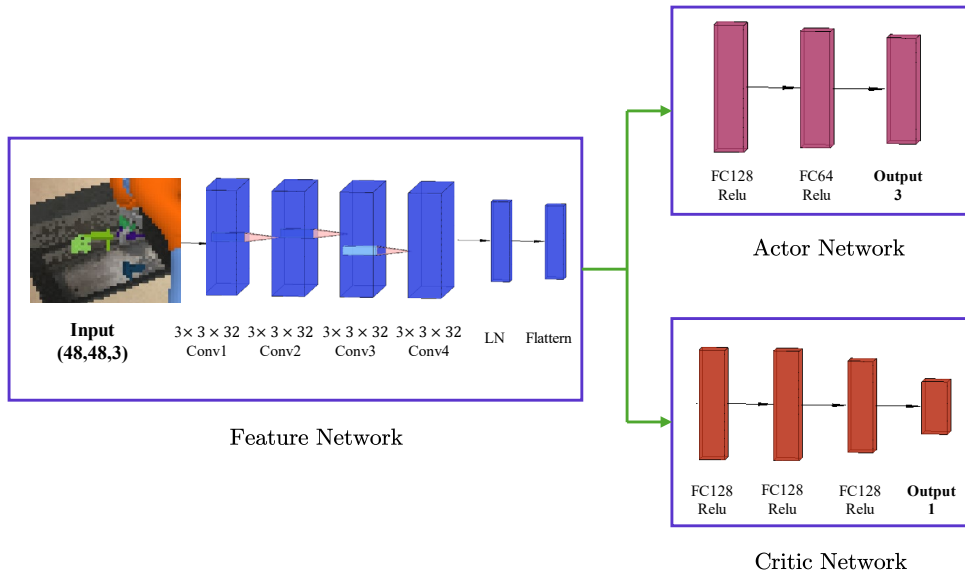


Fig. 2 Actor-critic architecture

2.3 Proximal Policy Optimization

PPO [18] is a widely used algorithm in the field of RL. It aims to optimize policy gradients in a more stable and efficient way than traditional policy gradient algorithms[19] by introducing a clipped objective function and a trust region. The core concept of PPO is to limit the amplitude of policy updates in each iteration using the probability ratio between the current policy and the old policy to avoid significant fluctuations in performance. This probability ratio is implemented through a clipping function, which is usually set to a small value, such as 0.1 or 0.2, to ensure the similarity between the new and old policies. The objective function of PPO can be expressed as:

$$L^{CLIP}(\theta) = E_t \left[\min \left(r_t(\theta) \hat{A}_t, \text{clip}(r_t(\theta), 1 - \epsilon, 1 + \epsilon) \hat{A}_t \right) \right] \quad (6)$$

$$r_t(\theta) = \frac{\pi_{\theta}(a_t | s_t)}{\pi_{\theta_{\text{old}}}(a_t | s_t)} \quad (7)$$

$$A_{\pi}(s_t, a_t) = Q(s_t, a_t) - V(s_t) \quad (8)$$

where $r_t(\theta)$ represents the ratio between the probability of action a_t given state s_t under the different policies (new and old one). \hat{A}_t denotes an estimate of the advantage, calculated using generalized advantage estimation, which can be calculated by Eq.(8). ϵ represents the hyper-parameter usually set as 0.2. Here, the expectation \hat{E}_t indicates the empirical average over a finite batch of samples, in an algorithm that alternates between sampling and optimization.

3. EXPERIMENTS

3.1 Experimental environment

The experimental environment was created using PyBullet[20], a real-time physics engine suitable for simulating robotic tasks. To be more specific, The Kuka-Diverse-Object-Env contains 8 different functions, which is designed for reinforcement learning experiments in object manipulation (see Fig. 3). It allows the robot to interact with randomly placed objects in both discrete and continuous action spaces. The environment includes customizable parameters for object placement, camera angles, and collision detection. Each episode terminates when the robot either attempts a grasp or exceeds the maximum step count.

There is a dataset including 900 various objects in Pybullet, which is used as training dataset (see Fig. 3). Each grasping system randomly places 5 different objects in a box for a 7-axis Kuka robot arm to grasp. 3D models of typical construction tools (e.g., hammer, drill, scaffolding) were loaded into the environment. In this study, we built a small dataset of construction tools with around 20. These models firstly were converted into URDF files, which allowed the definition of physical properties such as mass, friction, and dimensions. Furthermore, in order to characterize the common properties of the construction tools, they were divided into six shape-based categories (see Table 2). Then, we select one object in each type as the test object, as shown in Fig. 4.

Table 1 Different types of construction tools

Shape-based category	Construction tools dataset
Rod Shape	Hammer, Wrench, Crowbar
Cylindrical Shape	Flashlight, Paint Roller, Pipe
Flat Shape	Brush, Scraper, Scissors
Clamp Shape	Clamp, Pliers, Needlenose Pliers
Slender Shape	Scaffolding pin, Rebar, Screwdriver
Drill Shape	Hammer, Wrench, Crowba

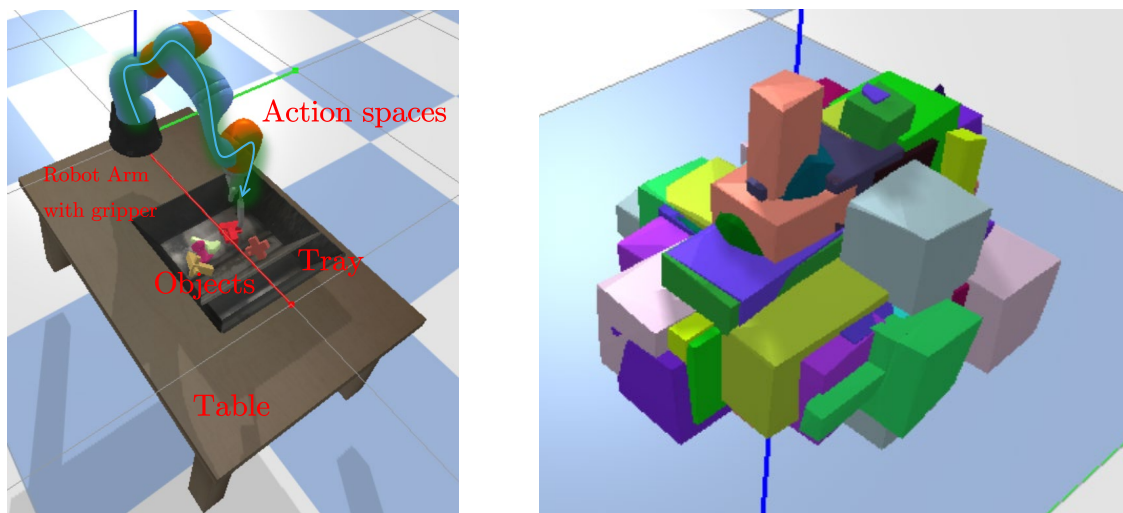
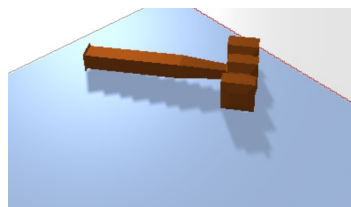
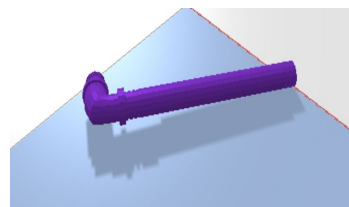


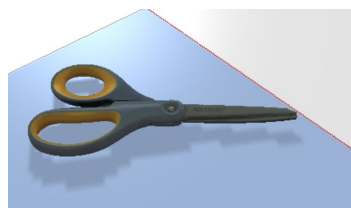
Fig. 3 Experimental environment (Left), Random object dataset (Right)



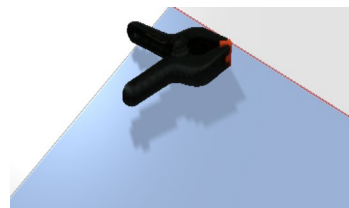
Hammer



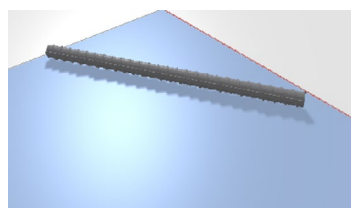
Pipe



Scissors



Clamp



Rebar



Drill

Fig. 4 Selected 6 types of test objects

3.2 Algorithm setting

The algorithm is based on Python 3.10 and Torch 2.4.1 and runs on Ubuntu 22.04 Linux system. Challenges with reward functions include sparse reward functions[11], which provide a reward only when a subtask is completed. While easy to define, they often result in limited guidance for policy updates, requiring a large amount of interaction data and leading to numerous failures. Continuous reward functions offer feedback for every action. This model adopts the sparse reward to decrease the exploring time. When grasping an object in one episode, the r_t is 1, otherwise it is 0. Table 1 shows the hyper-parameters used by PPO. The training took place in the PyBullet simulation, where a robotic arm was assigned the task of grasping various objects. Each season consisted of approximately 2,800 steps, corresponding to about 140 episodes, with each episode having a maximum of 20 steps. At the beginning of each episode, the robotic arm was randomly positioned, and the tools were randomly placed within the workspace. The agent learned to adjust its movements and grip based on feedback from the environment. In the resulting graph, the score represents the average of all episode scores within a season.

Table 2 Hyper-parameters used for PPO

Parameters	Value
Discount factor γ	0.993
Learning rate α	0.0004
Clip factor in PPO ϵ	0.07
Discount factor in GAE λ	0.95
Trajectory length N	2800
Training epochs	10
Batch size	128

3. RESULTS

The RL model was initially trained on a set of 900 randomly shaped objects, none of which were construction tools, over a short training period of approximately 12 hours. Following this, the model was tested in 100 trials to grasp unseen construction tools in order to evaluate its generalization capabilities. Here, if the construction tools can be lifted the position 0.2 m higher than the table, we assume the case is a successful grasping. The success ratios for each category of tools were assessed to determine the model's performance across these diverse types. The corresponding results for each category are shown in Figures 5–10.

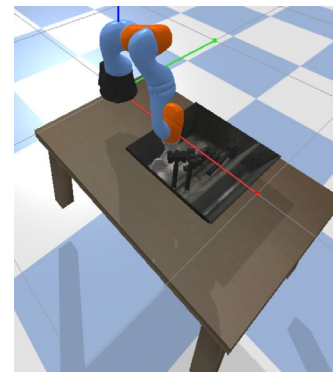
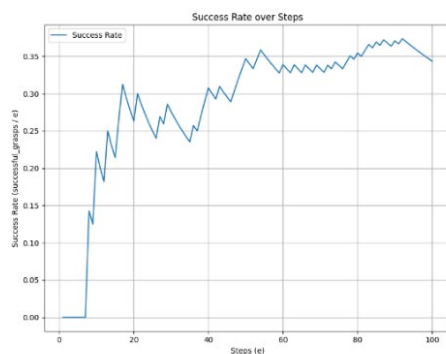


Fig. 5 The performance of RL model for hammer.

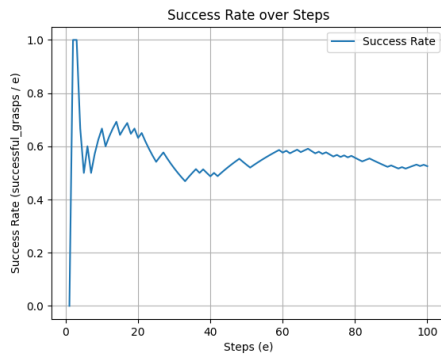


Fig. 6 The performance of RL model for pipe.

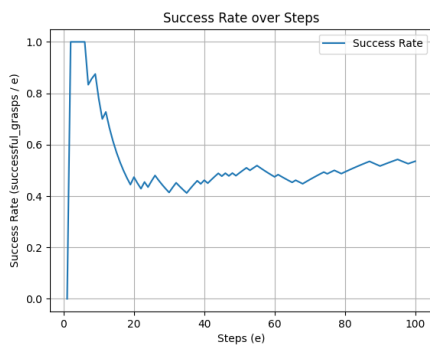
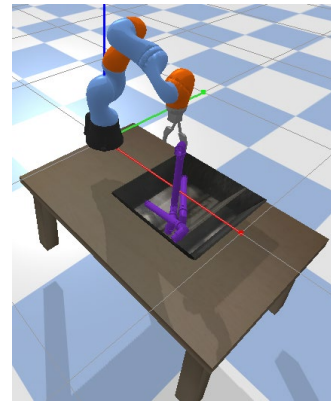


Fig. 7 The performance of RL model for scissors.

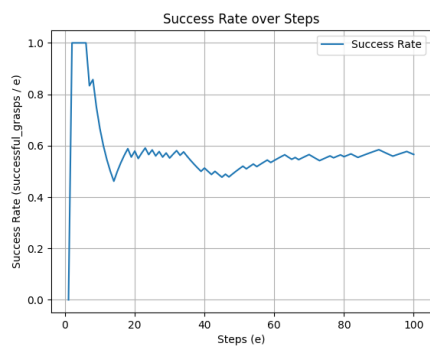
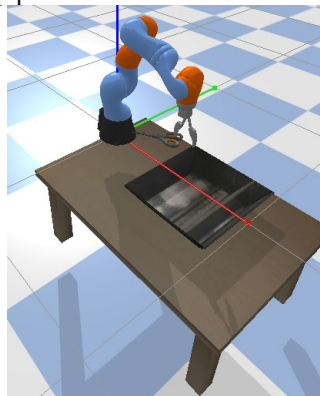


Fig. 8 The performance of RL model for clamp.

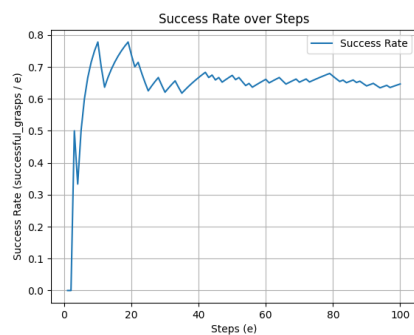
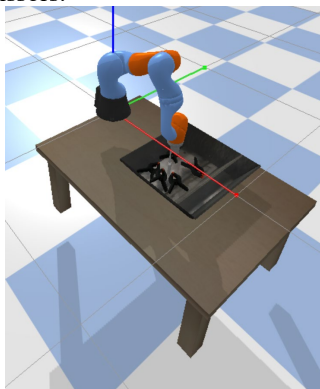
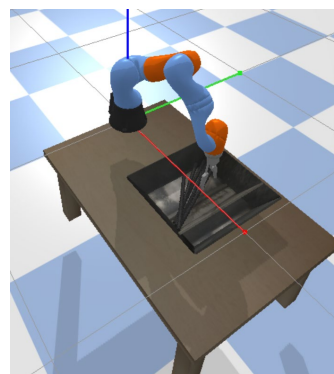


Fig. 9 The performance of RL model for rebar.



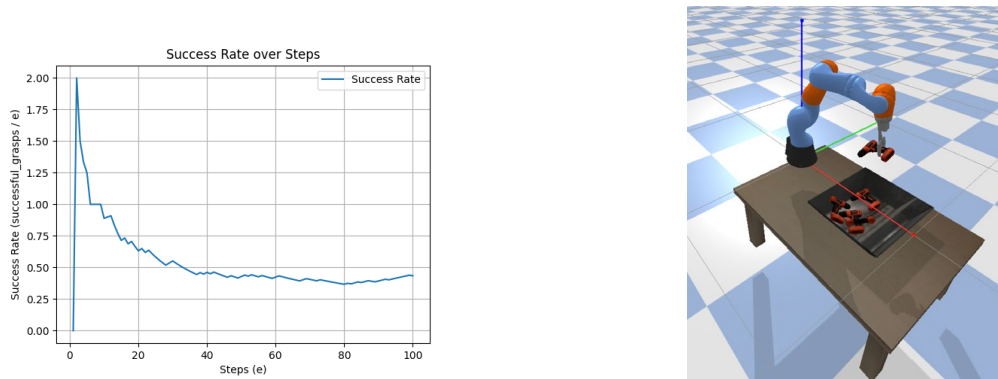


Fig. 10 The performance of RL model for drill.

Overall, as the number of test trials increased, the success ratios tended to converge towards specific values, which can be used to characterize the performance of the trained model. Notably, the model demonstrated the highest success rate for rebar grasping, achieving around a 65% success ratio, followed by pipe grasping as the second highest. The relatively high success rate for these two types of tools can likely be attributed to their cylindrical shape, which aligns well with the two-fingered gripper and lacks protruding features that might obstruct the grasping process. Although these construction tools are irregular, the gripper can exert sufficient frictional force to secure a grasp.

In contrast, the success rates for hammer and drill grasping were around 35%, lower than those for scissors (approximately 57%) and clamp (55%). The primary reason for the lower performance in grasping hammers and drills is the limited number of available grasping points for the gripper, as observed in failed trials. The RL model struggled to effectively learn how to grasp tools with fewer accessible grasping points. Nevertheless, these results provide valuable insights into the model's generalization capability and its potential to accurately grasp construction tools based on their shape characteristics.

4. CONCLUSIONS

The results from the RL model's testing across various construction tools highlight both its potential and the challenges involved in grasping irregular objects. While the model demonstrated a reasonable ability to generalize when grasping cylindrical objects like rebar and pipes, achieving success rates of 65% and slightly lower for pipes, it struggled with tools that have fewer or more complex grasping points, such as hammers and drills, which resulted in a lower success rate around 35%. Tools like scissors and clamps showed intermediate performance, with success rates of 57% and 55%, respectively. These findings suggest that the model's performance is closely linked to the geometric properties of the objects, with simple, cylindrical shapes being easier for the gripper to handle. The challenges in grasping tools with more complex shapes highlight the limitations of the current model in handling tools with fewer accessible grasping points or irregular designs.

From the perspective of RL, several promising avenues for improvement and future research can be considered. One potential enhancement involves integrating a self-attention mechanism into the RL framework by adding new layers to the model architecture. This would allow the model to better focus on relevant features during the learning process and improve its ability to handle complex grasping tasks. Additionally, by changing the input from a single camera to four camera images, the model can gain a more comprehensive view of the environment, reducing the limitations imposed by a restricted field of view and improving overall grasping performance. Another promising approach involves leveraging imitation learning to obtain a pretrained model, which can accelerate the training process. By using datasets generated from videos of construction tool grasping tasks and virtual reality (VR) environments, the model could learn from human demonstrations in tasks where real-world data is scarce or unavailable. This combination of self-attention mechanisms, multi-camera inputs, and imitation learning could significantly enhance the model's generalization ability and robustness, paving the way for more reliable and efficient robotic systems in real-world construction environments.

ACKNOWLEDGMENTS

This work was supported by the Collaborative Research Fund (C6044-23GF) supported by University Grants Committee of The Government of the Hong Kong Special Administrative Region, 30 for 30 Scheme (3030007) funded by The Hong Kong University of Science and Technology. The opinions presented in this paper are those of the authors and do not necessarily reflect the views of the sponsoring organizations. The authors would like to acknowledge helpful discussions with Zaolin Pan during the preparation of this manuscript.

REFERENCES

- [1] New Report Shows Construction Workforce Shortage Will Top Half a Million in 2023, Condustrial (2023). <https://condustrial.com/2023/03/02/new-report-shows-construction-workforce-shortage-will-top-half-a-million-in-2023/> (accessed October 5, 2024).
- [2] R.E. Chapman, D.T. Butry, A.L. Huang, Measuring and Improving U.S. Construction Productivity, (n.d.).
- [3] G. Biswas, A. Bhattacharya, R. Bhattacharya, Occupational health status of construction workers: A review, *Int. J. Med. Sci. Public Health* 6 (2017) 1. <https://doi.org/10.5455/ijmsph.2017.0745302112016>.
- [4] B. Ter Haar, J. Kruger, G. van Zijl, Off-site construction with 3D concrete printing, *Autom. Constr.* 152 (2023) 104906. <https://doi.org/10.1016/j.autcon.2023.104906>.
- [5] N. Melenbrink, J. Werfel, A. Menges, On-site autonomous construction robots: Towards unsupervised building, *Autom. Constr.* 119 (2020) 103312. <https://doi.org/10.1016/j.autcon.2020.103312>.
- [6] E. Gambao, C. Balaguer, F. Gebhart, Robot assembly system for computer-integrated construction, *Autom. Constr.* 9 (2000) 479–487. [https://doi.org/10.1016/S0926-5805\(00\)00059-5](https://doi.org/10.1016/S0926-5805(00)00059-5).
- [7] L.E. Bernold, Control schemes for tele-robotic pipe installation, *Autom. Constr.* 16 (2007) 518–524. <https://doi.org/10.1016/j.autcon.2006.09.002>.
- [8] B. Xiao, C. Chen, X. Yin, Recent advancements of robotics in construction, *Autom. Constr.* 144 (2022) 104591. <https://doi.org/10.1016/j.autcon.2022.104591>.
- [9] H. Wu, H. Li, X. Fang, X. Luo, A survey on teaching workplace skills to construction robots, *Expert Syst. Appl.* 205 (2022) 117658. <https://doi.org/10.1016/j.eswa.2022.117658>.
- [10] R.S. Sutton, A. Barto, Reinforcement learning: an introduction, Second edition, The MIT Press, Cambridge, Massachusetts London, England, 2020.
- [11] Grasping in the Wild: Learning 6DoF Closed-Loop Grasping From Low-Cost Demonstrations | IEEE Journals & Magazine | IEEE Xplore, (n.d.). <https://ieeexplore.ieee.org/document/9126187> (accessed October 5, 2024).
- [12] L. Huang, Z. Zhu, Z. Zou, To imitate or not to imitate: Boosting reinforcement learning-based construction robotic control for long-horizon tasks using virtual demonstrations, *Autom. Constr.* 146 (2023) 104691. <https://doi.org/10.1016/j.autcon.2022.104691>.
- [13] B. Belousov, B. Wibranek, J. Schneider, T. Schneider, G. Chalvatzaki, J. Peters, O. Tessmann, Robotic architectural assembly with tactile skills: Simulation and optimization, *Autom. Constr.* 133 (2022) 104006. <https://doi.org/10.1016/j.autcon.2021.104006>.
- [14] J. Cai, A. Du, X. Liang, S. Li, Prediction-Based Path Planning for Safe and Efficient Human–Robot Collaboration in Construction via Deep Reinforcement Learning, *J. Comput. Civ. Eng.* 37 (2023) 04022046. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0001056](https://doi.org/10.1061/(ASCE)CP.1943-5487.0001056).
- [15] Reinforcement Learning in Construction Engineering and Management: A Review | Journal of Construction Engineering and Management | Vol 148, No 11, (n.d.). <https://ascelibrary.org/doi/10.1061/%28ASCE%29CO.1943-7862.0002386> (accessed October 5, 2024).
- [16] D. Kalashnikov, A. Irpan, P. Pastor, J. Ibarz, A. Herzog, E. Jang, D. Quillen, E. Holly, M. Kalakrishnan, V. Vanhoucke, S. Levine, QT-Opt: Scalable Deep Reinforcement Learning for Vision-Based Robotic Manipulation, (2018). <https://doi.org/10.48550/arXiv.1806.10293>.
- [17] D. Quillen, E. Jang, O. Nachum, C. Finn, J. Ibarz, S. Levine, Deep Reinforcement Learning for Vision-Based Robotic Grasping: A Simulated Comparative Evaluation of Off-Policy Methods, *arXiv.Org* (2018). <https://arxiv.org/abs/1802.10264v2> (accessed September 30, 2024).
- [18] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, O. Klimov, Proximal Policy Optimization Algorithms, (n.d.).
- [19] J. Schulman, S. Levine, P. Moritz, M. Jordan, P. Abbeel, Trust Region Policy Optimization, (n.d.).
- [20] Bullet Real-Time Physics Simulation | Home of Bullet and PyBullet: physics simulation for games, visual effects, robotics and reinforcement learning., (2022). <https://pybullet.org/wordpress/> (accessed October 4, 2024).

MULTI-OBJECTIVE OPTIMIZATION OF TOWER CRANE LAYOUT PLANNING IN MODULAR INTEGRATED CONSTRUCTION CONSIDERING EFFICIENCY, COST, AND LIFTING SAFETY

Fangzheng LI¹, Hung-Lin CHI², Rongyan LI³, Roy Dong Wang⁴, Tarek Zayed⁵, and Bo Xiao⁶

1) Ph.D., Department of Building and Real Estate, The Hong Kong Polytechnic University, Hong Kong, SAR. Email: fangzheng.li@connect.polyu.hk

2) Ph.D., Assoc. Prof., Department of Building and Real Estate, The Hong Kong Polytechnic University, Hong Kong, SAR. Email: hung-lin.chi@polyu.edu.hk

3) Ph.D., Department of Building and Real Estate, The Hong Kong Polytechnic University, Hong Kong, SAR. Email: rongyan.li@connect.polyu.hk

4) Ph.D., Department of Building and Real Estate, The Hong Kong Polytechnic University, Hong Kong, SAR. Email: dongcq.wang@connect.polyu.hk

5) Ph.D., Prof., Department of Building and Real Estate, The Hong Kong Polytechnic University, Hong Kong, SAR. Email: tarek.zayed@polyu.edu.hk

6) Ph.D., Asst. Prof., Department of Civil, Environmental, and Geospatial Engineering, Michigan Technological University, USA. Email: boxiao@mtu.edu

Abstract: With the increase of modular integrated construction (MiC) projects, the planning of tower crane layout (TCLP) becomes vitally essential to achieve a balance among multiple goals, such as efficiency, economy, and safety. However, existing TCLP studies are usually formed as single-objective optimization based on total lifting time for conventional construction sites. There should be trade-offs among multiple goals, especially safety. In addition, the heavier components of MiC, requiring cranes with larger lifting capacity, pose a challenge in terms of cost. Therefore, it is necessary to propose a more general and reasonable model to assist managers in making better decisions for TCLP. This study aims to develop a multi-objective optimization model with efficiency, cost, and lifting safety considerations for MiC projects. Firstly, based on literature research and accident statistics, the total transportation time, total cost of tower cranes, and total lifting moment are chosen as the three optimization objectives. Then, the improved three-objective optimization model, considering module positioning time and separate movement features of tower cranes, is proposed. To solve the proposed multi-objective problem, the evolutionary algorithm NSGA-III is used for solutions. The proposed model can provide a series of trade-off solutions for efficiency, cost, and safety, representing different combinations of crane location, supply point location, and orientation. A MiC project in Hong Kong is studied as a case to verify the feasibility and effectiveness of the proposed model. The results show that the proposed model can determine the optimized layout plan with minimum time, cost, and lifting moment by locating the tower crane point, supply point, and supply point orientation. Disregarding the orientation of the supply point would result in an additional 18.2% transportation time, leading to increased costs. Compared to the original layout scheme, the developed model can save up to 41.7% in transportation time and improve safety by 27.4%.

Keywords: Multi-objective optimization, Tower crane layout planning, Lifting safety, MiC, NSGA-III.

1. INTRODUCTION

In Sep 2022, a fatal work accident occurred at a construction site in Hong Kong. A tower crane collapsed and struck a nearby office, resulting in the death of three employees and injuries to six employees (Labour Department Hong Kong 2022). The seriousness and danger of tower crane accidents are highlighted. Tower cranes, as one of the main vertical transportation equipment on construction sites, play a vital role in building construction, especially high-rise ones. However, it is estimated that tower crane-related accidents have occurred regularly in Hong Kong in the last 10 years, and more than 100 accidents occur globally each year. Construction crane safety has become one major concern of the Labour Department of HKSAR (Chen et al. 2022; Tam and Fung 2011). Accidents involving tower cranes not only threaten the lives of construction workers but also bring significant economic losses and schedule delays to construction projects. Therefore, strengthening the safety management of tower cranes and preventing and reducing accidents have become important remedies for the construction industry.

Modular integrated construction (MiC) has emerged as a cutting-edge and transformative technology in the construction industry in recent years (Darko et al. 2020). This innovative approach offers several distinct advantages over traditional construction methods. Firstly, MiC has the potential to reduce construction time by up to 30% (Tsz Wai et al. 2023). Secondly, the controlled factory environment and standardized manufacturing processes inherent in MiC ensure enhanced construction quality and precision. Thirdly, MiC greatly improves the working conditions and safety of construction sites. Furthermore, MiC aligns with the principles of sustainable development by promoting energy efficiency and reducing emissions.

The adoption of MiC in Hong Kong has been on the rise, with its application extending to various projects such as public housing, schools, and dormitories. Despite its growing popularity, MiC technology is still in its

nascent stages, presenting much room for further improvement and refinement. The cornerstone of on-site operations in MiC projects, particularly in high-rise buildings, lies in the installation of heavy-duty modules using tower cranes. Consequently, meticulous planning and optimization of tower cranes are paramount to ensure the smooth execution and success of MiC projects.

Tower crane layout planning (TCLP) is a key issue in construction site layout planning, aiming at determining the optimal tower crane layout scheme based on specific objectives and constraints in a complex construction environment. Optimal tower crane layout planning is of utmost importance for both conventional and modular construction projects. However, the inherent characteristics of MiC projects necessitate the development of a more general and reasonable model. The unique requirements of MiC, such as the need for precise module installation, the handling of heavy-duty components, and crane movement patterns, demand a tailored approach to tower crane planning. Specific challenges posed by modular construction should be taken into account, ensuring that the tower crane layouts are optimized to maximize efficiency, minimize costs, and guarantee the safe execution of the project.

Zhang et al. (1999) established the hook travel model for tower cranes, which laid the foundation for evaluating the travel time of tower cranes. Genetic algorithm and mixed-integer linear programming (MILP) were used to optimize the tower crane location in most later studies (Huang et al. 2011; Huang and Wong 2018; Tam et al. 2001). Zhang and Pan (2021) established the criteria considering high-rise MiC features and developed a two-step framework using a fuzzy-AHP-TOPSIS approach to optimize the tower crane layout. Unlike most studies that adopt the optimization-based approach, agent-based simulation (Younes and Marzouk 2018) was used to minimize operation time considering conflicts. BIM and VR-prototyping were used for planning interaction, avoiding clashes, and facilitating the decision-making process (Wang et al. 2015; Zhang and Pan 2021). The emergence of the GAN-based deep learning method using image-to-image translation greatly simplifies the decision-making process of TCLP (Li et al. 2023). However, existing TCLP studies are mostly about single-objective optimization based on total lifting time for conventional construction sites. Improving the existing models to meet the demand of MiC should be considered. The heavier components of the MiC require cranes with larger lifting capacity, safety, and cost should be put more emphasis on in addition to efficiency.

To fill the gaps in the current research and to address the realities in MiC, this paper develops a multi-objective optimization model with proper safety considerations for MiC projects and assists managers in making better decisions for TCLP. This study aims to provide the main contractors of MiC projects with a range of alternative tower crane layout plans. These plans offer multiple solutions that address the trade-offs among efficiency, cost, and safety. By presenting a variety of options, project managers can make informed decisions and select the optimal layout plan based on their specific project requirements and priorities.

2. METHOD

List of symbols

N	Total number of modules to be lifted
I	Total number of potential tower crane locations
J	Total number of potential supply points
K	Total number of demand points
n	ID of module $n \in N$
i	ID of tower crane locations $n \in N$
j	ID of potential supply points $j \in J$
k	ID of demand points $k \in K$
α	Coordination degree of trolleying and jib slewing
β	Coordination degree of hook movements in vertical and horizontal planes
γ	Safety factor to ensure a safety redundancy (e.g., 0.9)
w_n	Weight of n th module (t)
(X_{Ci}, Y_{Ci}, Z_{Ci})	Coordinate of i th tower crane location
(X_{Sj}, Y_{Sj}, Z_{Sj})	Coordinate of j th supply point
(X_{Dk}, Y_{Dk}, Z_{Dk})	Coordinate of k th demand point
V_h	Hoisting velocity of hook (m/min)
V_ω	Slewing velocity of jib (rad/min)
V_r	Radial velocity of trolley (m/min)
V_{ro}	Rotation velocity of hook for positioning a module (rad/min)

$T_{\omega(j,k)}^i$	Time for tangential movement of jib
$T_{r(j,k)}^i$	Time for radial movement of trolley
$T_{h(j,k)}^i$	Total horizontal movement time
$T_{v(j,k)}^i$	Total vertical movement time
$T_{ro(j,k)}^i$	Time for positioning a module
$T_{j,k,i}$	Total hook travel time

2.1 Problem Statement and Assumptions

The TCLP problem is formulated as a multi-objective problem with the aim to maximize efficiency, minimize costs, and maximize lifting safety. For information on a given construction site, the site is first simplified into a drawing with pre-determined demand, supply, and candidate tower crane locations. The data provided will be input into the optimization model developed. The model is then solved using a genetic algorithm.

In the model, the type of tower crane is given, with known crane configurations such as load charts and velocities. Available locations to set up the tower crane and areas for storing modules were pre-determined, and information on module requests, such as positions and weights, was provided. The model was proposed based on the following assumptions:

- (1) The module is provided in one supply location, and the orientation of the supply location represents that of the module.
- (2) The available areas for the module supply point and tower crane are pre-defined based on site conditions.
- (3) Only one tower crane is considered, which can be assigned to any of the available tower crane locations.

2.2 Proposed mathematical model

Time and cost are the single objectives selected for optimization in most of the current literature, so they are selected in this paper as the first two optimization objectives. The total cost of a tower crane is assumed to be the sum of the fixed and operation costs for that crane model. In addition, the lifting moment is adopted as an indicator of safety based on literature research, accident reports, and expert interviews. The establishment of each goal will be discussed specifically below:

- (1) Total transportation time and total cost

Transportation time has been consistently adopted as a primary optimization objective in crane layout optimization studies. This metric represents the total operation time required to complete all element transportation tasks from supply point(s) to demand point(s), indicating crane transport productivity. Zhang's model for calculating hook travel time has been extensively utilized. This model takes into consideration the tower crane's three degree-of-freedom motion pattern (radial, horizontal, and vertical), as shown in Figure 1.

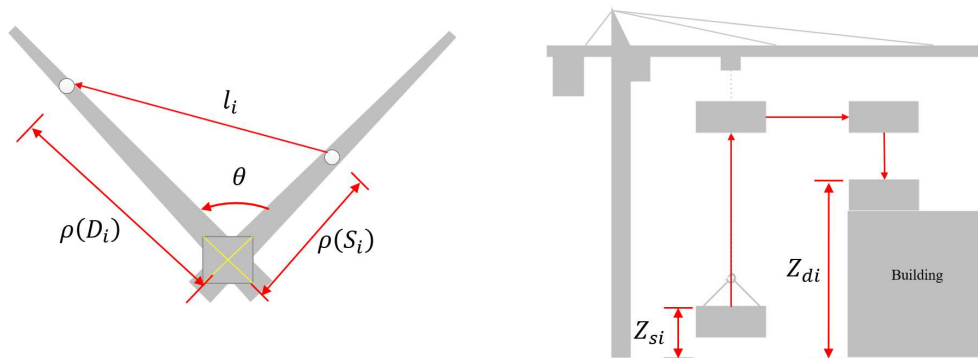


Figure 1. Mathematical model of tower crane radial, horizontal, and vertical movements.

For modular construction, since the modules need to be transported precisely to their final location and installed, the orientation of the modules needs to be rotationally aligned before the final lift-off, as shown in Figure 2. The final positioning angle of the module can be obtained from the orientation of the supply point, the angle of the module at the demand point, and the horizontal rotation angle of the tower crane, as shown in Eq. (6). TCLP should fully simulate the features of MiC, such as crane movement pattern (move the trolley first and then rotate the jib (or vice versa)) and rotation for the module positioning.

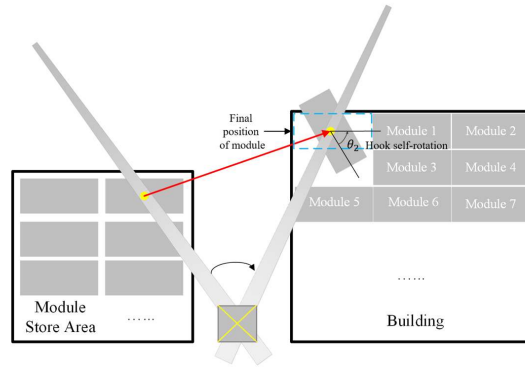


Figure 2. Tower crane module positioning for modular construction.

In this study, the time consumed for the module positioning module is appended in Zhang's model to more accurately characterize lifting in MiC. When the lifting of a module starts, the operators often move the trolley first and then rotate the jib (or vice versa) in MiC. As a result, the coordination degree α and β equal to 1.0 for modular element lifting. The traveling time for a module can be obtained using Eq. (7).

$$T_{r(j,k)}^i = \frac{|l_{i,j} - l_{i,k}|}{V_r} \quad (1)$$

$$T_{\omega(j,k)}^i = \frac{1}{V_{\omega}} \cdot \arccos \theta_{\omega(j,k)}^i \quad (2)$$

$$[0 \leq \arccos(\theta) \leq \pi]$$

$$T_{v(j,k)} = \frac{|Z_{Sj} - Z_{Dk}|}{V_v} \quad (3)$$

$$T_{h(j,k)}^i = \max(T_{r(j,k)}^i, T_{\omega(j,k)}^i) + \alpha \min(T_{r(j,k)}^i, T_{\omega(j,k)}^i) \quad (4)$$

$$T_{sd} = \max(T_{h(j,k)}^i, T_{v(j,k)}) + \beta \min(T_{h(j,k)}^i, T_{v(j,k)}) \quad (5)$$

$$T_{ro(j,k)}^i = \frac{1}{V_{hr}} \cdot |\theta_{Dk} - \theta_s - \theta_{\omega(j,k)}^i| \quad (6)$$

$$T_{j,k,i} = T_{sd} + T_{ro} \quad (7)$$

Thus, the objective function of minimizing total transportation time is shown in Eq. (8).

$$\min TT = \min \sum_j^J T_{j,k,i} \quad (8)$$

After obtaining the total transportation time, the cost can be calculated. For a given type of tower crane, the cost can be simplified into two parts: operation costs and fixed costs. The operation cost equals the total transportation time TT multiplied by unit time cost C . The fixed cost C_f incorporates the delivery, assembly, maintenance, and disassembly fee of the tower crane. Thus, the objective function of minimizing total cost is shown in Eq. (9).

$$\min TC = \min \sum_j^J T_{j,k,i} \cdot C + C_f \quad (9)$$

(2) Total lifting moment

The lifting moment of a crane is an essential safety consideration for enhanced operation safety (others include working conditions and crane stability). Project managers often choose the lifting moment and the lifting capacity to select the tower crane configurations. This ensures the safety of using the tower crane while controlling the rental costs. However, the lifting moment can either raise or lower the support reaction of the tower crane base, so it is inseparably linked to the safety of tower crane operation.

The high weight and size of the modular components impose higher requirements on the safety of the tower crane for module lifting. This is to prevent the tower crane from overturning and to ensure the safety of module lifting. The lifting moment equals the product of the module weight and the tower crane's working radius. It induces the force of rotation around the axis of a tower crane. Structural failure and collapse may occur if the

selected crane capacity does not satisfy the maximum lifting moment. Hebiba et al. (2022) used total lifting moments and total hook travel time as two optimization objectives to select an optimal crane location and supply locations, and the effectiveness of the proposed methodology is validated by a case study. As a result, the lifting moment LM is chosen to represent safety in this study. The lifting moment LM is calculated by Eq. (10), based on a single crane location associated with supply and demand locations. The lifting supply moments and lifting demand moments can be obtained from Eq. (11). Thus, the objective function of minimizing the total lifting moment is shown in Eq. (12).

$$LM_{j,k,i} = \sqrt{M_{n(i,j)}^2 + M_{n(i,k)}^2} \quad (10)$$

$$M_{n(i,k)} = w_n \cdot l_{i,k} = w_n \cdot \sqrt{(X_{Dk} - X_{Ci})^2 + (Y_{Dk} - Y_{Ci})^2} \quad (11)$$

$$M_{n(i,j)} = w_n \cdot l_{i,j} = w_n \cdot \sqrt{(X_{Sj} - X_{Ci})^2 + (Y_{Sj} - Y_{Ci})^2}$$

$$\min TLM = \min \sum_j^J LM_{j,k,i} \quad (12)$$

(3) Constraints

1. Given a tower crane location, the supply point and demand point of the module should be within the coverage of the tower crane.

$$R_{Cmin} \leq l_{i,j} = \sqrt{(X_{Sj} - X_{Ci})^2 + (Y_{Sj} - Y_{Ci})^2} \leq R_{Cmax} \quad (13)$$

$$R_{Cmin} \leq l_{i,k} = \sqrt{(X_{Dk} - X_{Ci})^2 + (Y_{Dk} - Y_{Ci})^2} \leq R_{Cmax} \quad (14)$$

2. For each module, the weight at the supply point and demand point should be within the loading capacity of the tower crane.

$$w_n \leq \gamma \cdot LC(\sqrt{(X_{Sj} - X_{Ci})^2 + (Y_{Sj} - Y_{Ci})^2}) \quad (15)$$

$$w_n \leq \gamma \cdot LC(\sqrt{(X_{Dk} - X_{Ci})^2 + (Y_{Dk} - Y_{Ci})^2})$$

$$LC(x) = \frac{a}{x} - b \quad (16)$$

3. For each module, the lifting moment at the supply point and demand point should be within the maximum lifting moment of the tower crane. The safety factor γ (0.9 here) is introduced to ensure a safety redundancy.

$$M_{n(i,j)} = w_n \cdot \sqrt{(X_{Sj} - X_{Ci})^2 + (Y_{Sj} - Y_{Ci})^2} \leq \gamma \cdot M_{max} \quad (17)$$

$$M_{n(i,k)} = w_n \cdot \sqrt{(X_{Dk} - X_{Ci})^2 + (Y_{Dk} - Y_{Ci})^2} \leq \gamma \cdot M_{max}$$

2.3 Multi-objective optimization evolutionary algorithms

Since the study in this paper can be regarded as a combinatorial optimization problem, evolutionary algorithms NSGA-III (Deb 2011; Deb and Jain 2014) are utilized to solve the three-objectives problem. The NSGA-III algorithm (as shown in Figure 3) is a relatively mature branch of the genetic algorithm and has great advantages in solving multi-objective combinatorial optimization problems. Compared with the exact solution method, NSGA-III can save computational cost. As the complexity of the project increases, the time taken for the exact solution will show a significant increase while the solution time of the evolutionary algorithm is still within acceptable limits. The basic operations of NSGA-III include Select, Crossover, Mutation, and Elitist Strategy, simulating the natural evolutionary process. NSGA-III is simple and efficient in achieving the Pareto front. With the introduction of a reference point mechanism, the selection process is guided to maintain the diversity of solutions when solving three or more optimization objective problems.

2.4 Case study using a MiC project

The selected project is a 28-storey subsidized sale house in Hong Kong, of which 25 floors are modular buildings, as shown in Figure 4. The house provides 300 room units in total using over 1200 concrete modules, ranging from 2.8 tons to 21.7 tons. Each floor was constructed using 49 modules around in-situ components. The modules required for the building are serviced by one tower crane. Detailed modular information is provided in Table 1. Therefore, an independent tower crane, POTAIN MCT805, is assigned, which has a maximum lifting weight of 25 t and coverage of 35 m. After a site investigation, the possible supply area and tower crane area were determined, and 11 potential tower crane locations and 34 supply points were selected, as shown in Figure 4. The

locations of the demand points are the 49 locations where modules need to be installed. The goal is to determine the optimal supply point location, supply point orientation, and tower crane location for trade-offs among efficiency, cost, and safety.

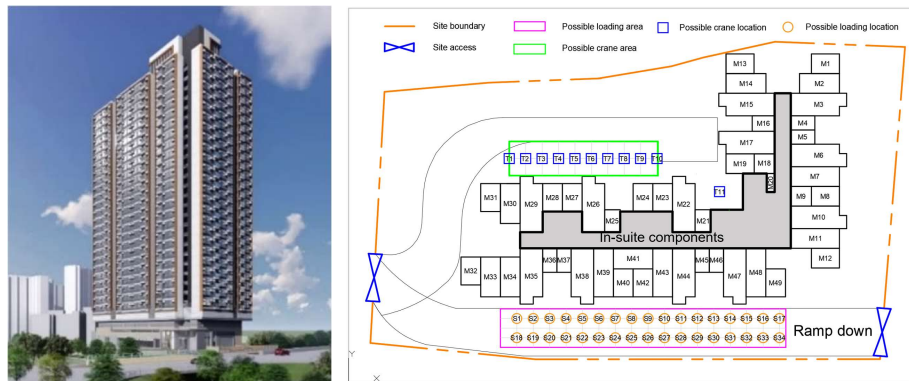


Figure 4. The exterior view of the construction site and site layout scheme

Table 1. Module information

Module IDs	Length (mm)	Width (mm)	Height (mm)	Weight (t)	Module position		
					x	y	Degree (rad)
M1	3410	2400	3000	9.8	58071	38787	0
M2	4860	2400	3000	15.57	57346	36387	0
M3	6800	2750	3000	21.7	57256	33812	0
...
M48	5920	2400	3000	18.33	49607	13317	1.57
M49	3410	2400	3000	9.8	52007	12062	1.57

3. RESULTS AND DISCUSSIONS

After the multi-objective optimization process, three non-dominated solutions to the Pareto frontier exist. This shows that the three solutions strictly outperform all other solutions in all three objective functions. In order to demonstrate the diversity of the Pareto frontiers and the selectivity of the results, we added the solution set of the second frontier as well, and the final results are shown in Figure 5. A total of 20 solution points were obtained. The three solutions to the Pareto frontier (P1, P2, and P3) and P4, P5, P6 from the other frontier are shown in Table 1 for comparison. Also, the original project's layout is listed. It is noted that a set of solutions are on the Pareto frontier that seeks an optimal balance between multiple goals. Afterwards, multiple criteria decision analysis (MCDA) methods such as weighted sum model (WSM), technique for order preference by similarity to ideal solution (TOPSIS), and analytic hierarchy process (AHP) can be used for weighting the three objectives based on specific project requirements and designer priorities.

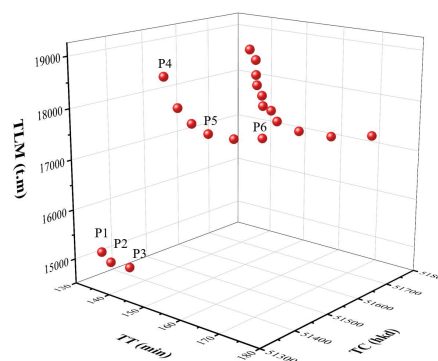


Figure 5. The obtained Pareto fronts

The original layout scheme is shown in Figure 6, where T8 and S3 are selected as the tower crane location

and supply location, and the default setting of supply point orientation is 0 degrees. All three solutions to the Pareto frontier choose T11 for locating the tower crane. The optimal layout calculated from our model is shown in Figures 7 and 8. Figure 7 shows the layout with minimum time and cost, where T11 and S11 are selected as the tower crane location and supply location, and the supply point orientation is 40 degrees. As for the layout with minimum lifting moment, T11, S13, and 40 degrees are determined.

Table 1. Solutions from Pareto fronts and project original scheme

Points	Tower crane location	Supply point	Supply point orientation	TT (min)	TC (HKD)	TLM (t.m)
P1	T11	S11	40°	148.4	51,484	15,130
P2	T11	S12	30°	151.2	51,512	14,899
P3	T11	S13	40°	155.4	51,554	14,792
P4	T9	S4	50°	143.0	51,433	18,575
P5	T10	S10	60°	157.2	51,572	17,329
P6	T10	S8	60°	162.0	51,628	17,443
The original scheme	T8	S6	0°	210.3	52,103	18,854


 Figure 6. The original site layout scheme ($\theta_S = 0^\circ$).

 Figure 7. Layout with minimum time & cost ($\theta_S = 40^\circ$).



Figure 8. Layout with minimum lifting moment ($\theta_s = 40^\circ$).

The time consumed by the positioning of the module is considered in the model, and the orientation of the module is thus optimized to minimize the time. Under the case P1, the total transportation time is 148.4 mins when the orientation is 40° . However, the total transportation time would be 181 mins if the orientation of the module is not considered, an 18.2% increase over the former. This means that it is essential to properly plan the orientation of the module to minimize total operation time and cost. It is worth noting that the time taken to position a module depends on the velocity of rotation. Faster rotation velocity will reduce the time and increase the risk of operations such as dropping modules (the rotation velocity is 0.5 rad/min in this paper). In terms of lifting safety, the total lifting moment of the original arrangement is 18,854 t.m, while the minimum total lifting moment of the P3 case is 14,792 t.m, a reduction of 27.4%, and the operating safety of the tower crane is further improved. Compared to the original layout scheme, the developed model can save up to 41.7% in transportation time and improve safety by 27.4%.

4. CONCLUSIONS

This study proposes a three-objective optimization model considering efficiency, cost, and lifting safety, which is applicable to MiC and traditional site conditions. The features and crane movement pattern of MiC are considered in the efficiency objective function, including the module rotation positioning and separate operation of tower crane motion. Besides, the lifting moment is selected as one of the objectives, representing the safety of the tower crane operation. The genetic algorithm NSGA-III is adopted to solve the proposed multi-objective problem. A case study using a MiC project in Hong Kong is conducted to verify the proposed model. Planners can choose from the solutions in the Pareto frontier based on the project needs and preferences for better decision-making. Based on the case study results, our model can determine the optimized layout plan with minimum time, cost, and lifting moment by locating the tower crane point, supply point, and supply point orientation. Disregarding the orientation of the supply point would result in an additional 18.2% transportation time, leading to increased costs. Compared to the original layout scheme, the developed model can save up to 41.7% in transportation time and improve safety by 27.4%.

ACKNOWLEDGMENTS

The author would like to thank Yau Lee Construction Co., Ltd. for providing MiC project data.

REFERENCES

- Chen, J., H.-L. Chi, Q. Du, and P. Wu. 2022. "Investigation of Operational Concerns of Construction Crane Operators: An Approach Integrating Factor Clustering and Prioritization." *J. Manage. Eng.*, 38 (4): 04022020. [https://doi.org/10.1061/\(ASCE\)ME.1943-5479.0001044](https://doi.org/10.1061/(ASCE)ME.1943-5479.0001044).
- Darko, A., A. P. C. Chan, Y. Yang, and M. O. Tetteh. 2020. "Building information modeling (BIM)-based modular integrated construction risk management – Critical survey and future needs." *Computers in Industry*, 123: 103327. <https://doi.org/10.1016/j.compind.2020.103327>.
- Deb, K. 2011. "Multi-objective Optimisation Using Evolutionary Algorithms: An Introduction." *Multi-objective Evolutionary Optimisation for Product Design and Manufacturing*, L. Wang, A. H. C. Ng, and K. Deb, eds., 3–34. London: Springer London.
- Deb, K., and H. Jain. 2014. "An Evolutionary Many-Objective Optimization Algorithm Using Reference-Point-Based Nondominated Sorting Approach, Part I: Solving Problems With Box Constraints." *IEEE Trans. Evol. Computat.*, 18 (4): 577–601. <https://doi.org/10.1109/TEVC.2013.2281535>.
- Hebiba, A. M., A. Bouferguene, S. Moon, and S. Han. 2022. "Wind-Wise Automated Stability Analysis for

- Selection of Tower Crane and Location." *J. Constr. Eng. Manage.*, 148 (11): 04022127. [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0002377](https://doi.org/10.1061/(ASCE)CO.1943-7862.0002377).
- Huang, C., and C. K. Wong. 2018. "Optimization of crane setup location and servicing schedule for urgent material requests with non-homogeneous and non-fixed material supply." *Automation in Construction*, 89: 183–198. <https://doi.org/10.1016/j.autcon.2018.01.015>.
- Huang, C., C. K. Wong, and C. M. Tam. 2011. "Optimization of tower crane and material supply locations in a high-rise building site by mixed-integer linear programming." *Automation in Construction*, 20 (5): 571–580. <https://doi.org/10.1016/j.autcon.2010.11.023>.
- Labour Department Hong Kong. (2022). Retrieved from Labour Department Hong Kong website: https://www.labour.gov.hk/common/osh/pdf/wsa/WSA-Collapse_of_a_Tower_Crane_ENG.pdf
- Li, R., H.-L. Chi, Z. Peng, X. Li, and A. P. C. Chan. 2023. "Automatic tower crane layout planning system for high-rise building construction using generative adversarial network." *Advanced Engineering Informatics*, 58: 102202. <https://doi.org/10.1016/j.aei.2023.102202>.
- Tam, C. M., T. K. L. Tong, and W. K. W. Chan. 2001. "Genetic Algorithm for Optimizing Supply Locations around Tower Crane." *J. Constr. Eng. Manage.*, 127 (4): 315–321. [https://doi.org/10.1061/\(ASCE\)0733-9364\(2001\)127:4\(315\)](https://doi.org/10.1061/(ASCE)0733-9364(2001)127:4(315)).
- Tam, V. W. Y., and I. W. H. Fung. 2011. "Tower crane safety in the construction industry: A Hong Kong study." *Safety Science*, 49 (2): 208–215. <https://doi.org/10.1016/j.ssci.2010.08.001>.
- Tsz Wai, C., P. Wai Yi, O. Ibrahim Olanrewaju, S. Abdelmageed, M. Hussein, S. Tariq, and T. Zayed. 2023. "A critical analysis of benefits and challenges of implementing modular integrated construction." *International Journal of Construction Management*, 23 (4): 656–668. <https://doi.org/10.1080/15623599.2021.1907525>.
- Wang, J., X. Zhang, W. Shou, X. Wang, B. Xu, M. J. Kim, and P. Wu. 2015. "A BIM-based approach for automated tower crane layout planning." *Automation in Construction*, 59: 168–178. <https://doi.org/10.1016/j.autcon.2015.05.006>.
- Younes, A., and M. Marzouk. 2018. "Tower cranes layout planning using agent-based simulation considering activity conflicts." *Automation in Construction*, 93: 348–360. Elsevier.
- Zhang, P., F. C. Harris, P. O. Olomolaiye, and G. D. Holt. 1999. "Location Optimization for a Group of Tower Cranes." *J. Constr. Eng. Manage.*, 125 (2): 115–122. [https://doi.org/10.1061/\(ASCE\)0733-9364\(1999\)125:2\(115\)](https://doi.org/10.1061/(ASCE)0733-9364(1999)125:2(115)).
- Zhang, Z., and W. Pan. 2021. "Multi-criteria decision analysis for tower crane layout planning in high-rise modular integrated construction." *Automation in Construction*, 127: 103709. <https://doi.org/10.1016/j.autcon.2021.103709>.
- Zhang, Z., and W. Pan. 2021. "Virtual reality supported interactive tower crane layout planning for high-rise modular integrated construction." *Automation in Construction*, 130: 103854. <https://doi.org/10.1016/j.autcon.2021.103854>.

A STUDY OF HUMAN-ROBOT COLLABORATIVE CONSTRUCTION TASK ALLOCATION PROBLEM CONSIDERING THE LEVEL OF TASK DANGER

Yilin Wang¹, Chao Mao², and Tingpeng Wang³

1) Graduate Student, School of Management Science and Real Estate, Chongqing University, Chongqing, China. Email: nicole_wyl@163.com

2) Ph.D., Prof., School of Management Science and Real Estate, Chongqing University, Chongqing, China. Email: maochaol201@126.com

3) Ph.D. Candidate, School of Management Science and Real Estate, Chongqing University, Chongqing, China. Email: wangtingpeng@stu.cqu.edu.cn

Abstract: In recent years, the construction industry has been expanding its production scale, however the industry is facing problems such as low productivity and frequent safety accidents. With the potential to improve construction quality and safety, construction robots have become an important means of solving the industry's problems. However, given that construction robots are built for specific operating environments, they are not yet able to fully replace manual labor, and thus human-robot collaboration is bound to be the mainstay of construction for quite some time to come. Therefore, this paper focuses on the problem of human-robot collaboration on construction sites, and innovatively proposes a task allocation framework based on the degree of danger. Taking concrete floor slab construction as an example, this paper firstly identifies the dangers in traditional construction scenarios and robotic construction scenarios, and then quantitatively evaluates each danger using the likelihood, exposure, and consequence level evaluation method. Further, this paper designs a task assignment framework that aims to minimize the degree of danger while taking into account the cost-effectiveness. The framework is intended to assist the construction team in making scientific decisions to accurately select the appropriate robot type for introduction and ensure that the overall cost is maintained within an economically feasible interval.

Keywords: Construction robot, Human-robot collaboration, Task allocation, Danger analysis.

1. INTRODUCTION

The construction industry, as an important contributor to the global economy, has been expanding its production scale in recent years, and the global construction market is expected to grow to \$13.9 trillion by 2037 from \$9.7 trillion in 2022 (Oxford Economics, 2023). However, the industry is currently facing problems such as low productivity, frequent safety accidents, serious energy consumption, and low levels of information technology and industrialization (Zeng et al., 2003). Due to the aging of the construction workforce, the high exit rate from the industry, and the slowing down of the upgrading of workers' skill levels (Kumarage et al., 2023), surveys have already indicated that by 2035, the labor shortage in China's construction industry will reach 10.97 million (Wang et al., 2021), and the problem of imbalance in the supply and demand of labor will become increasingly serious. In addition, the construction industry has been recognized as the most dangerous industry in the world after mining and quarrying (International Labour Organization, 2023), and nearly 3 million workers worldwide die from workplace accidents or occupational diseases every year, of which 1 in 6 die in the construction industry (Cheng et al., 2020). Therefore, there is an urgent need to address the safety and labor dilemma in the construction industry. Construction robots, which can assist and replace "dangerous, laborious, dirty, and heavy" construction work, have the potential to improve construction efficiency and ensure construction safety and quality, and have become a mainstream solution to address these challenges facing the industry (Cai et al., 2019; Gharbia et al., 2020).

Various construction robots have been developed and applied (Zhang et al., 2022). Examples include measuring robots (Oshio et al., 2022), bricklaying robots (Melenbrink et al., 2020), and others. However, most of these robots are designed for specific environments, and it is difficult to quickly adapt to highly unstructured construction environments with only one type of robot (Liang et al., 2021). Therefore, humans are needed to utilize their strengths in responsiveness and intuitive decision making to assist construction robots in accomplishing construction tasks effectively. In the field of architecture, engineering and construction, human-robot collaboration is defined as workers and construction robots working together to accomplish a specific construction task, and each role must be able to contribute to the overall system (Zhang et al., 2023). Workers can play a variety of roles during their collaboration with construction robots, including supervisors, operators, teammates, programmers, and in some cases, as bystanders. In contrast, operator and teammate, which require deep interaction, are the most researched human-robot collaboration problems. In this context, operator refers to a construction worker who uses an operating device to control a robot to complete a construction task (Lee et al., 2011), and teammate refers to a worker who works with a robot to complete a construction task (Xiang et al., 2021). This study also focuses on these two types of roles in human-robot collaboration. Compared to traditional construction methods, human-robot collaboration avoids the problems faced when performing tasks

with separate humans or robots and can achieve significantly better human-robot team performance than separate human or robot operations through complementary human-robot strengths (Seeber et al., 2020). However, the division of labor between construction workers and robots is not clear in human-robot collaborative systems. How workers and robots should cooperate to improve the safety and efficiency of the system has also become one of the challenging issues in the new on-site construction human-robot collaboration management model (Wang et al., 2024). This is exactly the human-robot collaboration task allocation problem discussed in this study.

The problem of task allocation for human-robot collaboration has become an increasingly popular research area (Tsarouchi et al., 2016), and reasonable task allocation between humans and robots can improve the productivity and stability of construction. However, the problem of human-robot collaborative task allocation is mainly focused on product assembly (Chen et al., 2014), and is rarely mentioned in the construction field. Although methods for task allocation among multiple construction robots have been proposed (Ye et al., 2024), it did not include workers at the construction site in the task allocation, and was proposed based on the assumptions of the capabilities of construction robots, which were not applicable to existing construction robots in the market. Meanwhile, while previous task allocation problems for humans and robots have focused on dimensions such as efficiency (Bogner et al., 2018), operational comfort (Kim et al., 2018), and resource utilization (Malik & Bilberg, 2019), this study focuses on the inherently hazardous nature of the construction industry and aims to use the task danger level as the main measure to rationalize the division of work tasks and provide a basis for human-robot collaboration in the construction industry.

2. LITERATURE REVIEW

Research related to human-robot collaboration in the construction field focuses on human-robot interaction and safety of human-robot collaboration. As a realization of human-robot collaboration, human-robot interaction is mainly used for the execution and scheduling of collaborative tasks to improve construction efficiency (Wilcox & Shah, 2012). Wang & Zhu (2021) proposed a vision-based framework to capture and interpret workers' gestures aimed at controlling construction machinery, which has significant advantages in the noisy context of building construction. Fang et al. (2018) developed a computer vision-based approach that utilizes two convolutional neural network models to determine whether a worker is wearing a seatbelt while working at height, enabling automated detection of workers who are not wearing a seatbelt.

As a guarantee for human-robot collaboration, safety is a key factor to consider for human-robot collaboration systems. ISO 10218-1/2 has specified safety requirements for industrial robot collaboration with four control modes including safety monitoring stationary, hand-held guidance, speed and distance monitoring, and power and force limitation (Kim et al., 2019; Yan et al., 2020). However, it focuses mainly on industrial robotic systems, and it cannot be fully generalized to the construction industry's robots and work environments. You et al. (2018) proposed a robot acceptance safety model to explore different approaches to overcome the challenges of human-robot collaborative workspace and to enhance safety perception. Kim et al. (2019) proposed a drone-based visual monitoring approach to detect impact hazards around workers in advance and achieve timely intervention to improve the safety of construction workers' work environment. However, the above studies mainly focus on the development of systems that can monitor the status of construction sites in real time from a technological perspective to improve construction safety, and do not consider the impact of the task allocation situation on construction safety from the perspective adopted by the industry, and the design of the human-robot model to reduce construction hazards is still subject to further research. You et al. (2018) showed that during the human-robot collaboration process, human mental stress and health condition are in need of special attention, and that a separated workspace is more likely to increase humans' trust in the robot, thus improving their perceived safety and reducing their worries and fears.

3. METHODOLOGY

The purpose of this paper is to establish a set of danger level metrics applicable to human-robot collaborative construction, and then propose a human-robot task allocation framework based on this danger level, and apply corresponding algorithms to solve the task allocation model. Eventually, we will apply this danger level-based human-robot collaborative construction task allocation method to real-world scenarios to verify its effectiveness and practicality.

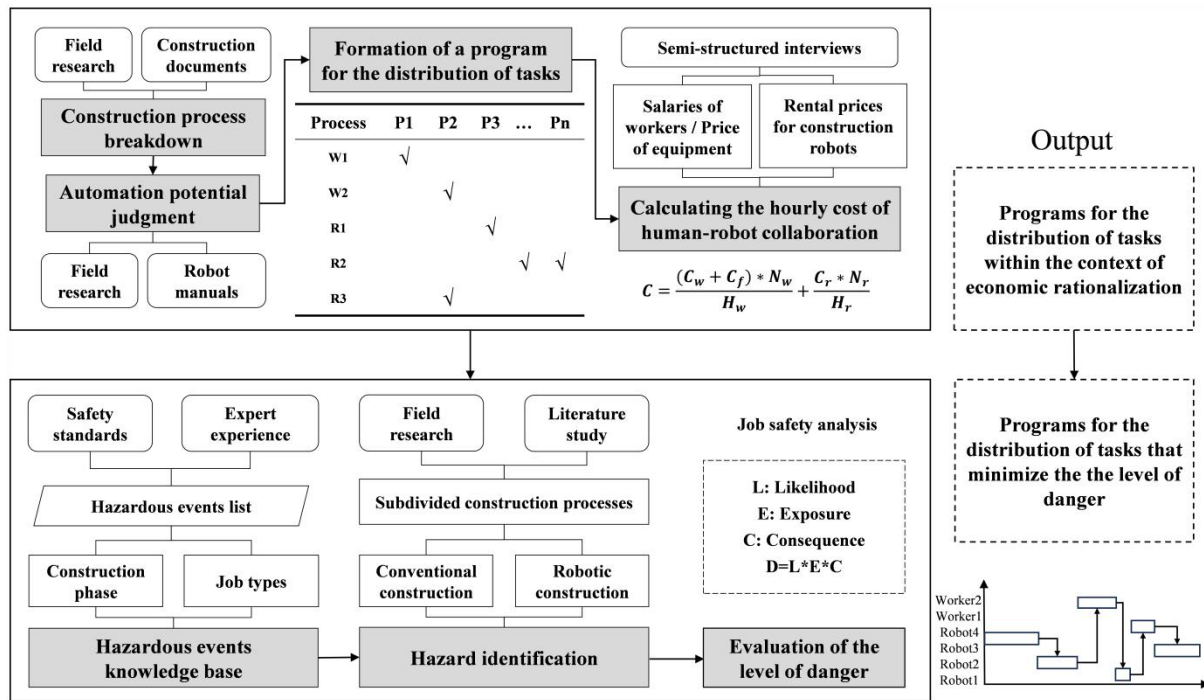


Figure 1. Methodology

The construction task allocation framework is shown in Figure 1. Firstly, by systematically sorting out relevant documents such as the *Building Construction Manual* and combining the actual situation of the on-site research, we carried out a comprehensive and detailed division of the construction process. This step is the basis for ensuring the accuracy of the subsequent task allocation. For the segmented processes, we comprehensively considered the technical capabilities and application status of the existing construction robots in the market, and judged whether the process is suitable to be completed by construction robots. For those construction processes that can be performed by robots, we adopt a permutation and combination approach to form multiple task allocation schemes. In order to determine the optimal task allocation plan under the premise of economic rationality, we made full use of the cost data obtained from the research, such as workers' salaries and robot rental prices, and evaluated the cost-effectiveness of each plan one by one. This step not only helped us to identify economically viable options, but also provided strong data support for subsequent optimisation.

However, cost was not the only factor we considered. While ensuring economic rationality, we also conducted an in-depth analysis and calculation of the danger level of each tasking option. In order to comprehensively and accurately assess the degree of danger, we have extensively collected valuable information such as accident cases, safety standards and expert experience. Through comprehensive analysis and generalisation, we came up with a list of hazardous events that each construction stage and each construction work type may face. Based on this, we further identified potential danger sources for each construction process in both traditional construction scenarios and robotic construction scenarios. The identification of these hazardous sources provides an important basis for our subsequent evaluation of the degree of danger. In order to arrive at a scientific and accurate evaluation result, we used the likelihood, exposure, and consequence level evaluation method to evaluate each danger source one by one. Through this step, we ultimately identified the task assignment programme with the lowest level of task danger.

3. CONSTRUCTION TASK ALLOCATION MODEL

Human-robot collaborative construction task allocation is the process of decomposing a complex construction task into sub-tasks of appropriate granularity and assigning the task to a construction worker or a construction robot to achieve the global optimum under consideration of specific constraints and objectives. In this paper, we have evaluated the scheme of assigning construction tasks based on two indicators: the cost and the degree of danger of the task.

3.1 The Cost of Human-Robot Collaboration

In the construction industry, human-robot collaborative automation is emerging as a key tool for improving construction safety and productivity. In order to ensure that this automation modification is economically viable, we can assess the cost-effectiveness of a human-robot collaboration system by developing a mathematical model. This model will help us to compare the costs of traditional construction methods with those

of human-robot collaborative methods to ensure that the costs are kept within reasonable limits. The cost of working with human-robot collaboration system depends on the cost of equipment and labour (Sun et al., 2024). We focus on analysing equipment and labour costs to provide data support to construction project managers so that they can determine the optimal number of workers and construction robots required for a given project, thus achieving the optimal balance of cost-effectiveness and productivity. In short, this model is designed to help us find cost-effective configurations of human-robot collaboration to improve the overall economics of construction projects. Table 1 shows the parameters involved in the model.

Table 1. Parameters in the cost model	
Symbol	Definition
C	The hourly cost of human-robot collaboration
C _w	Salaries of workers
C _f	Price of facilities
C _r	Rental prices for construction robots
N _w	Number of workers
N _r	Number of robots
H _w	Working hours of workers
H _r	Working hours of construction robots

Here, the parameters that affect the cost are the price of the construction robot, the price of the equipment used by the workers, and the salary of the workers. Therefore, the following equation (1) should be used to calculate the cost per hour of work for the human-robot collaboration system, where each parameter is defined as shown in Table 1.

$$C = \frac{(C_w + C_f) * N_w}{H_w} + \frac{C_r * N_r}{H_r} \quad (1)$$

3.2 The Danger of Human-Robot Collaboration

In the environment where workers and construction robots work together, with the application of construction robots, some of the construction tasks originally undertaken by workers are transferred to the robots. At this time, the task dangers faced by workers in the human-robot collaboration system contain two main aspects: the dangers in the traditional construction environment, and the dangers arising from working together with construction robots, as shown in Figure 2.



Figure 2. Components of danger for human-robot collaborative systems

In human-robot collaboration systems, construction workers are still required to perform part of the work. Danger assessment for this part of the task mainly focuses on unsafe human behavior, unsafe state of objects and unsafe environment in traditional construction scenarios. This study thoroughly considered the specific characteristics of construction sites and synthesized information from relevant technical specifications of construction projects, past accident cases, and other information to conduct a detailed job safety analysis of each specific construction process segment. Through in-depth communication with a number of relevant personnel, including safety officers, production managers, project managers, etc., we have compiled a list of hazardous events in concrete floor slab construction as follows, as shown in Table 2.

Table 2. List of hazardous events in concrete floor slab construction	
Construction process	Hazardous events list
Fabric pouring	(1) The support surface of the fabric machine is not flat and solid enough, and

	<p>more than 4 wire ropes are not used to tie diagonally, which may cause overturning phenomenon.</p> <p>(2) When installing, dismantling, or shifting the manual fabricator, there are unrelated people staying within the operation range.</p> <p>(3) When the pump pipe is clogged and needs to be flushed, the pump pipe port is not equipped with a reliable protective cover, and the operator stays at the front end of the pump pipe and within the rotating radius.</p> <p>(4) People stay under the overhang and around the discharge hose during operation of the fabricator.</p> <p>(5) Fabric machine operator violates rules during operation.</p> <p>(6) Failure to stop the construction operation of the fabricator in inclement weather, such as gusts of wind of grade 5 or above, thunderstorms, heavy snow, fog, etc., and the possibility of tipping over.</p>
Vibration	<p>(1) Failure of the operator of the concrete vibrator to wear insulated shoes and gloves.</p> <p>(2) Workers' hands or other parts may be pinched or hit by the moving parts of the vibrator.</p> <p>(3) Concrete vibration during work at height or at the edge of the work, if the safety protection measures are not in place, collapse or worker fall accidents may occur.</p>
Paving	<p>(1) During concrete paving, the ground can become very slippery and workers may slip or fall due to loss of footing.</p>
Aluminum ruler scraping	<p>(1) Prolonged bending, squatting, or repeating the same movement may lead to damage to a worker's musculoskeletal system and chronic pain or injury.</p> <p>(2) May slip and fall due to slippery floors, resulting in a fall or other bodily injury.</p>
Compacting, grinding	<p>(1) The use of machinery and equipment, such as compactors and mills, can lead to worker injuries if they are not operated properly or if the equipment is poorly maintained.</p>

In addition, construction robots still need the assistance of workers in performing their tasks, and construction robots themselves, as a newly introduced work environment, bring new safety risks to workers. For the danger assessment of this part of the task, we focus on the unsafe behaviors of people and the unsafe state of objects in the construction robot operating scenario. This study follows the ISO 10218 standard and combines the specific situation of the actual application of construction robots to list the list of hazardous events that may be involved in various construction robots for each construction robot in the concrete engineering category, as shown in Table 3.

Table 3. List of hazardous events for robots in the concrete engineering category

Type of construction robots	Hazardous events list
Intelligent follower fabricator	<p>(1) Loss of control of fabric robot system poses risks to workers, including bumps and bruises.</p> <p>(1) During the operation of the robot, staying, playing or fooling around within 3m in the forward direction, 1m on both sides and the turning radius of the robot may pose a risk of being hit.</p>
Ground leveling robot	<p>(2) Operation of this product by untrained or inexperienced persons.</p> <p>(3) Failure to wear the required labor protection equipment before operating the equipment: helmet, reflective clothing, labor protection shoes, noise-reducing earplugs, and protective masks.</p>
Crawler trowel robot	<p>(1) When the robot is operating, there is a person within 1m of the operating range.</p> <p>(2) Operation of this product by untrained or inexperienced persons.</p>
Polishing robot	<p>(1) During the operation of the robot, staying, playing or fooling around within 3m in the forward direction, 1m on both sides and the turning radius of the robot may pose a risk of being hit.</p> <p>(2) Operation of this product by untrained or inexperienced persons.</p> <p>(3) Failure to wear the required labor protection equipment before operating the equipment: helmet, reflective clothing, labor protection shoes, noise-reducing earplugs, and protective masks.</p>

While the introduction of construction robots has eased the burden of risky labor on workers to some extent, the change has not been without its pitfalls. It is also accompanied by new challenges and risks, such as accidents that may be caused by robot mishandling or poor maintenance.

In view of this, this study considers the level of danger to workers in completing tasks as a core assessment indicator, aiming to minimize this level of danger through in-depth research. This study draws on the traditional method of danger assessment of operating conditions-the likelihood, exposure, and consequence level evaluation method to quantitatively analyze the overall safety condition of the construction site. In calculating the degree of danger for workers to complete tasks in the human-robot collaboration system, we use the following equation (2).

$$D = D_w + D_r \quad (2)$$

Where, D represents the level of danger to the worker in the human-robot collaboration system to accomplish the task; D_w represents the level of danger faced by the worker from the traditional construction environment; and D_r represents the level of danger faced by the worker from the construction robot.

3.3 Mathematical Model

The mathematical model for task allocation is as follows. By equation (3), we can make the danger level minimized. By equation (4), we can ensure that the cost of the human-robot collaboration system is less than the traditional construction.

$$\min D = \sum_{i=1}^n (D_{wi} + D_{ri}) \quad (3)$$

$$s. t. \sum_{i=1}^n \left(\frac{(C_w + C_{fi}) * N_{wi}}{H_w} + \frac{C_{ri} * N_{ri}}{H_r} \right) \leq C_\alpha \quad (4)$$

Where, the definition of each parameter is the same as Eq. (1)(2), i represents the i^{th} process ($i=1, 2, 3, \dots, n$), C_α represents the cost of the work of the human-robot collaboration system when the construction task is traditionally done manually only.

3.4 Solution Algorithm

For the above mathematical model, we set up a specific encoding method for the task allocation solution of human-robot collaborative construction, and realized the task scheduling of human-robot collaborative construction by using the violent search algorithm. Among the optimization objectives are that the cost of the human-robot collaboration system is less than the cost of conventional construction, i.e., the cost of doing it all manually, and that the optimal solution is the least hazardous.

The following is the design of the key computational process, as shown in Table 4.

Table 4. The key computational process	
Algorithm for Optimal Subtask Allocation	
1	Input: process data, robot data, cost, and danger
2	Output: optimal allocation solution, allocation cost, allocation danger
3	Generate all possible allocation solutions:
4	For each process, generate allocations (0 for manual, 1 for robot)
5	Define function to calculate total cost and total danger:
6	Function: calculate_cost_and_danger(allocation)
7	Initialize total_cost = 0, total_danger = 0
8	For each process in the allocation:
9	If manual (0):
10	Add manual cost to total_cost
11	Add manual danger to total_danger
12	If robot (1):
13	Add robot cost to total_cost
14	Add robot danger to total_danger
15	Record robot usage
16	Return total_cost, total_danger
17	Find optimal allocation solution:
18	Initialize best_index = None, min_danger = ∞
19	For each allocation:
20	Get its cost and danger
21	If cost < manual cost and danger < min_danger, update

	optimal solution
22	Output optimal solution:
23	If optimal allocation is found, output allocation, cost, and danger
24	Otherwise, output "No optimal solution found"

4. CASE STUDY

In this paper, the concrete floor slab construction process is chosen as a case background to verify the feasibility of the proposed method. The concrete floor slab construction process includes pre-construction preparation, concrete mixing, fabric pouring, vibration and trimming. This paper focuses on the concrete construction robots of Bright Dream Robotics, a leading company in China's construction robotics industry, which produces a wide range of construction robots, among which the concrete construction robots include intelligent follower fabricators, ground leveling robots, crawler trowel robots, and polishing robots. These advanced robots have been practically applied in several construction projects. Given the scarcity and confidentiality of data on construction robots, cost data on construction robots in this study were collected from interviews and publicly shared cases. As a preliminary exploration, a hospital project in Chongqing is planned to carry out a robot trial on a 1000m² site at a manageable cost, so only one of each type of construction robot will be introduced for the trial. The correspondence between the concrete floor construction process and the construction robot is shown in Figure 3 below.

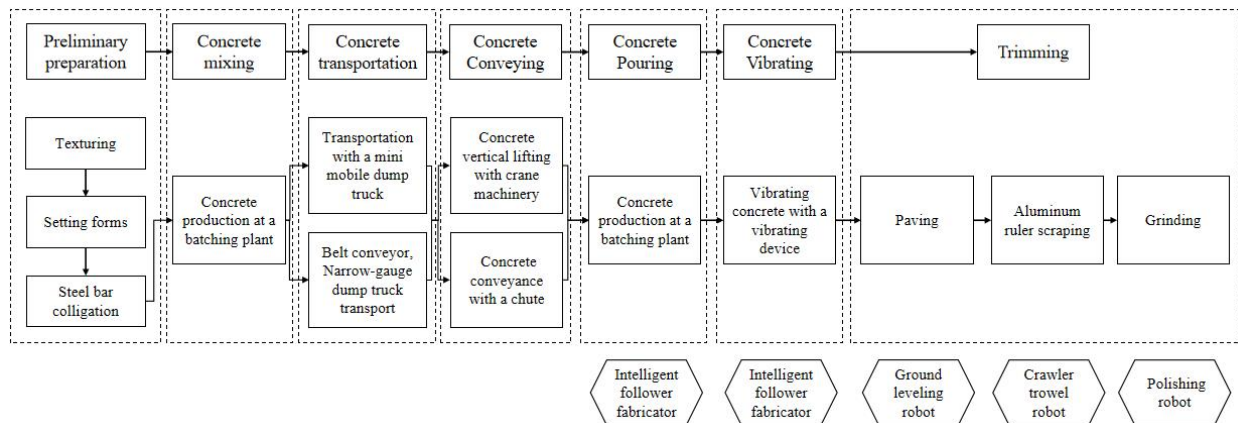


Figure 3. Construction process of concrete floor slabs

In this case, 4 front-line personnel in construction and safety, 4 managers such as technical leaders and production managers are invited, and a total of 8 people score the hazardous events of each construction process and each construction robot. According to the 8 scoring results, the danger coefficient of each dangerous event can be derived in accordance with the formula for calculating the danger of each type of dangerous event.

To establish a model for solving the task allocation problem based on the violent search algorithm, this paper uses the python platform to write the algorithmic program to derive the optimal task allocation scheme with the minimum degree of danger and the cost within a reasonable range, as shown in Figure 4 below. In the distribution scheme, preliminary preparation is assigned to worker1, concrete transportation is assigned to worker2, concrete pouring and concrete vibrating is assigned to intelligent follower fabricators, paving is assigned to worker3, AI ruler scraping is assigned to crawler trowel robot, grinding is assigned to worker4. Ground leveling robot and polishing robot are not used in this case.

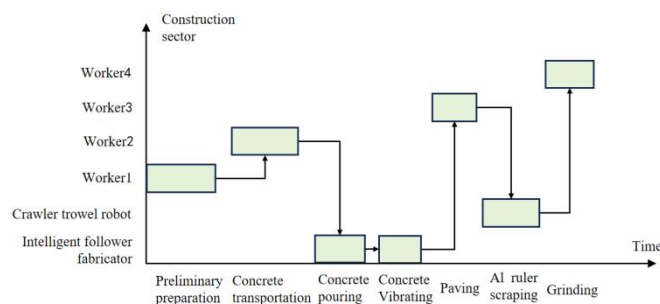


Figure 4. Optimal distribution of tasks program

5. CONCLUSIONS

Efficient cooperation between man and robot can improve construction productivity and safety. Appropriate human-robot collaboration task allocation is the premise and foundation for realizing good results of human-robot cooperative work. Meanwhile, the presentation of human-robot task assignment results can be considered as a strong motivation for construction teams to introduce construction robots, i.e., to reduce the level of danger and to keep the cost within reasonable limits. In addition, the design of construction robots for adapting to diversified and variable construction needs provides strong theoretical and technical support for the government to accelerate the construction of digital talent echelon, optimize the existing workforce structure, and reduce the low labor fitness caused by technological friction.

However, human-robot collaborative construction task assignment is a very complex process. This study establishes a human-robot collaborative construction task allocation model to solve this problem by taking the task danger level as the goal and taking into account the cost, and the case study results show that the human-robot collaborative task allocation method proposed in this study can effectively find the optimal solution. However, the worker and robot task allocation problem still needs further research. An effective human-robot collaborative system should consider the strategy of human-robot collaborative task allocation from multiple dimensions, not only considering the economic efficiency and safety, but also considering the time efficiency and completion quality of the workers and robots when they work together.

ACKNOWLEDGMENTS

This work was supported by the Chongqing Natural Science Fund Project (Grant Number CSTB2022NSCQ-MSX1622).

REFERENCES

- Bogner, K., Pferschy, U., Unterberger, R., and Zeiner, H. (2018). Optimised scheduling in human–robot collaboration—a use case in the assembly of printed circuit boards. *International Journal of Production Research*, 56(16), 5522–5540.
- Cai, S., Ma, Z., Skibniewski, M. J., and Bao, S. (2019). Construction automation and robotics for high-rise buildings over the past decades: A comprehensive review. *Advanced Engineering Informatics*, 42, 100989.
- Chen, F., Sekiyama, K., Cannella, F., and Fukuda, T. (2014). Optimal subtask allocation for human and robot collaboration within hybrid assembly system. *IEEE Transactions on Automation Science and Engineering*, 11(4), 1065–1075.
- Cheng, M.-Y., Kusoemo, D., and Gosno, R. A. (2020). Text mining-based construction site accident classification using hybrid supervised machine learning. *Automation in Construction*, 118, 103265.
- Fang, W., Ding, L., Luo, H., & Love, P. E. D. (2018). Falls from heights: A computer vision-based approach for safety harness detection. *Automation in Construction*, 91, 53–61.
- Gharbia, M., Chang-Richards, A., Lu, Y., Zhong, R. Y., and Li, H. (2020). Robotic technologies for on-site building construction: A systematic review. *Journal of Building Engineering*, 32, 101584.
- International Labour Organization. (2023). *A Call For Safer And Healthier Working Environments*. Retrieved from International Labour Organization website: <https://www.ilo.org/publications/call-safer-and-healthier-working-environments>
- Kumarage, N., Gill, S. P. K., Gill, A., Kaluthantirige, P., Silva, L., Hewage, K., and Ruwanpura, J. (2023). Construction labour shortage, challenges, and solutions: A survey-based approach. In *Canadian Society of Civil Engineering Annual Conference* (pp. 399–413). Cham: Springer Nature Switzerland.
- Kim, W., Lee, J., Peternel, L., Tsagarakis, N., and Ajoudani, A. (2018). Anticipatory robot assistance for the prevention of human static joint overloading in human–robot collaboration. *IEEE Robotics and Automation Letters*, 3(1), 68–75.
- Kim, D., Liu, M., Lee, S., & Kamat, V. R. (2019). Remote proximity monitoring between mobile construction resources using camera-mounted UAVs. *Automation in Construction*, 99, 168–182.
- Lee, S., Yu, S., Choi, J., and Han, C. (2011). A methodology to quantitatively evaluate the safety of a glazing robot. *Applied Ergonomics*, 42(3), 445–454.
- Liang, C.-J., Wang, X., Kamat, V. R., and Menassa, C. C. (2021). Human–robot collaboration in construction: Classification and research trends. *Journal of Construction Engineering and Management*, 147(10), 03121006.
- Malik, A. A., and Bilberg, A. (2019). Complexity-based task allocation in human-robot collaborative assembly. *Industrial Robot: The International Journal of Robotics Research and Application*, 46(4), 471–480.
- Melenbrink, N., Werfel, J., and Menges, A. (2020). On-site autonomous construction robots: Towards unsupervised building. *Automation in Construction*, 119, 103312.
- Oshio, K., Tsujimoto, M., Taniguchi, K., Obo, T., and Kubota, N. (2022). 2D-SLAM of illuminance

- measurement robot using 3D-LiDAR and IMU on slopes. *2022 Joint 12th International Conference on Soft Computing and Intelligent Systems and 23rd International Symposium on Advanced Intelligent Systems (SCIS&ISIS)*, 1–6.
- Oxford Economics. (2023). *Global Construction Futures*. Retrieved from Oxford Economics website: <https://www.oxfordeconomics.com/resource/global-construction-futures/>
- Seeber, I., Bittner, E., Briggs, R. O., De Vreede, T., De Vreede, G.-J., Elkins, A., Maier, R., Merz, A. B., Oeste-Reiß, S., Randrup, N., Schwabe, G., and Söllner, M. (2020). Machines as teammates: A research agenda on AI in team collaboration. *Information & Management*, 57(2), 103174.
- Sun, B., Mao, C., Wang, T., and Li, Z. (2024). Cost assessment framework for construction robots: comparative study of robotic and traditional construction. *Journal of Management in Engineering*, 40(5), 05024009.
- Tsarouchi, P., Makris, S., and Chrysosolouris, G. (2016). Human–robot interaction review and challenges on task planning and programming. *International Journal of Computer Integrated Manufacturing*, 29(8), 916–931.
- Wang G., XU K., and Cao D. (2021). Scenario prediction and corresponding measures of labor demand in China's construction industry for 2035. *Journal of Civil Engineering and Management*, 38(4), 15–22.
- Wang, T., Mao, C., Sun, B., and Li, Z. (2024). Genealogy of construction robotics. *Automation in Construction*, 166, 105607.
- Wang, X., & Zhu, Z. (2021). Vision–based framework for automatic interpretation of construction workers' hand gestures. *Automation in Construction*, 130, 103872.
- Wilcox, R., & Shah, J. (2012). Optimization of Multi-Agent Workflow for Human-Robot Collaboration in Assembly Manufacturing. In *Infotech@Aerospace 2012*. American Institute of Aeronautics and Astronautics.
- Xiang, S., Wang, R., and Feng, C. (2021). Mobile projective augmented reality for collaborative robots in construction. *Automation in Construction*, 127, 103704.
- Ye, X., Guo, H., and Luo, Z. (2024). Two-stage task allocation for multiple construction robots using an improved genetic algorithm. *Automation in Construction*, 165, 105583.
- Yan, X., Zhang, H., & Li, H. (2020). Computer vision-based recognition of 3D relationship between construction entities for monitoring struck-by accidents. *Computer-Aided Civil and Infrastructure Engineering*, 35(9), 1023–1038.
- You, S., Kim, J.-H., Lee, S., Kamat, V., & Robert, L. P. (2018). Enhancing perceived safety in human–robot collaborative construction using immersive virtual environments. *Automation in Construction*, 96, 161–170.
- Zeng, S. X., Tam, C. M., Wang, H. C., and Deng, Z. M. (2003). Overcoming problems associated with sustainable development of the construction industry in China. *Architectural Science Review*, 46(4), 353–361.
- Zhang, J., Luo, H., and Xu, J. (2022). Towards fully BIM-enabled building automation and robotics: A perspective of lifecycle information flow. *Computers in Industry*, 135, 103570.
- Zhang, M., Xu, R., Wu, H., Pan, J., and Luo, X. (2023). Human–robot collaboration for on-site construction. *Automation in Construction*, 150, 104812.

Energy Consumption Modeling for a Robot Arm in 3D Concrete Printing

Tong Liu¹, Song Du², Fei Teng³, and Yiwei Weng⁴

- 1) Ph.D. Candidate, Department of Building and Real Estate, The Hong Kong Polytechnic University. Hong Kong SAR. Email: tong3213.liu@connect.polyu.hk
- 2) Ph.D. Candidate, Department of Building and Real Estate, The Hong Kong Polytechnic University. Hong Kong SAR. Email: song.du@connect.polyu.hk
- 3) Ph.D. Candidate, Department of Building and Real Estate, The Hong Kong Polytechnic University. Hong Kong SAR. Email: fei0708.teng@connect.polyu.hk
- 4) Ph.D. Assistant Professor, Department of Building and Real Estate, The Hong Kong Polytechnic University. Hong Kong SAR. Email: yiwei.weng@polyu.edu.hk

Abstract: The construction sector is a major contributor to global CO₂ emissions and energy consumption. 3D concrete printing (3DCP) provides sustainable solutions to tackle the environmental challenges. However, the long-time and continuous operation of robotic arm printers in 3DCP incur critical challenges in energy efficiency. To address these challenges, this study aims to develop an energy consumption (EC) model of a robotic arm in 3DCP. The proposed EC model has desirable agreement compared to the experimental result, achieving an accuracy 99.51%. The impact of the proposed EC model is evaluated by printing a pre-designed path with various positions. Results reveal that the EC reduction can achieve up to 53.72% with varying positions. The findings reveal that the proposed EC model has the potential to reduce the EC for energy efficiency in 3DCP.

Keywords: 3D concrete printing, energy consumption, positions, robots

1. INTRODUCTION

The construction sector is a main contributor to global carbon emissions and energy consumption due to the huge amount of cement production (Li et al., 2024). 3D concrete printing (3DCP) provides the potential to address these challenges with enhanced productivity and sustainability (Mechtcherine et al., 2018). Previous research has demonstrated that 3DCP outperforms precast concrete by reducing 25.4% the overall cost, 85.9% in CO₂ emissions, and 87.1% in energy consumption concerning the production of a prefabricated bathroom unit (Weng et al., 2020). However, 3DCP typically exhibits a higher electricity consumption than conventional construction since a long operation of the robotic arm is required for large-scale printing (Zhang et al., 2018). It is pressingly needed to enhance the energy efficiency of robotic operation during printing.

Kajzr et al., (2024) investigated the energy consumption (EC) of a robotic arm in 3DCP in a simulated environment. The results demonstrate that the motion characteristics of robotic arms and their operational scenarios significantly influence energy consumption, presenting challenges for accurate energy modeling. The EC behavior of a robotic arm is determined by the joint configurations, motion trajectories, and tasks (Wang et al., 2018). Most of the existing EC optimization methods mainly focus on specific applications such as welding (Zhou et al., 2022), polishing (Cao et al., 2020), and other assembly processes (Soori et al., 2023). For instance, Pellicciari et al., (2013) and Feng et al., (2021) proposed methods to optimize robotic trajectories for energy-saving in pick-and-place tasks. The results show that a 10% reduction of EC reduction can be achieved based on flexible joint configurations (Feng et al., 2021). Liu et al., (2018) investigated the correlation between robot speed and EC, identifying an optimal speed range for energy minimization. Gadaleta et al., (2021) conducted extensive experiments on industrial robots, revealing an energy-saving potential exceeding 50%.

However, few studies have been conducted with respect to the application of the EC model in 3DCP. 3DCP typically exhibits unique characteristics compared to the processes abovementioned, including 1) constrained joint configurations due to a vertically fixed concrete extruder, 2) high gravitation and fractional force, and 3) the requirement of a continuous toolpath. To fill the above research gaps, the first step is to construct the EC model, taking 3DCP characteristics into account. In this study, firstly, an EC model compatible with 3DCP is proposed. The EC model is established by determining a parameter set of a robotic arm. The effect of printing positions on the EC behavior of the robotic arm is evaluated using a pre-designed printing path.

2. METHOD

The movement of a robot is powered by joint torque. Joint torques are associated with the robot's dynamic parameters, including mass, the center of mass, inertia tensor, and frictional joint torque. The robot dynamic model establishes the relationship between joint torque and joint configuration, expressed by:

$$\tau = M(q)\ddot{q} + C(q, \dot{q})\dot{q} + G(q) + \tau_f(q) \quad (1)$$

where τ represents the joint torque vector, q , \dot{q} , and \ddot{q} are joint position vector, joint velocity vector, and joint acceleration vector, respectively. $M(q)$ is the positive definite inertia matrix, $C(q, \dot{q})$ is the Coriolis-Centrifugal matrix, $G(q)$ is the torque vector induced by gravity. $\tau_f(q)$ is the frictional vector. In this study, the friction vector can be formulated as (Paes et al., 2014):

$$\tau_f(\dot{q}) = f_v\dot{q} + f_c \text{sign}(\dot{q}) \quad (2)$$

where the f_v and f_c represent viscous and Coulomb friction coefficients, respectively.

For each link, a set of dynamic parameters needs to be determined, including the mass (m_i), center of mass ($C_{m,i}$), the inertia tensor ($I_{m,i}$), as well as viscous (f_{vi}) and Coulomb friction items (f_{ci}). The dynamic parameter vector (π_i) of link typically contains parameters related to the mass, center of mass, and inertia:

$$\pi_i = [m_i, C_{m,i}, I_{m,i}, f_{ci}, f_{vi}] \quad (3)$$

and a dynamic parameter set (Π) to represent the collection of robot dynamic parameters:

$$\Pi = [\pi_1, \pi_2, \dots, \pi_6]^T \quad (4)$$

where Π defines the used to create the robot links object. The robot object can be created based on the determined parameter set Π . Afterward the modeled joint torque τ^{mod} can be calculated by the recursive Newton-Euler method *rne*, given joint position q , velocity \dot{q} , and acceleration \ddot{q} (Chignoli et al., 2023). This operation can be represented by:

$$\tau = rne(q, \dot{q}, \ddot{q}) \quad (5)$$

the robot real-time power $P(t)$ satisfies:

$$P(t) = \tau(t)q(t) \quad (6)$$

during a time period of robotic motion T , the total EC can be calculated with Eq. (6):

$$E = \int_T P(t)dt \quad (7)$$

Rapid is a programming language designed to control robotic systems. In the motion of a robot, the joint data can be collected via the function of *GetJointData* in Rapid and sent to the computer. The collected data, including joint torque τ^{col} and joint configuration $(q, \dot{q}, \ddot{q})^{col}$, can be used to calculate the collected power $P^{exp}(t)$ and total EC E^{exp} by Eqs. (6) and (7) from experiment, respectively. It should be noted that *GetJointData* cannot provide the data of joint acceleration, and therefore, the collected joint acceleration (\ddot{q}_{col}) is calculated numerically through a central finite difference method (Teixeira et al., 2023). The modeled power $P^{mod}(t)$ and EC E^{mod} from the experiment is calculated by Eqs. (5)~(7) using the collected joint configurations $(q, \dot{q}, \ddot{q})^{col}$.

The dynamic parameter set Π can be determined by minimizing the variance between the collected real-time power $P^{col}(t)$ and modeled power $P^{mod}(t)$, which are discrete data. The total number of collected discrete data is labeled as N , P_n^{col} , and P_n^{mod} , utilized to represent n -th collected and modeled instantaneous power. To determine the dynamic parameter set, an objective function is defined as below:

$$C = \sqrt{\frac{1}{N} \sum_{n=1}^N (P_n^{exp} - P_n^{mod})^2} \quad (8)$$

where the objective function C indicates the difference between the collected and modeled power. Once the dynamic parameters set Π is determined, the objective function C reaches its minimum.

In this study, the test trajectory is pre-defined for dynamic parameter determination. As shown in Figure 1 (a), a pre-designed printing path is adopted for experiment, and the position is represented by a central reference point. Figure 1 (b) illustrates the printed concrete structures along the pre-designed path. Figure 1 (c) shows the setup of the printing task, and Figure 1 (d) presents the start point (Ps), middle point (Pm), and end point (Pe) in the trajectory for power comparison. The real-time power data (P_i^{col}) will be collected during the robotic operation. The dynamic parameter set of the robot arm will be determined to construct the EC model by minimizing the defined objection function C in Eq. (8).

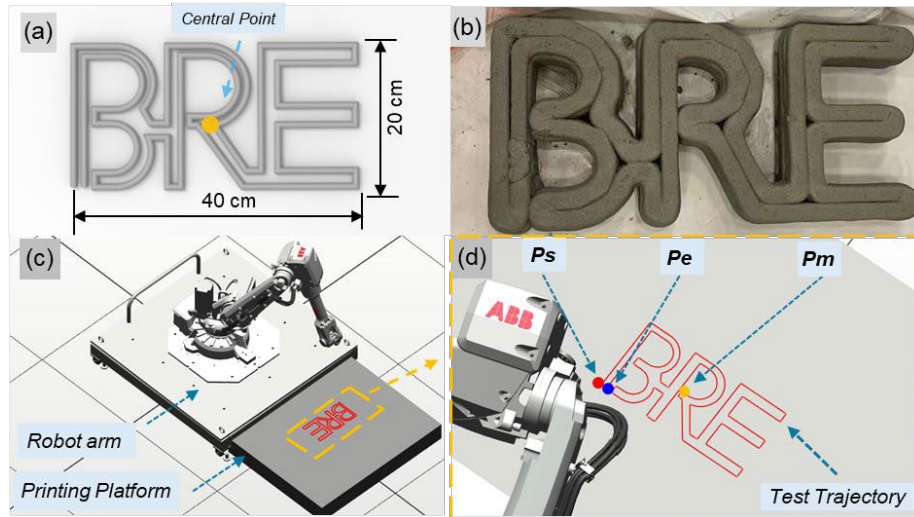


Figure 1. Motion trajectory illustration: (a) dimensions of the printing path; (b) printed specimens; (c) printing setup and area; (d) a continuous robot trajectory for printing.

3. RESULTS AND DISCUSSION

3.1 Comparison between modeled and collected test trajectory

In this work, Π and C are regarded as independent and dependent variables, respectively, and the function of *fminunc* in Matlab-2024 is applied to optimize C and determine the parameter set Π . Then, the real-time power and the EC behavior during robot motion can be determined in theoretical models.

Figure 2 (a) shows the comparison between the theoretical and experimental instantaneous power-time curves. The local maximum can be detected when the end effector moves at the P_m . At least five experiments were conducted for the test trajectories. The EC can be calculated by employing Eq. (6), and experimental and simulation results can be seen in Figure 2 (b). The accuracy of EC prediction is 99.51%, indicating that the proposed EC model is reliable in predicting the EC values during the motion of the robot arm.

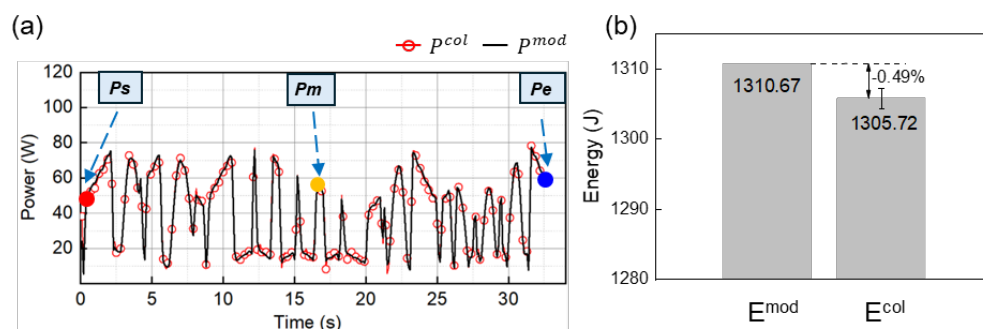


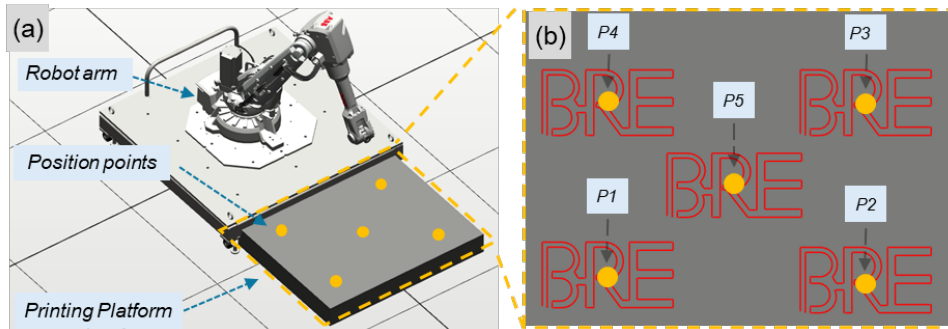
Figure 2. Results to determine model parameters. (a) comparison between the theoretical and experimental results of real-time power p^{mod} and p^{col} ; (b) comparison of the theoretical and experimental results for a printed BRE path.

3.2 Impact of position on EC behavior

As stated above, for a given trajectory, the EC value can be calculated via the proposed EC models with the determined parameter set Π . Then, the effect of printing positions on EC was evaluated. Five different positions within the working area of the robot were used in this study. Figure 3 (a) shows the printing setup and five different positions. The coordinate data of these positions are referenced to the base frame and are listed in Table 1. Figure 3 (b) shows the printing paths with different positions $P1 \sim P5$, which are located at the four corners and the central location of the working area.

Table 1. Coordinates of different positions

Points	Position (mm)
P1	(1100, 400, 0)
P2	(1100, -400, 0)
P3	(1400, -400, 0)
P4	(1400, 400, 0)
P5	(1250, 0, 0)


 Figure 3. Diagram for various positions: (a) printing setup and printing platform; (b) pre-designed path at different printing positions of $P1 \sim P5$ on the printing platform.

During the robot motion in the virtual environment, the collected joint configuration $(q, \dot{q}, \ddot{q})^{col}$ can be obtained for EC calculations. Figure 4 presents the simulated result of EC values for positions $P1 \sim P5$. The results show that the EC value in position $P4$ is the highest, 2,497.6J, comparable to the EC value in position $P3$. The EC value in position $P2$ is the lowest, 1,151.2J, comparable to the EC value in positions $P1$ and $P5$, 1,185.4J and 1,237.8J, respectively. Results reveal that the EC value can be reduced by 53.72% between the maximum ($P4$) and minimum ($P2$) values among different positions. This is mainly attributed to the different robot configurations during the printing. In practical applications, printing positions can be regarded as an optimal factor in the EC model for printing path-planning to realize energy efficiency and lower carbon emissions in the 3DCP process.

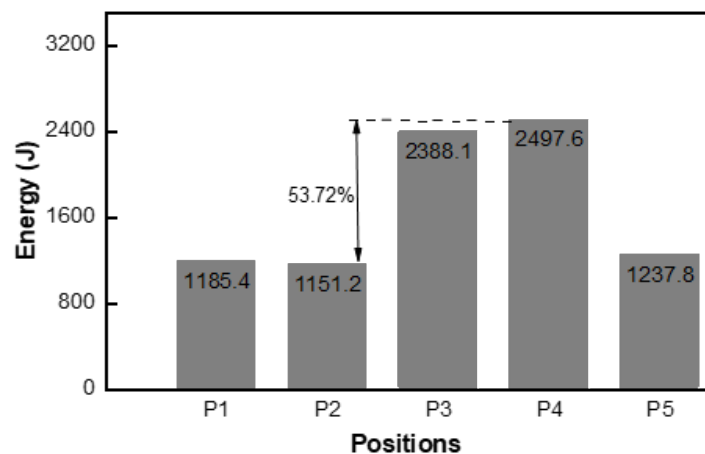


Figure 4. Simulated EC at different printing positions.

4. CONCLUSIONS

This study integrates the construction characteristics of 3D concrete printing (3DCP) with the dynamic parameters of robotic arms to develop an energy consumption model tailored for 3DCP operations. The proposed model provides energy consumption analyses for both printing positions and printing paths in 3DCP, thereby facilitating the achievement of low-carbon construction. The EC model is constructed by the determination of the dynamic parameter set Π . Then, the effect of the printing positions on EC is explored in 3DCP. Results show that a 99.51% accuracy of the modeled EC behavior can be achieved compared to that of the experimental results. The EC can be reduced by 53.72% with the optimization of the printing position for the pre-designed path. This proposed EC model for 3DCP

provides a novel design method for researchers and designers to enhance energy efficiency in 3DCP processes.

ACKNOWLEDGMENTS

The author would like to gratefully acknowledge the Project No.52308284 supported by National Natural Science Foundation of China and the Hong Kong Polytechnic University.

REFERENCES

- Cao, H., Zhou, J., Jiang, P., Hon, K. K. B., Yi, H., & Dong, C. (2020). An integrated processing energy modeling and optimization of automated robotic polishing system. *Robotics and Computer-Integrated Manufacturing*, 65, 101973.
- Chignoli, M., Adrian, N., Kim, S., & Wensing, P. M. (2023). Recursive Rigid-Body Dynamics Algorithms for Systems with Kinematic Loops. *arXiv preprint arXiv:2311.13732*.
- Feng, Y., Ji, Z., Gao, Y., Zheng, H., & Tan, J. (2021). An energy-saving optimization method for cyclic pick-and-place tasks based on flexible joint configurations. *Robotics and Computer-Integrated Manufacturing*, 67, 102037.
- Gadaleta, M., Berselli, G., Pellicciari, M., & Grassia, F. (2021). Extensive experimental investigation for the optimization of the energy consumption of a high payload industrial robot with open research dataset. *Robotics and Computer-Integrated Manufacturing*, 68, 102046.
- Kajzr, D., Myslivec, T., & Černohorský, J. (2024). Modelling, Analysis and Comparison of Robot Energy Consumption for Three-Dimensional Concrete Printing Technology. *Robotics*, 13(5), 78.
- Li, K., Zou, Z., Zhang, Y., & Shuai, C. (2024). Assessing the spatial-temporal environmental efficiency of global construction sector. *Science of The Total Environment*, 951, 175604.
- Liu, A., Liu, H., Yao, B., Xu, W., & Yang, M. (2018). Energy consumption modeling of industrial robot based on simulated power data and parameter identification. *Advances in Mechanical Engineering*, 10(5), 1687814018773852.
- Mechtcherine, V., Grafe, J., Nerella, V. N., Spaniol, E., Hertel, M., & Füßel, U. (2018). 3D-printed steel reinforcement for digital concrete construction—Manufacture, mechanical properties and bond behaviour. *Construction and Building Materials*, 179, 125-137.
- Pellicciari, M., Berselli, G., Leali, F., & Vergnano, A. (2013). A method for reducing the energy consumption of pick-and-place industrial robots. *Mechatronics*, 23(3), 326-334.
- Teixeira, F., Sarris, C., Zhang, Y., Na, D.-Y., Berenger, J.-P., Su, Y., Okoniewski, M., Chew, W., Backman, V., & Simpson, J. J. (2023). Finite-difference time-domain methods. *Nature Reviews Methods Primers*, 3(1), 75.
- Wang, L., Mohammed, A., Wang, X. V., & Schmidt, B. (2018). Energy-efficient robot applications towards sustainable manufacturing. *International Journal of Computer Integrated Manufacturing*, 31(8), 692-700.
- Weng, Y., Li, M., Ruan, S., Wong, T. N., Tan, M. J., Yeong, K. L. O., & Qian, S. (2020). Comparative economic, environmental and productivity assessment of a concrete bathroom unit fabricated through 3D printing and a precast approach. *Journal of Cleaner Production*, 261, 121245.
- Zhang, X., Li, M., Lim, J. H., Weng, Y., Tay, Y. W. D., Pham, H., & Pham, Q.-C. (2018). Large-scale 3D printing by a team of mobile robots. *Automation in Construction*, 95, 98-106.



Part 6

**Large Model and Intelligent
Building Operation and Maintenance**

ENHANCING KNOWLEDGE CAPTURE AND REUSE FOR OFFSITE CONSTRUCTION THROUGH A CONTAINER-BASED KNOWLEDGE APPROACH

Zhen Zhang¹, Yang Zou², Brian H.W. Guo³, Lixin Jiang⁴, Johannes Dimyadi⁵, and Roy Davies⁶

1) Ph.D. Student, Department of Civil and Environmental Engineering, University of Auckland, New Zealand. Email: zhang.zhen@auckland.ac.nz

2) Ph.D., Senior Lecturer, Department of Civil and Environmental Engineering, University of Auckland, New Zealand. Email: yang.zou@auckland.ac.nz

3) Ph.D., Senior Lecturer, Department of Civil and Natural Resources Engineering, University of Canterbury, New Zealand. Email: brian.guo@canterbury.ac.nz

4) Ph.D., Assoc. Prof., School of Psychology, University of Auckland, New Zealand. Email: l.jiang@auckland.ac.nz

5) Ph.D., CEO, Codify Asset Solutions (CAS) Ltd, New Zealand. Email: jdimyadi@cas.net.nz

6) Ph.D., Senior Technician, Department of Civil and Environmental Engineering, University of Auckland, New Zealand. Email: roy.davies@auckland.ac.nz

Abstract: Offsite construction (OSC) requires an integrated design and delivery system. Reusing prior OSC knowledge is paramount to the success of new OSC projects. However, a gap remains in managing and reusing this knowledge across the broader industry. The challenge lies in the fragmented nature of project-based organisations with isolated knowledge systems, which often lack integration and advanced capabilities for knowledge-based collaboration. This paper proposes a novel framework that leverages cutting-edge knowledge-based methods and artificial intelligence (AI) technologies to create a container-based knowledge system (CBKS) that can enhance knowledge capture and reuse towards collaborative OSC. Specifically, the semantic web stack is adopted to construct multimodal knowledge containers for both human users and AI agents. In addition, GPT-4o, a large language model (LLM), is embedded into the knowledge system for better knowledge querying, matching and retrieving. By using this framework, constructed modular knowledge units can integrate product, process and organisation factors to address specific OSC problems. To evaluate the technical feasibility of the proposed framework, a prototype is developed and illustrated through a modular connection design. This illustrative case study demonstrates how knowledge is captured for product representation under manufacturing and assembly constraints, enabling its reuse in different projects. Moreover, the usefulness of GPT-4o for enhancing this process is also tested.

Keywords: Offsite construction, Integrated knowledge base, Information container, Large language model.

1. INTRODUCTION

Offsite Construction (OSC) refers to the manufacturing of building components away from the site, followed by their transportation to and assembly at the final location. In recent years, OSC has gained significant attention as a solution to address inefficiencies in productivity, affordability, and sustainability within the traditional construction industry (Wuni & Shen, 2019). OSC offers numerous benefits, including reduced construction time, improved quality control, and minimised on-site disruptions (Razkenari et al., 2020). The adoption of OSC relies on an integrated design and delivery system, where clients, designers, manufacturers and contractors need to collaboratively work together at the early stage of a project to reach the goals (Hosseini et al., 2018). Therefore, the reuse of prior knowledge is crucial for the success of new projects. However, there remains a gap in effectively managing and reusing OSC knowledge across the broader industry. The challenge lies in the fragmented nature of project-based organisations with isolated knowledge systems, which often lack integration and advanced capabilities for knowledge-based collaboration (Z. Zhang et al., 2024).

To address these challenges, effective Knowledge Management (KM) is essential for OSC. Managing OSC knowledge can facilitate the sharing of best practices, enhance process efficiency, and promote continuous improvement across projects. However, previous research on KM in OSC does not fully integrate technology, process, and people-related factors that are critical to successful knowledge capture and reuse (Z. Zhang et al., 2024). It is argued that this gap can be bridged by developing an integrated knowledge system for combining technological tools, process-oriented structures, and the necessary human factors so that OSC knowledge can be effectively captured, shared and reused. For this purpose, a novel approach leveraging the power of semantic web technology and artificial intelligence (AI) is proposed.

Firstly, involving the semantic web technology increases the level of integration and flexibility of KM in a wide range of OSC project lifecycles. The semantic web has increasingly attracted interest in the construction industry (Pauwels et al., 2017). It includes technologies such as Resource Description Framework (RDF), Web Ontology Language (OWL), and SPARQL. RDF provides a framework for representing information in a graph format, enabling data interoperability. OWL is used for defining complex relationships between concepts and facilitating reasoning over the knowledge base. SPARQL is a query language used to retrieve and manipulate data stored in RDF format, allowing efficient knowledge querying and integration. More importantly, knowledge in the

real-world construction project is usually used in multimodal formats, such as texts, 2D drawings, 3D-models, etc., with highly dynamic requirements in different scenarios (Bilal et al., 2016), whereas current knowledge approaches for OSC mainly focus on specific data format (Z. Zhang et al., 2024). To flexibly integrate the multimodal knowledge for specific usage, a container-based approach can be adopted according to the international standard series ISO 21597 to link various data and documents together using ontologies and links datasets (ISO, 2020a; 2020b). Though semantic web technology offers significant advantages for representing and linking knowledge in a structured manner, the application it in KM for OSC has not been widely adopted in the OSC industry, leaving a gap in understanding its full potential in this domain. The reasons can include high cost of ontology development, the complexity of SPARQL queries, and the need to develop rule sets.

Additionally, large language model (LLM) is embedded in this framework to promote the management of OSC knowledge. The construction industry has witnessed the emergence of AI technologies, especially LLM, which show promising potential for enhancing KM in OSC. LLMs are AI models trained on vast amounts of text data, capable of understanding, generating, and translating human language. They leverage deep learning to comprehend complex patterns and semantics, making them powerful tools for a wide range of natural language tasks, such as automated construction reporting (Pu et al., 2024). However, challenges like ensuring accuracy, improving interpretability and maintaining context relevance in the construction domain persist (Abioye et al., 2021). This is where the synergy between LLMs and semantic web technology becomes crucial (Pan et al., 2024). For example, ontologies, which can formally represent domain-specific knowledge, offer a solution to enhance the performance of LLMs, ensuring accurate and context-sensitive knowledge retrieval. Furthermore, the semantic web can serve as a foundation for creating interconnected knowledge containers, enabling LLMs to deliver more accurate and informed responses. These knowledge containers enable LLMs to understand relationships between different OSC elements, such as components, processes, and stakeholders. By leveraging these interconnected datasets, LLMs can generate more precise and insightful responses to user queries, supporting decision-making processes in OSC (Wu et al., 2023). In this way, the knowledge hidden in previous OSC cases can be derived into knowledge containers for effective reuse in new projects enhanced by LLMs.

To explore the mechanism and feasibility of this approach, this study developed a theoretical framework of a container-based knowledge system and developed a system prototype that combines semantic web technology and LLM with effective human-computer interaction to facilitate knowledge capture, reuse, and management throughout the OSC lifecycle. The feasibility of this approach is illustrated through a case study, showcasing its potential to improve KM processes in OSC projects. In the next section, we introduce the method adopted in this study and outline the research questions. Next, this paper details the design and development of the proposed solution. After that, the system test is illustrated through a case study on modular building connections in Section 3. Finally, the discussion and conclusion sections evaluate and summarize the research findings.

2. DESIGN AND DEVELOPMENT OF CBKS

2.1 Research method and questions

A design science research methodology has been applied for developing and evaluating the prototype of proposed framework (Peffer et al., 2007). Firstly, the research problems and motivation were identified through literature review and our previous study (Z. Zhang et al., 2024). It was found that the study of a container-based knowledge system (CBKS) has been ignored. In addition, no research has explored the collaboration between CBKS and LLM, particularly how LLM can improve KM for OSC. To narrow down the scope, this study mainly focuses on knowledge capture and reuse enhanced by information containers and LLM. The following research questions (RQs) are investigated:

RQ1: How can a container-based knowledge system (CBKS) enhance knowledge capture and reuse for OSC?

RQ2: What are the potential ways of the proposed CBKS collaborating with LLM to facilitate OSC knowledge querying, manipulating and retrieving?

Therefore, the objectives were defined for developing the corresponding solution, namely, LLM-augmented CBKS. Then, this study conducted an iterative design and development process followed by a case study to illustrate and test the feasibility of the proposed CBKS. Finally, the test outcomes were discussed for directing future research.

2.2 The theoretical framework

The theoretical framework of CBKS uses a problem-solving-oriented method to make sure each container is a best practice guide for achieving its value (Barak & Goffer, 2002). The OSC knowledge in this framework is modelled at two layers, namely, the universal layer and the project-specific layer. The rationality of this method comes from the modular nature of OSC as the knowledge that can be modularised and standardised should be stored at a universal layer for maximum reusability on the top, whereas the knowledge that can only be used in the specific project should be treated case by case at the bottom. There are more opportunities for leveraging the value of universal knowledge in OSC projects. Moreover, knowledge modelling based on the semantic web requires

high expertise (Davies et al., 2006). Therefore, KM professionals are necessary to maintain knowledge from the backend so that the CBKS can provide knowledge services to the frontend users. Meanwhile, embedded LLM is expected to promote this system by collaborating with the constructed knowledge container (Pan et al., 2024).

As shown in Figure 1, the knowledge container is configured by knowledge managers through a set of Terminology Box (T-Box), Assertion Box (A-box), and Rule Box (R-box). T-Box stands for the taxonomy of the concepts and their attributes and relations in the scope of the knowledge container. Based on the defined T-Box, instance facts can be asserted to form A-Box manually or through the execution of SPARQL code according to the specific cases and requirements. Then, classes and assertion facts can work with predefined rules in the R-Box to infer new facts, returning the expected outcomes to a knowledge user (L. Zhang & Lobov, 2024). The knowledge user can also interact with the CBKS through a user interface to assert case-specific facts that will be processed by the reasoning engines (R-Box) for knowledge reuse (Pérez et al., 2009). In addition, there is an opportunity to facilitate the transformation of required queries between natural language and SPARQL code by using GPT-4o as it is one of the most powerful LLMs developed by OpenAI (2024). This can potentially improve knowledge sharing and reuse because most of the industrial professionals (knowledge users) may not be familiar with semantic web technologies.

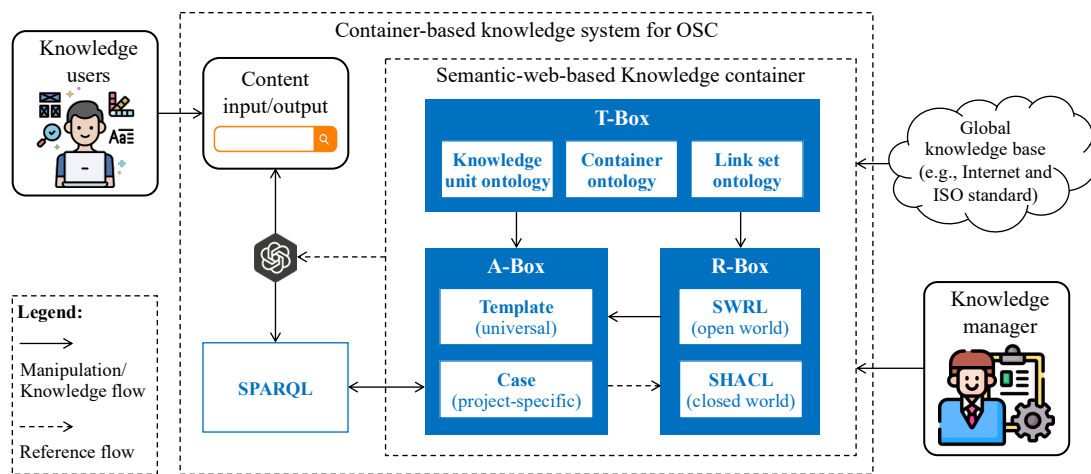


Figure 1. The theoretical framework of LLM-augmented CBKS

2.3 Semantic modelling of data and rules

The core of the proposed CBKS is the semantic modelling of data and rules. The processes of building the knowledge base (KB) include the construction of the OSC knowledge model, container model, and rules in T-Box, A-Box, and R-Box.

There are three ontology resources in the T-Box: OSC knowledge ontology, container ontology and link set ontology. The OSC knowledge ontology aims to present domain knowledge for solving specific problems, such as OSC product design, quantity take-off, project scheduling, etc. The role of container ontology and link set ontology is to govern the actual data saved as document format and link these documents and their internal elements together (Hagedorn, Liu, et al., 2023). This is designed to present case-based experiential knowledge generated in previous OSC projects.

The A-Box contains instantiated data models presenting templates that can be used universally in different scenarios and the cases that have occurred in previous projects. After defining the classes, attributes and relations in T-Box, knowledge managers can start the construction of A-Box. At this stage, both the universal assertions in templates and case-specific facts can be input manually for debugging purposes. After the validation of semantic rules, some case-specific assertions can be derived by the reasoning engine embedding in the proposed CBKS. In addition, external documents can also be linked to the knowledge model through the link dataset asserted in the A-Box for providing case-specific references, which has been proved in many other studies (Hagedorn, Liu, et al., 2023; Hölting et al., 2021; Liu et al., 2023).

In the R-Box, semantic rules are formed to provide a reasoning role that can conduct compliance checks and generate new facts according to assertions in the A-Box. Shap Constraint Language (SHACL) and Semantic Web Rule Language (SWRL) are two representative types of semantic rules. SHACL defines rules in the form of information availability requirements, while SWRL defines rules in the form of an implication between a set of antecedents and consequents (Nuyts et al., 2024). They stand for two different knowledge representation patterns, which will be described further in Section 2.4.

2.4 Mechanism of knowledge sharing and reuse

For knowledge sharing and reuse in OSC projects, there are opportunities to utilise reusable ontologies defined globally, for example, the standardised product representation defined in the international or national standards (L. Zhang & Lobov, 2024). In this approach, the container ontology and link set ontology resources are directly reused from the international standard ISO 21597 about information container for linked document delivery (ISO, 2020a; 2020b). For a specific knowledge domain, knowledge managers can develop customised local ontologies and combine them with global ones to form the basis of T-Box. When facing a specific problem, knowledge users can assess the reusability of the components in global and local ontologies and directly create an instance of them to reuse predefined concepts, attributes and relations.

These reusable knowledge contents can be universally applied to different projects sharing the same scope. However, it is inevitable for projects to have their characters and unique environments, which means that project-specific data and rules must be taken into consideration. Hence, the proposed CBKS applies this kind of knowledge directly to the instances in A-Box rather than the classes in T-Box. In this way, universal knowledge and project-based knowledge can be modelled and stored in the template library and case library of the specially designed KB respectively.

Theoretically, this CBKS framework can be driven by various reasoning engines with corresponding rule languages. These rule languages generally fall into two categories of knowledge-representing mechanisms: *Closed-World Assumption (CWA)* and *Open-World Assumption (OWA)*. Under a *CWA*, all entities and relationships in the KB cannot be changed but discovered (Reiter, 1981), whereas unseen entities, relationships and their attributes can be predicted and reasoned in an *OWA* (Shi & Weninger, 2018). As mentioned in Section 3.2, there are two representative rule languages in the semantic web world: SHACL and SWRL, in which SHACL naturally operates under a *CWA* while SWRL originally operates under an *OWA*. It is noticed that, however, the current research on SWRL in the construction field mainly applies it under the *CWA* patterns. In the present work, it is argued that utilising the *OWA* nature of SWRL can facilitate knowledge sharing and reuse in the OSC processes as universal knowledge and case-specific knowledge are used in hybrid in the real-world OSC projects.

2.5 User interactions with CBKS using LLM

There are two key roles: knowledge managers and knowledge users who are responsible for building and using this proposed CBKS. The overview of the general work processes and data flows for applying CBKS is shown in Figure 2 in the form of Business Process Model and Notation (BPMN) (Chinosi & Trombetta, 2012). The diagram contains three pools representing the key players responsible for knowledge use and management to interact with CBKS.

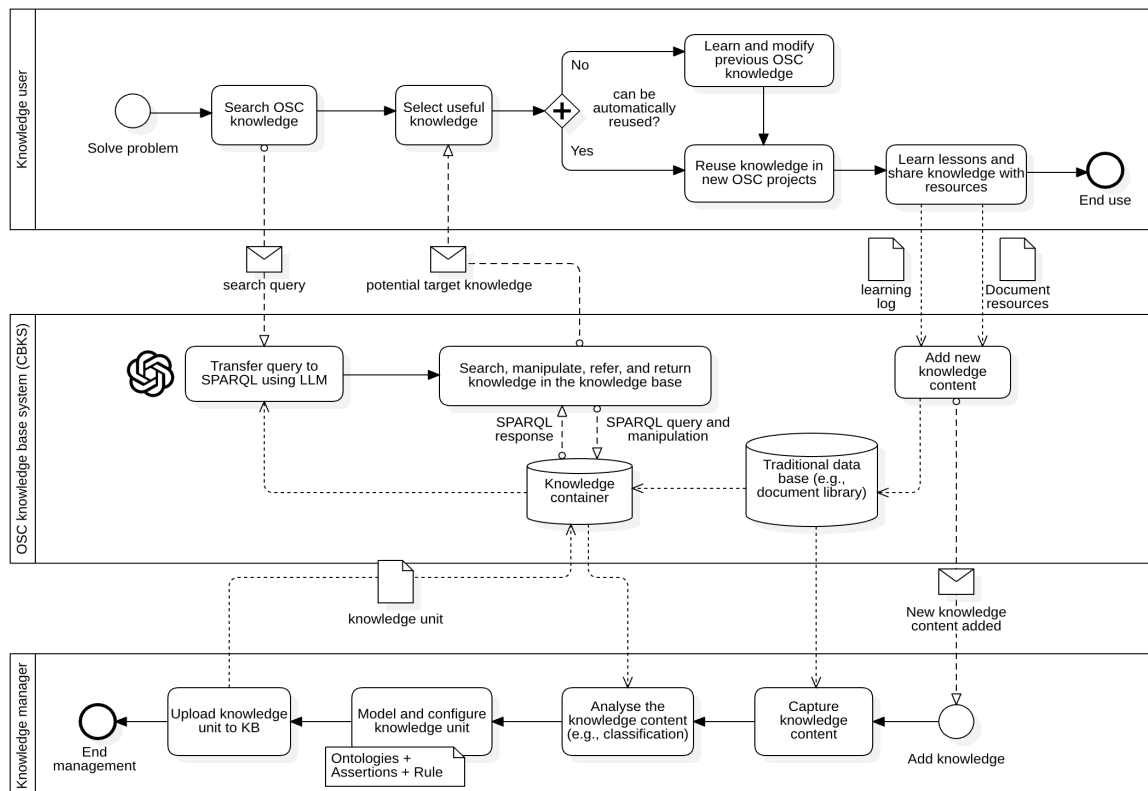


Figure 2. BPMN diagram of the application of CBKS

The CBKS enables knowledge users to search for knowledge and manipulate case-specific facts using SPARQL code. However, most professionals from the construction industry may not be familiar with writing the SPARQL query and update (Pérez et al., 2009). To decrease this barrier, the proposed CBKS leverages the power of LLM to conduct a mutual transformation of the professional content between natural language and SPARQL code. By providing proper prompts and the context of knowledge models, effective knowledge query and manipulation and high-quality responses can be expected from the LLM-augmented CBKS. Then, the knowledge users can select useful knowledge responses to directly use in their problem-solving process or make any necessary modifications. After that, new lessons may be learned and shared with partners and retained in the KB.

Universally applicable and case-based knowledge needs to be managed and integrated into the KB. In this process, knowledge managers first capture knowledge from a global knowledge base and a traditional database in CBKS. Then, the knowledge content will be analysed and formalised to construct knowledge units in the form of ontologies, assertions and rules. Subsequently, these knowledge units are uploaded to the CBKS for utilisation.

3. ILLUSTRATION OF THE CBKS FRAMEWORK

This section illustrates the development and test of the core functions provided by CBKS through a case study, including semantic-based knowledge inference and LLM-embedded knowledge manipulation and query. On the one hand, how OSC knowledge is modelled, stored, and inferred with restrictions of manufacturing and assembly to drive information delivery in project documents is described in detail (RQ1). On the other hand, a series of tests are conducted to explore the potential ways of, and prove the feasibility of, embedding LLM into the proposed CBKS (RQ2).

3.1 Semantic web-based knowledge inference

A design case of modular building connection is adopted for study because the connection is more likely to be universally standardised but still affected by case-based factors in OSC projects. Figure 3 demonstrates the mechanism of CBKS performing knowledge representation and reasoning of the selected modular building connection design (Chen et al., 2017). This case consists of four main sections including a set of stay bolts, cover plates, intermediate plates, and a plug-in device, which are assembled with modular space modules.

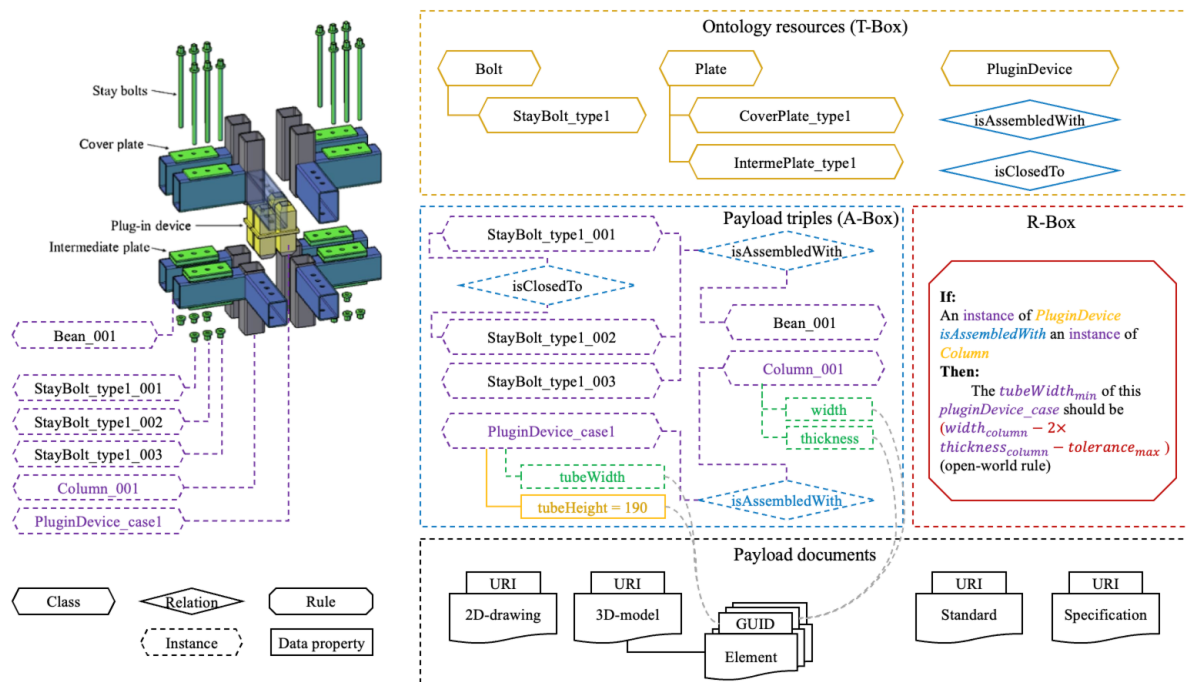


Figure 3. Case study for the CBKS

The semantic definition of the modular connection is achieved through classes and relations in T-Box, and its configuration is realised by asserting corresponding instances in A-Box. The first step, for example, is defining classes “Bolt”, “Plate”, “PluginDevice”, and relations “isAssembledWith” and “isClosedTo”. Then their instances are modelled in an object-oriented manner, presenting the actual solution of this modular connection. These instances can directly inherit the attributes defined in T-Box that are universally applicable for different projects, while the attributes depending on the actual project circumstances are derived from running of the predefined rules.

By way of illustration, the tube of the plug-in device needs to be inserted into the steel columns of related module frameworks for building assembly. The width and thickness of the modular steel columns are determined case by case as there are many uncontrollable effects from different projects. Consequently, external documents and their internal elements can be linked to the related instances in A-Box through identifiers, such as Universal Resource Identifier (URI) and Globally Unique Identifier (GUID), in which the inferred outcomes can drive the generation and adaptation of the business objects in external documents like a plug-in device in BIM model. It should be noticed that this deep link mechanism has been proved in (Hagedorn, Pauwels, et al., 2023), whereas the generation and adaptation process still need to be explored in the future.

This paper only focuses on the test of knowledge inferring part as it is the semantic foundation of up-layer applications such as generative OSC product design. For validating this scenario, Protégé, an open-source ontology editor, is adopted to implement the test. As shown in Figure 4, when the classes of the test connection case are defined and enough facts have been asserted, the predefined rule can derive implicit facts from these universal restrictions and case-based assertions.

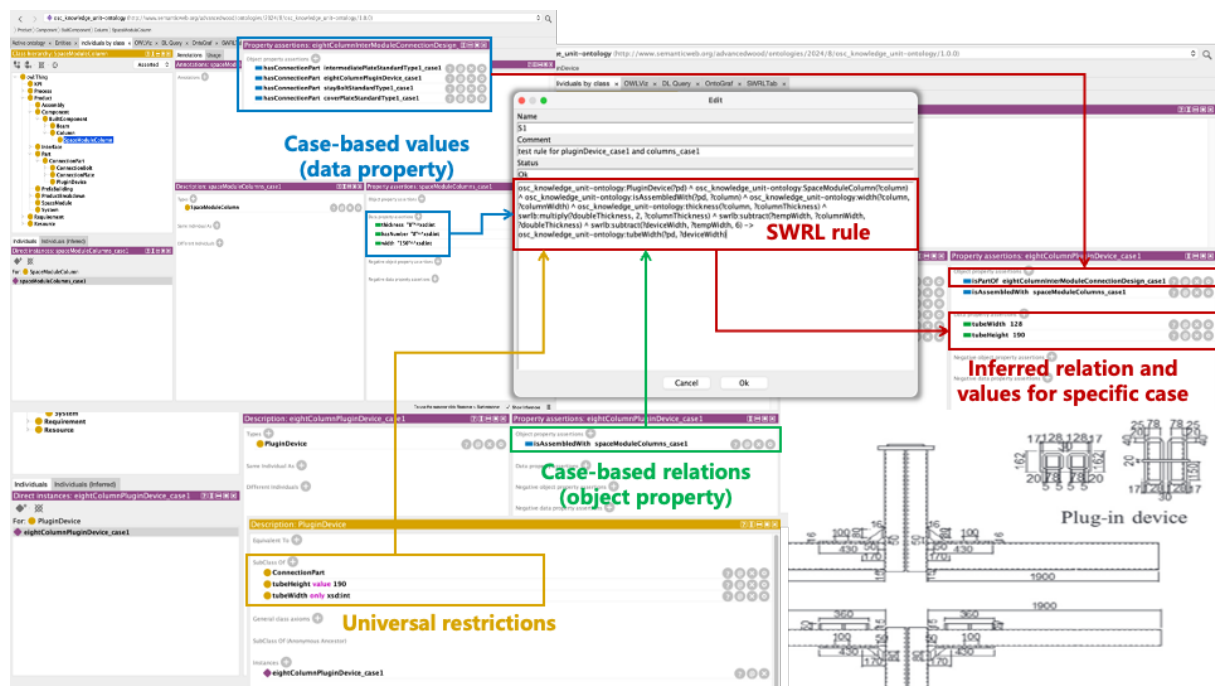


Figure 4. Knowledge reasoning of the CBKS

In this case, the tube width of the selected plug-in device is affected by the width and thickness of the modular columns directly assembled with it. The sizes of modular columns will inevitably change case by case in different projects. Under this CBKS framework with *OWA*, knowledge can be easily shared and reused by users asserting case-based facts from new conditions. However, for traditional industry professionals who are not familiar with semantic web technologies, it is hard to interact with this new knowledge base. Therefore, this study attempted to embed LLM to facilitate knowledge manipulation and query by transferring natural language to SPARQL code.

3.2 LLM-embedded knowledge manipulation and query

This section demonstrates the potential opportunity to embed LLM for easier knowledge manipulation and query using GPT-4o. For knowledge manipulation, a prompt was formed as a template to insert a new instance for class “SpaceModuleColumn” named “spaceModduleColumn_case2” and adapt its “width” and “thickness” with two new values. Figure 5 shows one of the test results. It can be noticed that GPT-4o can provide acceptable responses followed by a set of interpretations as the contextual information of the given ontology is provided in the template prompts.

like “What are key components of a type of modular connection according to the existing cases?”. In the given knowledge model, as mentioned in Section 4.1, an “EightColumnInterModuleConnection” was defined and has a configuration type assembled with four main components: “EightColumnPluginDevice”, “CoverPlate”, “IntermediatePlate” and “StayBolt”. This was asked to assess the response accuracy of GPT-4o. The tests were run ten times. Though GPT-4o sometimes used different methods to parse the given knowledge at the backend, it could always return the same and accurate answer in each time (see Figure 7).

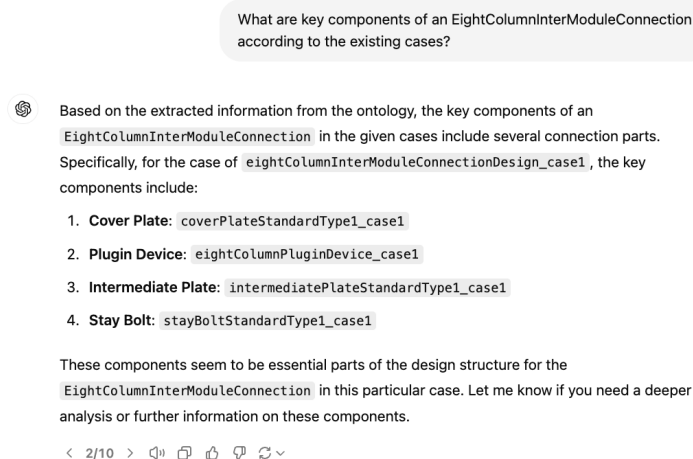


Figure 7. Augmentation of LLM in a specific knowledge problem

4. DISCUSSION

The contribution of this study lies in proposing an innovative knowledge system, CBKS, and proving its feasibility of leveraging the capability of semantic web-based containers and LLMs for enhancing the management of OSC knowledge. Through the proposed CBKS, the universal and case-specific knowledge can be successfully captured and reused by modelling them into global and local ontologies and rules. The semantic web stack combining ontologies of OSC knowledge, information container and link set can provide linked multimodal data for satisfying dynamic knowledge demands ranging from documents such as a BIM model or a CAD drawing to their internal elements (e.g., beam, wall, connection, etc.). For the rule construction, the case study adopted SWRL to conduct the knowledge inferring under an *OWA* instead of adopting it under *CWA*, which can facilitate knowledge reuse in dynamic project conditions.

Additionally, LLM such as GPT-4o can facilitate the knowledge manipulation and query through transferring natural language to SPARQL code and parsing the ontology-based knowledge model to answer professional questions. In the test of understanding and answering questions, GPT-4o showed a stable performance for returning correct responses derived from the given knowledge model. However, more tests are needed for different scenarios and questions. It should also be noted that prompt templates are needed for better generation in the transformation from natural language to the SPARQL code, which leads to lower flexibility in the application.

Considering this paper serves as a proof-of-concept to propose a knowledge system framework for enhancing knowledge capture and reuse for OSC, the knowledge model in the case study was simplified and a limited number of scenarios were tested. Moreover, the effectiveness of deploying the proposed framework was not evaluated, and this will be addressed by our future research.

5. CONCLUSION

This paper introduces a knowledge base framework called CBKS that utilises semantic web and LLM technologies to enhance the knowledge capture and reuse for OSC. This framework is illustrated through a case study involving the representation of an OSC product (eight-column modular building connection). The findings are summarised as follows: 1) The semantic web stack as a core can link external documents and elements to the OSC objects driven by knowledge in CBKS; 2) Compared to SHACL, SWRL can support more flexible knowledge use by capturing open-world rules rather than asserting all constraints in a fixed pattern; 3) LLM, especially GPT-4o, has significant potential to augment the knowledge manipulation and query based on the given ontology model. In future research, it is essential to apply the proposed CBKS to a well-defined OSC problem and connect it to the application layer to evaluate the knowledge-driven automation processes, such as the generative design of an OSC product. Furthermore, the completed system will be deployed to an open server for more widely and systematic evaluation by industry experts.

ACKNOWLEDGEMENT

This research is sponsored by BRANZ (Building Research Association of New Zealand) from the New Zealand Building Research Levy (Ref. LR16302). The authors would express their gratitude to Laura Tammaro and Narrel Brogan, Research Advisors at BRANZ, for their continuous support and guidance.

REFERENCES

- Abioye, S. O., Oyedele, L. O., Akanbi, L., Ajayi, A., Davila Delgado, J. M., Bilal, M., Akinade, O. O., & Ahmed, A. (2021). Artificial intelligence in the construction industry: A review of present status, opportunities and future challenges. *Journal of Building Engineering*, 44, 103299. <https://doi.org/10.1016/j.jobbe.2021.103299>
- Barak, M., & Goffer, N. (2002). Fostering Systematic Innovative Thinking and Problem Solving: Lessons Education Can Learn From Industry. *International Journal of Technology and Design Education*, 12(3), 227–247. <https://doi.org/10.1023/A:1020259623483>
- Bilal, M., Oyedele, L. O., Qadir, J., Munir, K., Ajayi, S. O., Akinade, O. O., Owolabi, H. A., Alaka, H. A., & Pasha, M. (2016). Big Data in the construction industry: A review of present status, opportunities, and future trends. *Advanced Engineering Informatics*, 30(3), 500–521. <https://doi.org/10.1016/j.aei.2016.07.001>
- Chen, Z., Liu, J., & Yu, Y. (2017). Experimental study on interior connections in modular steel buildings. *Engineering Structures*, 147, 625–638. <https://doi.org/10.1016/j.engstruct.2017.06.002>
- Chinosi, M., & Trombetta, A. (2012). BPMN: An introduction to the standard. *Computer Standards & Interfaces*, 34(1), 124–134. <https://doi.org/10.1016/j.csi.2011.06.002>
- Davies, J., Studer, R., & Warren, P. (2006). *Semantic Web Technologies: Trends and Research in Ontology-based Systems*. John Wiley & Sons.
- Hagedorn, P., Liu, L., König, M., Hajdin, R., Blumenfeld, T., Stöckner, M., Billmaier, M., Grossauer, K., & Gavin, K. (2023). BIM-Enabled Infrastructure Asset Management Using Information Containers and Semantic Web. *Journal of Computing in Civil Engineering*, 37(1), 04022041. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0001051](https://doi.org/10.1061/(ASCE)CP.1943-5487.0001051)
- Hagedorn, P., Pauwels, P., & König, M. (2023). Semantic rule checking of cross-domain building data in information containers for linked document delivery using the shapes constraint language. *Automation in Construction*, 156, 105106. <https://doi.org/10.1016/j.autcon.2023.105106>
- Höltgen, L., Cleve, F., & Hagedorn, P. (2021). Implementation of an Open Web Interface for the Container-based Exchange of Linked Building Data. *Proceedings of the 32 Forum Bauinformatik*. https://www.researchgate.net/profile/Philipp-Hagedorn/publication/354464413_Implementation_of_an_Open_Web_Interface_for_the_Container-based_Exchange_of_Linked_Building_Data/links/613a14f1d1bbee063c5c887b/Implementation-of-an-Open-Web-Interface-for-the-Container-based-Exchange-of-Linked-Building-Data.pdf
- Hosseini, M. R., Martek, I., Zavadskas, E. K., Aibinu, A. A., Arashpour, M., & Chileshe, N. (2018). Critical evaluation of off-site construction research: A Scientometric analysis. *Automation in Construction*, 87, 235–247. <https://doi.org/10.1016/j.autcon.2017.12.002>
- ISO. (2020a). *ISO 21597-1:2020 Information container for linked document delivery—Exchange specification—Part 1: Container*. <https://www.iso.org/standard/74389.html>
- ISO. (2020b). *ISO 21597-2:2020 Information container for linked document delivery—Exchange specification—Part 2: Link types*. <https://www.iso.org/standard/74390.html>
- Liu, L., Hagedorn, P., & König, M. (2023). *Definition of a container-based machine-readable IDM integrating level of information needs*. 4, 0–0. <https://doi.org/10.35490/EC3.2023.221>
- Nuyts, E., Bonduel, M., & Verstraeten, R. (2024). Comparative analysis of approaches for automated compliance checking of construction data. *Advanced Engineering Informatics*, 60, 102443. <https://doi.org/10.1016/j.aei.2024.102443>
- OpenAI. (2024). <https://openai.com/>
- Pan, S., Luo, L., Wang, Y., Chen, C., Wang, J., & Wu, X. (2024). Unifying Large Language Models and Knowledge Graphs: A Roadmap. *IEEE Transactions on Knowledge and Data Engineering*, 36(7), 3580–3599. <https://doi.org/10.1109/TKDE.2024.3352100>
- Pauwels, P., Zhang, S., & Lee, Y.-C. (2017). Semantic web technologies in AEC industry: A literature overview. *Automation in Construction*, 73, 145–165. <https://doi.org/10.1016/j.autcon.2016.10.003>
- Peffer, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, 24(3), 45–77. <https://doi.org/10.2753/MIS0742-1222240302>
- Pérez, J., Arenas, M., & Gutiérrez, C. (2009). Semantics and complexity of SPARQL. *ACM Trans. Database Syst.*, 34(3), 16:1–16:45. <https://doi.org/10.1145/1567274.1567278>
- Pu, H., Yang, X., Li, J., & Guo, R. (2024). AutoRepo: A general framework for multimodal LLM-based automated construction reporting. *Expert Systems with Applications*, 255, 124601.

- <https://doi.org/10.1016/j.eswa.2024.124601>
- Razkenari, M., Fenner, A., Shojaei, A., Hakim, H., & Kibert, C. (2020). Perceptions of offsite construction in the United States: An investigation of current practices. *Journal of Building Engineering*, 29, 101138. <https://doi.org/10.1016/j.jobbe.2019.101138>
- Reiter, R. (1981). ON CLOSED WORLD DATA BASES. In B. L. Webber & N. J. Nilsson (Eds.), *Readings in Artificial Intelligence* (pp. 119–140). Morgan Kaufmann. <https://doi.org/10.1016/B978-0-934613-03-3.50014-3>
- Shi, B., & Weninger, T. (2018). Open-World Knowledge Graph Completion. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1), Article 1. <https://doi.org/10.1609/aaai.v32i1.11535>
- Wu, Y., Zhou, S., Liu, Y., Lu, W., Liu, X., Zhang, Y., Sun, C., Wu, F., & Kuang, K. (2023). *Precedent-Enhanced Legal Judgment Prediction with LLM and Domain-Model Collaboration* (arXiv:2310.09241). arXiv. <https://doi.org/10.48550/arXiv.2310.09241>
- Wuni, I. Y., & Shen, G. Q. P. (2019). Holistic Review and Conceptual Framework for the Drivers of Offsite Construction: A Total Interpretive Structural Modelling Approach. *Buildings*, 9(5), Article 5. <https://doi.org/10.3390/buildings9050117>
- Zhang, L., & Lobov, A. (2024). Semantic Web Rule Language-based approach for implementing Knowledge-Based Engineering systems. *Advanced Engineering Informatics*, 62, 102587. <https://doi.org/10.1016/j.aei.2024.102587>
- Zhang, Z., Zou, Y., Guo, B. H. W., Dimyadi, J., Davies, R., & Jiang, L. (2024). Knowledge management for off-site construction. *Automation in Construction*, 166, 105632. <https://doi.org/10.1016/j.autcon.2024.105632>

Automated text classification of construction inspection report: A small samples training approach

Kai Li¹, Chao Dong², Xueqing Fang³, and Da Li⁴

1) Master Student, Department of Civil Engineering and Architecture, Wuhan University of Technology, Wuhan, Hubei, China. Email: 2664361126@qq.com

2) Ph.D., Assoc. Prof., Department of Civil Engineering and Architecture, Wuhan University of Technology, Wuhan, Hubei, China. Email: chaodong@whut.edu.cn

3) Master Student, Department of Civil Engineering and Architecture, Wuhan University of Technology, Wuhan, Hubei, China. Email: 2545711470@qq.com

4) Ph.D., Assoc. Prof., School of Civil Engineering, Henan University of Science and Technology, Zhengzhou, Henan, China. Email: 9905639@haust.edu.cn

Abstract: Risk management is crucial for construction safety, but safety risk assessment often relies on experts' knowledge, which makes automatic risk management in engineering projects still a big challenge. Fortunately, for large-scale infrastructure construction, on-site inspection is required, and the conditions on-site are recorded in text format, which provides an opportunity to learn risk information from inspection reports. To improve document processing efficiency, automatic text classification plays an important role. However, currently, automatic text classification requires large scale training datasets. It is a big challenge for the engineering industry, especially for the fields which heavily rely on the experts' knowledge, such as risk assessment. Limited data sources, high time and labor costs make it not practical to establish a large-scale dataset. This work proposes a BERT-based ensemble model for small-sample text classification, leveraging the Focal loss function to address data imbalance issues. Concurrently, an ensemble strategy is employed to enhance the model's generalization capabilities, while the learning rate gradient descent method is applied to mitigate the risk of model overfitting. The efficacy of the proposed framework is validated through a four-classification task about identifying risk levels based on the inspection reports of a metro construction project. The BERT-based ensemble model proposed in this paper achieves an accuracy of 96.24% on the test set, surpassing other pre-trained classification models and excelling in automated text classification tasks.

Keywords: Small sample training, Multi-label text classification, Construction risk management, BERT

1. INTRODUCTION

Construction is an inherently complex process characterized by long construction cycle, unique on-site conditions, sophisticated construction techniques, and significant environmental dependency. Compared to other industries, accidents happen frequently in the construction sector, leading to casualties and property damage. In 2017, the U.S. Bureau of Labor Statistics reported over 950 fatal injuries and more than 200,000 non-fatal injuries in the construction industry, accounting for more than 21% of all occupational fatalities in the United States (BLS&OSHE, 2018). To mitigate the harm caused by accidents, implementing safety risk management during construction process is crucial.

The primary objective of safety risk management is to assess the safety risk status of construction projects, which helps allocating appropriate resources to mitigate risk. Currently, safety risk assessment in construction industry still heavily relies on experts' experience. For example, for the risk assessment during foundation pits excavation, safety management engineers conduct on-site exploration daily, inspecting the on-site conditions (support structures, surrounding environment, and monitoring facilities etc.). Then, based on the exploration results and monitoring data, a construction site inspection report is created. Finally, experienced engineers make a comprehensive assessment of the construction safety risk based on the inspection report. This process is time-consuming and heavily relies on the engineers' expertise. The construction industry is labor-intensive, the shortage of experienced engineers, and the lengthy transmission time of risk information further hampers timely response to risks. Therefore, an automated method for safety risk assessment based on inspection reports needs to be developed to enable timely risk response.

The rapid advancement of artificial intelligence provides new possibilities for construction safety risk management, with Natural Language Processing (NLP) technologies demonstrating high accuracy and applicability in processing textual data (Kim & Chi, 2019). To address the limitations of manual risk assessment, automated methods for safety risk assessment in construction projects have begun to emerge. Automated safety risk assessment based on construction inspection is a text classification problem. Currently, the most popular automated text classification methods can be classified into Shallow Machine Learning models and Deep Learning models. Abderrahim Zermene adopted Random Forest model to classify the causes of falls from height (Zermene & Mohd Tohir, 2023). Fan Zhang et al. proposed an approach integrating multiple machine learning models to classify the causes of construction accident reports (Zhang et al., 2019). However, these algorithms have limited learning capabilities, require manually provided features, and exhibit high error rates. In contrast, deep learning

algorithms can identify features automatically. Zhang et al. enhanced a CNN model by introducing multi-channel input to classify unstructured construction quality records (Zhang et al., 2022). Similarly, Baker et al. employed a CNN and a hierarchical attention network (HAN), incorporating RNNs, to automatically extract accident precursors from construction accident report datasets (Baker et al., 2020). Even many attempts have been made in text mining in the construction industry, several challenges remain: (1) Semi-structured or unstructured natural language reports make it difficult for deep learning models to achieve a high accuracy; (2) Limited data sources and the high cost of manual labeling makes it a big challenge to develop large-scale domain-specific datasets in construction industry; (3) Even construction domain is high risk, severe risk incidents are still rare, leading to data imbalance issues. Given these circumstances, the methods mentioned earlier struggle to handle text classification for imbalanced small-sample datasets in the construction domain.

In recent years, the emergence of BERT pre-trained models has offered a promising solution for addressing challenges associated with small datasets. As a pre-trained model, BERT requires minimal domain-specific data for fine-tuning to achieve the desired results. Thus, this paper proposes a BERT-based ensemble model. First, data augmentation techniques are employed to enhance sample diversity. Then, leveraging BERT's powerful text processing capabilities, features are extracted from construction inspection reports, followed by fine-tuning multiple models. The final prediction result is determined by aggregating the predictions from these models. Furthermore, to address the data imbalance issues inherent in small-sample datasets, the Focal Loss function is employed to mitigate the impact of class imbalance. Ultimately, an automated warning level identification mechanism for construction inspection reports is developed, assisting project managers conducting timely risk control.

2. LITERATURE REVIEW

2.1 Application of text classification in construction safety management

Nowadays, there are many studies on automated text classification in the construction field. Typical text classification methods can be divided into knowledge-based methods and machine learning-based methods. Ontology is considered one of the most common knowledge-based approaches. Seokho Chi et al. developed an ontology-based text classification method to support automated job hazard analysis, identify key risk factors associated with each construction incident, and determine the critical risk combinations leading to accidents (Chi & Han, 2013). Salama and El-Gohary proposed a semantic machine learning-based text classification algorithm for categorizing clauses and sub-clauses in textual documents, facilitating compliance checking for construction regulations and contracts (Salama & El-Gohary, 2016).

With advancements in machine learning, new possibilities have emerged for text classification. For instance, Yang Miang Goh applied algorithms such as Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF), K-Nearest Neighbors (KNN), Decision Tree (DT), and Naive Bayes (NB) to classify construction incident narratives. The F1 score of SVM ranged from 0.45 to 0.92, outperforming other classifiers (Goh & Ubeynarayana, 2017). Building on this, Fan Zhang et al. proposed an ensemble model based on multiple machine learning algorithms to classify the causes of construction accident reports (Zhang et al., 2019).

However, a limitation of machine learning is the requirement for manual feature extraction. Compared to traditional machine learning, deep learning is an end-to-end process capable of automatically learning features from training datasets. Today, research on text classification using deep learning is increasingly prevalent. In the construction domain, Botao Zhong et al. proposed a method combining Natural Language Processing (NLP) and Convolutional Neural Networks (CNN) to analyze and classify construction incident texts, while using an LDA model to explore the intrinsic relationships between different categories of accident causes (Zhong et al., 2020). Dan Tian et al. employed a CNN-based text classification model to categorize textual descriptions of construction site conditions into six classes, demonstrating the proposed model's reliability and applicability in handling large-scale construction site texts (Tian et al., 2021). Hrishikesh Gaddekar and Nikhil Bugalia proposed a semi-supervised YAKE-GLDA method for the automatic classification of construction safety reports. However, the YAKE-GLDA method is suitable only for medium-sized databases and not for smaller ones (Gaddekar & Bugalia, 2023). Although there have been studies on text classification in the construction field, there is still a gap in research on small-sample text classification.

2.2 Application of BERT in Text Classification

In most cases, obtaining large domain-specific datasets is challenging. Google has released a new language representation model based on the Transformer, Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2018). BERT is pre-trained on a large corpus and fine-tuned by adding additional output layers. Therefore, advanced models built on BERT can accomplish NLP tasks without the need for training on large-scale datasets.

BERT has been successfully applied across various domains, demonstrating outstanding performance in text classification tasks. Chi Sun investigated a series of fine-tuning approaches, providing a general framework for fine-tuning BERT across different text classification tasks (Sun et al., 2019). Fang et al. developed a novel

BERT-based model for the automatic classification of near-miss information in safety reports. The model was validated using a database of near-miss incident reports from real projects, achieving a 10% increase in accuracy after fine-tuning (Fang et al., 2020).

Several studies have enhanced model performance by modifying BERT's classification head and integrating it with other neural networks. Kamaljit Kaur et al. proposed a Bidirectional Encoder-Decoder Transformer-Convolutional Neural Network (BERT-CNN) model for requirement classification, which improves model performance by stacking convolutional layers on top of the BERT layer. Experiments conducted on the PROMISE dataset containing 625 requirements demonstrated that the proposed model outperformed state-of-the-art baseline methods (Kaur, K., & Kaur, P., 2023). Nishant Rai attempted to connect the output layers of an LSTM model and a BERT model for fake news classification of news headlines. Testing on the PolitiFact and GossipCop datasets showed an improvement in accuracy (Rai et al., 2022).

When facing challenging tasks or insufficient training data, ensemble methods are often employed to enhance model performance. J. Briskilal integrated BERT and RoBERTa models for the classification of idiomatic and literal text (Briskilal & Subalalitha, 2022). Rohan Singh Wilkho utilized the Bagging ensemble method, combining different models to achieve superior predictive performance. He fine-tuned 50 models for each architecture type on different subsets of the training-validation set and experimentally determined the optimal number of models to ensemble (Wilkho & Gharaibeh, 2024).

Based on previous works, we propose a BERT-based ensemble model for classifying imbalanced small-sample text datasets. This model can automatically assess the risk status of construction sites based on monitoring reports.

3. MODEL ARCHITECTURE

Our model is composed of multiple BERT classification models. Each model consists of a pre-trained BERT model and a classification head, both of which are fine-tuned on our dataset. The specific architecture is shown in Figure 1. The model takes a piece of text as input and outputs a label based on the content of the text. The following section introduces our model architecture.

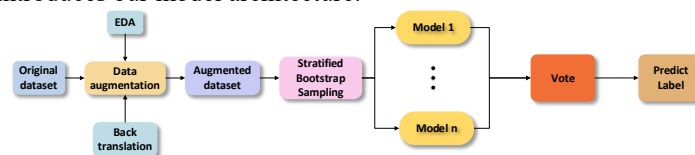


Figure 1. BERT-Based Ensemble Model Framework for Small-Sample Text Classification

3.1 Data Augmentation

This research focuses on text classification problems for small-sample datasets. To increase sample diversity, we used a data augmentation strategy combining back-translation and Easy Data Augmentation (EDA). EDA is a method for expanding text data through simple operations such as synonym replacement, random insertion, random swapping, and random deletion. By making slight modifications to the original text, the generated samples help the model learn richer textual expressions, improving its robustness. Compared to other data augmentation methods, this approach is simple, requiring no complex algorithms or extensive computational resources, while increasing sample diversity at minimal cost.

3.2 Bagging Ensemble

We used the Bagging ensemble method in our model. Bagging (Bootstrap Aggregating) is a widely used ensemble learning method aimed at improving machine learning models by combining the predictions of multiple models, reducing the risk of overfitting, and enhancing generalization. We trained different classifiers on various subsets of the same training dataset, with these subsets generated from the original dataset using Bootstrap sampling (which allows samples to be selected multiple times). After creating the subsets, we trained multiple models on them, which together formed the Bagging ensemble model. The final prediction is made through a simple majority voting process.

3.3 BERT Classification Model

Our model is built on the integration of multiple BERT classification models. This section focuses on explaining the individual BERT models in the ensemble, with the architecture of a single model shown in Figure 2. In each individual model, we use the Focal loss function to address the issue of data imbalance, and apply a learning rate decay strategy along with the AdamW optimizer to enhance model accuracy. Further explanation of the individual models will be provided in later sections.

(1) Input Embedding

BERT's input embedding consists of three main components: token embedding, segment embedding, and

position embedding. These embeddings work together to transform the text into a numerical form that the model can process. The BERT model only accepts fixed-length input sequences (with a maximum length of 512 tokens). We have fixed the input sequence length to 512 tokens, padding shorter sequences with empty tokens.

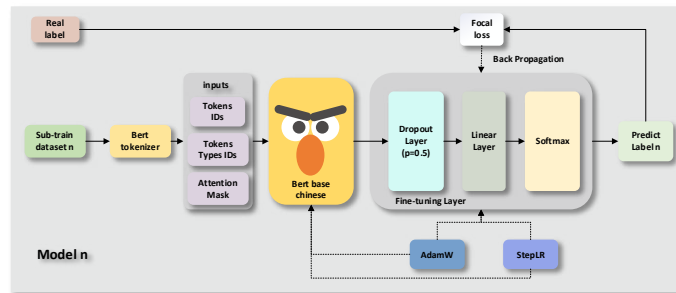


Figure 2. Structure of an individual BERT classification model

(2) Multi-layer Bidirectional Transformer Encoder

The multi-layer bidirectional Transformer encoder in BERT is the core of the model, responsible for processing text data through the input embedding layer. The encoder is based on the Transformer architecture and is composed of multiple identical stacked layers. Each layer consists of two main components: a multi-head self-attention mechanism and a position-wise feed-forward network.

(3) Classification Head

We use a pre-trained BERT model as a feature extractor, and add a fully connected layer with a softmax activation function on top for classification. The fully connected layer and activation function together form the classification head of the model, responsible for converting BERT's high-dimensional feature representation into the final classification prediction. Through the fully connected layer, the features extracted by BERT are mapped to the final classification labels. The output dimension of this layer equals the number of categories in the classification task. To improve the model's generalization ability, dropout is applied before the fully connected layer, randomly "dropping" some neuron outputs to reduce overfitting. Finally, the softmax function is used to calculate the predicted probability for each category. Additionally, Focal Loss is used to calculate the loss during model training, optimizing the process.

4. EXPERIMENTS

4.1 Dataset

Our dataset consists of 158 monitoring reports from a metro construction project, with experts manually labeling four warning levels: no warning, yellow warning, orange warning, and red warning. To increase sample diversity, we used back-translation and EDA methods for data augmentation. Table 1 shows the sample count for each warning level. The data is highly imbalanced, so we have applied two methods to cope with data imbalance: the Focal Loss function and the Bagging ensemble.

Table 1. The distribution of labels in the dataset

Dataset Labels	The distribution of labels in the dataset	
	Original Dataset	Augmented Dataset
No Warning	24	48
Yellow Warning	119	238
Orange Warning	9	108
Red Warning	4	48

When using BERT for text classification, preprocessing steps such as tokenization, adding special tokens, generating Token IDs and Type IDs, and constructing attention masks are necessary to meet the model's input requirements.

4.2 Training, Validation, Testing

This paper applies fully supervised training of the BERT model on a small dataset, fine-tuning task-specific patterns to improve training speed and accuracy. By analyzing weights and the attention mechanism, the model provides interpretability for small datasets. The data is split into two groups: 70% for training and 30% for testing, with the training set further divided into two groups: 80% for training and 20% for validation.

We used the training set to perform bagging ensemble learning. We split the training dataset into 5 different subsets to train 5 different models (for each architecture).

The training and validation of BERT involve the fine-tuning process. The general fine-tuning parameters are: training epochs = 8, initial learning rate = $8e-6$, optimizer = AdamW, and the loss function is FocalLoss with class weights of 5:1:30:1.

(1) Focal Loss Function

The cross-entropy loss function is commonly used when training deep learning models for classification, as it measures the difference between two probability distributions for a given random variable or set of events. However, this research focuses on imbalanced datasets. In such cases, training with the cross-entropy loss function can lead to models that disproportionately favor the majority class, resulting in suboptimal performance for the minority class. To mitigate this issue, the focal loss function is employed in this paper as a replacement for the cross-entropy loss during model training. Focal Loss function is calculated using Equation (1).

$$FL(p_i) = -\alpha_i(1 - p_i)^\gamma \log(p_i) \quad (1)$$

Where p_i corresponds to the predicted probability of the true class, α_i denotes the modulating factor to handle class imbalance. γ is a parameter that adjusts the attention given to easy and hard samples. Overall, the focal loss function addresses data imbalance by modulating the contribution of each sample to the loss. α_i alleviates the imbalance from the perspective of sample class by adjusting the weight of the loss calculation for different classes, while γ enhances the contribution of hard-to-learn samples to improve model performance. Typically, $\gamma=2$ is set to a standard value in the field of computer vision.

(2) Comparative experiment and sensitivity analysis

To validate the effectiveness of Focal Loss function, comparative experiments were conducted. The results show that, compared to the traditional cross-entropy loss function, Focal Loss function demonstrates stronger classification performance on imbalanced datasets (see Table 2). In addition, we also studied the impact of the α_t parameter on the model's classification performance. With $\gamma = 2$, the α_t parameter was systematically adjusted in the range from 0 to 30, with intervals of 10, and the classification performance was recorded after each training session. The experimental results show that as the α_t value increases, the model's classification performance on imbalanced classes improves, further confirming the model's sensitivity to the α_t parameter. This finding suggests that selecting an appropriate α_t parameter helps improve the model's ability to identify minority classes.

Table 2. Comparative experiment and sensitivity analysis of Focal Loss function

Types of Loss Functions	Index	No warning	Yellow warning	Orange warning	Red warning
cross-entropy loss function	F1score	0.8959	0.9270	0.8616	1
Focal Loss function (class weight =1:1:1:1)	F1score	0.8959	0.9329	0.8808	1
Focal Loss function (class weight =1:1:10:1)	F1score	0.8887	0.9404	0.8925	1
Focal Loss function (class weight =1:1:20:1)	F1score	0.9280	0.9464	0.8925	1
Focal Loss function (class weight =1:1:30:1)	F1score	0.9280	0.9592	0.9356	1

(3) Learning Rate Decay

Learning rate is a key hyperparameter that controls how the model's weights are adjusted. An appropriate learning rate can speed up convergence and avoid oscillations or overly slow progress. A traditional constant learning rate often fails to meet the needs of deep networks, requiring numerous experiments and making it difficult to find the optimal value. This paper uses a time decay strategy, starting with a larger learning rate to quickly approach the optimal solution, and then gradually reducing it to allow fine-tuned adjustments in later stages and avoid oscillations.

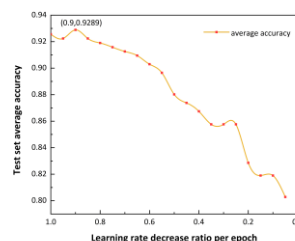


Figure 3. Optimal learning rate decay ratio

For small datasets, we used cross-validation to determine the hyperparameters. After multiple tests, the

learning rate was reduced to 90% of its original value after each epoch, achieving the highest accuracy, as shown in Figure 3.

4.3 Evaluation Metrics

Since our research involves a multi-label classification problem, it is important to evaluate not only the overall performance of the model, but also its ability to correctly identify each label. Therefore, we use the F1 score to assess the model's performance on each individual label and the micro-F1 score to evaluate the model's overall performance across all labels. The F1 score is calculated using Equations (2), (3), and (4).

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

$$F1\text{score} = 2 \times \frac{\text{precision} * \text{Recall}}{\text{precision} + \text{Recall}} \quad (4)$$

5. CASE STUDY

To validate the effectiveness of the proposed small-sample text classification framework in automatic text classification, we created two datasets. The data samples are shown in Table 3. The model's performance was evaluated through 7-fold cross-validation, and we compared our model with other models suitable for small-sample classification as benchmarks.

Table 3. Labels used for the data, their criteria and sample narrative.

ID	Label	Criteria	Narrative example
0	no warning	Based on the four aspects included in the inspection daily reports—construction conditions, support structure, surrounding environment, and monitoring facilities—experts were asked to score the reports and classify the foundation pit risk warning levels into four categories.	<p>Narrative: 开挖面土体为中密卵石，基坑开挖区域为0-24轴，1-8轴制作顶板模板，8-12轴中板浇筑，12-16轴搭设中板脚手架，16-19轴制作底板垫层，19-24轴已开挖至设计深度，停止开挖，第四层钢支撑、第三层钢支撑1-18轴、第二层钢支撑1-10轴已拆除，东端盾构井有渗水，系用坑内降水，降水设施运转正常，支护状体无裂缝、无明显缺陷、隆起，基坑侧壁及坑内无涌水、流沙、管涌，基坑周围无超载。支护桩无裂缝、侵陷等情况，冠梁已施工完成，连续性完好，无过大变形、裂缝基坑防坠措施完好，钢支撑架设及时。基坑周边土体存在细小裂缝、车辆碾压等情况，无明显沉降隆起。地表竖向位移监测点DBC29-01、DBC29-02等破坏，桩顶位移监测点ZQS27无法观测。</p> <p>Label: yellow warning</p>
1	yellow warning		
2	orange warning		
3	red warning		

5.1 Summary of the data set

For training and testing purposes, we divided 156 original monitoring inspection reports from a metro project into a total training dataset and a test dataset in a 7:3 ratio. The total training dataset was further split into a training set and a validation set at a ratio of 8:2. The inspection reports are formulated by manual daily inspection about construction conditions, support structure, surrounding environment, and monitoring facilities, which provide a comprehensive view of the foundation pit's risk conditions.

We invited experts with safety management experience over 15 years in metro construction, to label inspection reports by evaluating risk levels based on each inspection report. The risk levels are divided into four categories: no warning, yellow warning, orange warning, and red warning. No warning indicates that the project is in a low-risk state, and risk response measures are generally not required. Yellow warning indicates that the project's risk level is slightly higher than no warning, requiring minimal resources to control the risk. Orange warning indicates that the project is in a medium-high risk state, necessitating considerable resources and proper risk control measures to control the risk. Red warning indicates that the project is in a high-risk state, which typically leads to severe engineering accidents and requires close attention from project managers. Due to the specific nature of the construction industry, serious risks are undesirable. Reasonable risk control measures are often taken before risk evolves to high levels, leading to significant data imbalance. Normal and yellow warning samples make up the majority, while orange and red warnings are rare. Additionally, our dataset faces the issue of being a small sample dataset.

To improve the model's training performance and reliability, we applied data augmentation method to original dataset. In the following sections, we will use both the original data and the augmented data to train and evaluate the model separately.

5.2 Model Performance Based on Small-Sample Training

The BERT ensemble model was implemented using the Pytorch deep neural network platform. The hyperparameters of the deep neural network were determined based on the best results from multiple rounds of

experimental testing. The specific parameters are shown in Table 4.

80% of the training set was used for model training, while the remaining 20% was used for model validation. The training loss from the original dataset is shown in Figure 4. After 20 epochs, the training loss fluctuated slightly, indicating that the model had reached optimal performance on the current dataset. In the final stage of training, the recall rate for the predefined labels stabilized at a relatively steady level. After 25 epochs of training, the recall rates for the normal and yellow warning labels stabilized above 0.9. Although there were only 13 samples for the normal label, the prediction accuracy was high, indicating that reports without warnings were more clearly expressed. However, due to significant data imbalance in the original training dataset, the red and orange warning samples were scarce (2 red samples and 4 orange samples), making them hard-to-learn samples and resulting in considerable fluctuations during training.

Due to the limited sample size, we conducted a 7-fold stratified cross-validation to verify the reliability of the trained model. We randomly divided the total training set consisting of 109 samples into 7 parts, aiming to maintain the original sample distribution as much as possible. In each round, one part was selected as test set, while the remaining 6 parts were combined as the training data set. Figure 5 (a, b) shows the precision and recall distributions of the classification labels from the 7-fold cross-validation. Since the number of orange and red warning samples in the dataset is extremely limited, and the distinction between orange and yellow warnings is not clear, random partitioning cannot ensure that each fold in cross-validation contains all four label categories. Therefore, in Figure 5 (a, b), the lowest training precision and recall rates for the orange and red warning labels can be zero.

The precision and recall distributions from the 7-fold cross-validation show that the range for orange and red warning labels is quite large, while the precision and recall for no warning and yellow warning labels are distributed at a higher level. The training results indicate that the size of the training dataset directly affects the model's robustness. Due to the lack of training samples, the model's robustness is weak, especially for sample categories with lower proportions in the training dataset. However, the metrics for no warning and yellow warning labels are distributed at a high level, indicating that the model can perform text classification tasks well once the dataset reaches a sufficient size.

Table 4. Bert integrated model training parameters.

Hyperparameters	Value
Batch Size	2
Dropout Ratio	0.5
Optimizer	AdamW
Initial learning rate	8e-6
Optimal number of integration models	5

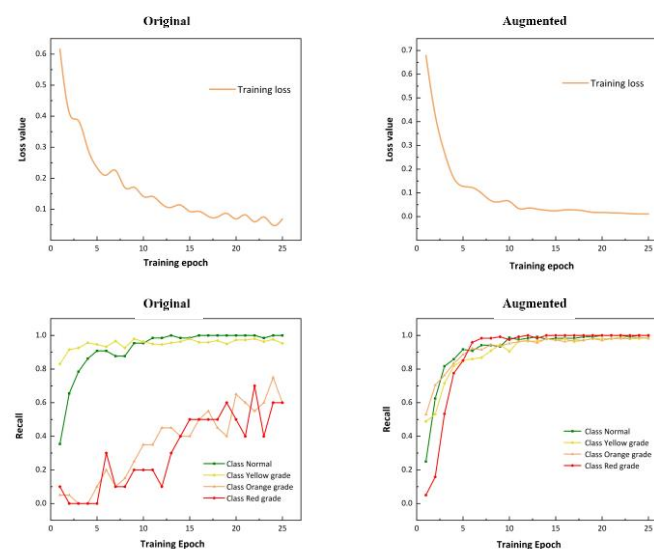


Figure 4. Training loss curve and Training recall curve

5.3 Model Performance Based on Data-Augmented Training

In the initial stage, the total training dataset contained only 109 samples, with very limited samples for each category. Therefore, we expanded our dataset using a combination of EDA and back-translation data augmentation methods. The expanded training dataset included 309 samples.

Training on the augmented dataset improved the model's robustness. As shown in Figure 4, after 15

training epochs, the training loss stabilized without fluctuation. The recall rate during the training process stabilized after 10 epochs. The expanded training dataset made the model training more stable, with faster convergence, and there were no significant fluctuations in any class labels during training. Evaluating the model's performance in the final training stage, the recall rates for all class labels exceeded 0.9, indicating a significant improvement in the model's training performance.

To verify the model's robustness, a 7-fold cross-validation was performed on the augmented dataset, with precision and recall rates shown in Figure 5 (d, e). Except for the orange warning, the average precision and recall rates for the other warning categories were above 0.95. Although the recall rate for the orange warning fluctuated, its average reached 0.8.

5.4 Model's robustness validation

To accurately evaluate the model's performance, we validated the model on a completely new test set, and the test results are shown in Figure 5 (c, f). On the original dataset, although the overall test accuracy reached 87.23%, for the orange warning category, all 3 test samples were predicted incorrectly, resulting in an accuracy of 0 for this label. As for the red warning samples, although they were all predicted correctly, there was only one red sample in the original test set, so the accuracy result lacks statistical significance due to the small sample size.

On the augmented dataset, we clearly saw that although the overall accuracy maintains 96.24%, the model also achieved high accuracy for each individual category, with only few yellow and orange samples incorrectly predicted. While ensuring high accuracy, the augmented test set contains enough samples for each category, making the results more convincing.

To examine the advantage of the model in small-sample text classification, we compared the ensemble model with other BERT-based text classification models. We trained and tested the models using both the original dataset and the augmented dataset. For a more comprehensive evaluation, we calculated the precision, recall, and F1 score for each category in the test set, and the results are shown in Table 5.

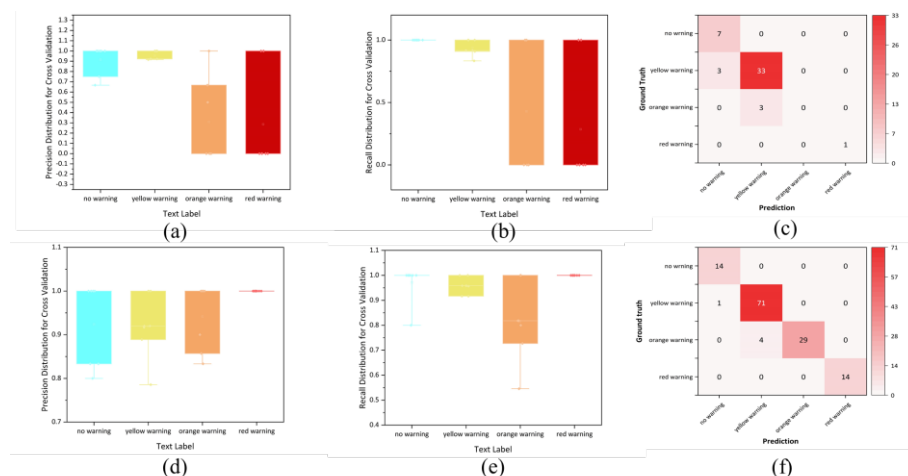


Figure 5 Performance comparison of the model on the original (a, b, c) and augmented (d, e, f) datasets.

First, the models were trained on the original dataset. For the normal and yellow warning categories, the F1 scores of the models were almost identical and performed well. Due to the small number of orange samples, all models performed poorly in this category, but BERT + DPNN and BERT + RCNN showed some improvement in predicting the orange warnings. Although the number of red warning samples was also small, all models performed well in identifying the red warnings.

On the augmented dataset, the F1 scores of all models are relatively high, indicating that each model performs well in classifying imbalanced small samples. However, the F1 scores of the ensemble model are higher than those of the other models, demonstrating that our proposed ensemble model offers significant advantages and better performance in small-sample text classification tasks.

6. DISCUSSION

Small-sample datasets are prone to overfitting during model training, making it difficult to learn features accurately. This is often accompanied by data imbalance issues, as seen in our dataset, where the model tends to favor the majority classes and ignore the minority classes. As shown in Figure 4, during the convergence process, the model's training loss on the original dataset did not stabilize at a low level in the later stages and exhibited some fluctuations. The recall for the orange and red warning categories fluctuated significantly during training. On the original dataset, as shown in Figure 5 (c), the ensemble model achieved an overall test accuracy of 87.23%, but for the orange warning category, all 3 test samples were predicted incorrectly, resulting in an accuracy of 0 for

this category. As for the red warning samples, all were predicted correctly, there was only one sample, making the result less reliable due to randomness. To address the issue of insufficient data, we applied data augmentation techniques to expand our dataset, increasing sample diversity and providing more learnable samples. During the convergence process, compared to the original dataset, the model's training loss on the augmented dataset was smoother and stabilized at a lower level, with recall for all categories converging to 0.95, as shown in Figure 4. The generalization ability of the ensemble model was also improved, with the average metrics for all categories in 7-fold cross-validation exceeding 0.8. Additionally, the BERT-based model more accurately identified category features on the augmented test set.

In Section 5.4, when comparing with other models, the performance differences between models on the original dataset were small, mainly due to the limited dataset size, which restricted the models' capabilities. However, on the augmented dataset, the ensemble model outperformed other models on all metrics, indicating that our multi-model ensemble architecture and the Focal Loss function performed well in coping with data imbalance in small datasets.

Table 5. Original Data Set Training based Model Performances.

Model name	Index	No warning		Yellow warning		Orange warning		Red warning	
		Original	Augmented	Original	Augmented	Original	Augmented	Original	Augmented
BERT-based ensemble model	Precision	0.7000	0.9333	0.9167	0.9467	0.0000	1.0000	1.0000	1.0000
	Recall	1.0000	1.0000	0.9167	0.9861	0.0000	0.8788	1.0000	1.0000
	F1score	0.8235	0.9655	0.9167	0.9660	—	0.9355	1.0000	1.0000
BERT+CNN	Precision	0.7000	0.8235	0.9167	0.8718	0.0000	0.9583	1.0000	1.0000
	Recall	1.0000	1.0000	0.9167	0.9444	0.0000	0.6970	1.0000	1.0000
	F1score	0.8235	0.9032	0.9167	0.9066	—	0.8070	1.0000	1.0000
BERT+RNN	Precision	0.7000	0.8235	0.9167	0.9571	0.0000	0.9375	1.0000	1.0000
	Recall	1.0000	1.0000	0.9167	0.9306	0.0000	0.9091	1.0000	1.0000
	F1score	0.8235	0.9032	0.9167	0.9461	—	0.9231	1.0000	1.0000
BERT+DPN N	Precision	0.7000	0.9286	0.9697	0.9583	0.6667	0.9375	1.0000	1.0000
	Recall	1.0000	1.0000	0.8889	0.9583	0.6667	0.9091	1.0000	1.0000
	F1score	0.8235	0.9630	0.9275	0.9583	0.6667	0.9231	1.0000	1.0000
BERT+RCN N	Precision	0.7000	0.8235	0.9412	0.9189	1.0000	0.9643	1.0000	1.0000
	Recall	1.0000	1.0000	0.9143	0.9444	0.3333	0.8182	1.0000	1.0000
	F1score	0.8235	0.9032	0.9276	0.9315	0.5000	0.8853	1.0000	1.0000

However, there are certain limitations to our research. First, the model training still needs manual labels, which is a time-consuming process. With the accumulation of domain-specific datasets, manual work can be reduced with the help of unsupervised training on large datasets. Additionally, we only trained and tested the model's effectiveness on the inspection reports from one real-world project. Future work will be requires testing the model on other text classification problems in the construction industry.

7. CONCLUSION

During construction, a large amount of unstructured and semi-structured text is generated, providing project managers with valuable information. However, reading and identifying useful information in inspection reports and classifying warning levels is typically a manual and time-consuming process. Additionally, due to the limited data sources and the high cost of manual labeling, it is difficult to form large-scale domain datasets.

In summary, this paper proposes a BERT ensemble model to address the issue of imbalanced small-sample datasets. First, a combination of EDA and back-translation data augmentation techniques is applied to enhance the diversity of the small-sample dataset without collecting additional data. The training dataset is then divided into multiple subsets, with each model trained on a different subset. The final prediction label is determined by a voting mechanism. For individual models, learning rate gradient descent is used to reduce the risk of overfitting, and the Focal Loss function assigns weights to different classes during loss calculation, improving the model's ability to learn from minority, hard-to-learn samples. Finally, to verify the feasibility of the proposed model, the paper classifies real textual data from inspection reports of a real project. The data was split into training and test sets, with cross-validation performed on the training set, showing that the model performs well on different datasets and has strong generalization capabilities. Additionally, the ensemble model achieved a classification accuracy of 96.24% on the test set, slightly outperforming other classification models and demonstrating good robustness.

This model framework provides a viable path for automated risk assessment from construction site

inspection reports, supporting intelligent construction management, enabling timely responses to safety risks.

ACKNOWLEDGMENTS

This research is supported by the National Natural Science Foundation of China (Grant Nos. 72201198).

REFERENCES

- Baker, H., Hallowell, M. R., & Tixier, A. J.-P. (2020). Automatically learning construction injury precursors from text. **Automation in Construction*, 118*, 103145. <https://doi.org/10.1016/j.autcon.2020.103145>
- Briskilal, J., & Subalalitha, C. N. (2022). An ensemble model for classifying idioms and literal texts using BERT and RoBERTa. *Information Processing and Management*, 59, 102756. <https://doi.org/10.1016/j.ipm.2021.102756>.
- Bureau of Labor Statistics (BLS), & Occupational Safety and Health Administration (OSHA). (2018). Commonly used statistics. OSHA.gov. <https://www.osha.gov/oshstats/commonstats.html> (Accessed: May 11, 2019).
- Chi, S., & Han, S. (2013). Analyses of systems theory for construction accident prevention with specific reference to OSHA accident reports. *International Journal of Project Management*, 31(7), 1027-1041. <https://doi.org/10.1016/j.ijproman.2012.12.004>.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. [arXiv Preprint arXiv:1810.04805](https://arxiv.org/abs/1810.04805).
- Fang, W., Luo, H., Xu, S., Love, P. E. D., Lu, Z., & Ye, C. (2020). Automated text classification of near-misses from safety reports: An improved deep learning approach. *Advanced Engineering Informatics*, 44, 101060. <https://doi.org/10.1016/j.aei.2020.101060>.
- Gadekar, H., & Bugalia, N. (2023). Automatic classification of construction safety reports using semi-supervised YAKE-guided LDA approach. *Advanced Engineering Informatics*, 56, 101929. <https://doi.org/10.1016/j.aei.2023.101929>.
- Goh, Y. M., & Ubeynarayana, C. U. (2017). Construction accident narrative classification: An evaluation of text mining techniques. *Accident Analysis and Prevention*, 108, 122-130. <https://doi.org/10.1016/j.aap.2017.08.026>.
- Kaur, K., & Kaur, P. (2023). BERT-CNN: Improving BERT for requirements classification using CNN. *Procedia Computer Science*, 218, 2604-2611. <https://doi.org/10.1016/j.procs.2023.01.234>.
- Kim, T., & Chi, S. (2019). Accident case retrieval and analyses: Using natural language processing in the construction industry. *Journal of Construction Engineering and Management*, 145(3), 04019004. [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0001625](https://doi.org/10.1061/(ASCE)CO.1943-7862.0001625).
- Rai, N., Kumar, D., Kaushik, N., Raj, C., & Ali, A. (2022). Fake news classification using transformer-based enhanced LSTM and BERT. *International Journal of Cognitive Computing in Engineering*, 3, 98-105. <https://doi.org/10.1016/j.ijcce.2022.03.003>.
- Salama, D. M., & El-Gohary, N. M. (2016). Semantic text classification for supporting automated compliance checking in construction. *Journal of Computing in Civil Engineering*, 30(1), 04014106. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000301](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000301).
- Sun, C., Qiu, X., Huang, X., & Xu, Y. (2019). How to Fine-Tune BERT for Text Classification?. Springer, Cham, arxiv.org/abs/1905.05583.
- Tian, D., Li, M., Shi, J., Shen, Y., & Han, S. (2021). On-site text classification and knowledge mining for large-scale projects construction by integrated intelligent approach. *Advanced Engineering Informatics*, 49, 101355. <https://doi.org/10.1016/j.aei.2021.101355>.
- Wilkho, R. S., Chang, S., & Gharaibeh, N. G. (2024). FF-BERT: A BERT-based ensemble for automated classification of web-based text on flash flood events. *Advanced Engineering Informatics*, 59, 102293. <https://doi.org/10.1016/j.aei.2023.102293>.
- Zermane, A., & Mohd Tohir, M. Z. (2023). Predicting fatal fall from heights accidents using random forest classification machine learning model. *Safety Science*, 159, 106023. <https://doi.org/10.1016/j.ssci.2022.106023>.
- Zhang, D., Li, M., Tian, D., Song, L., & Shen, Y. (2022). Intelligent text recognition based on multi-feature channels network for construction quality control. *Advanced Engineering Informatics*, 53, 101669. <https://doi.org/10.1016/j.aei.2022.101669>.
- Zhang, F., Fleyeh, H., Wang, X., & Lu, M. (2019). Construction site accident analysis using text mining and natural language processing techniques. *Automation in Construction*, 99, 238-248. <https://doi.org/10.1016/j.autcon.2018.12.016>.
- Zhong, B., Pan, X., Love, P. E. D., & Ding, L. (2020). Deep learning and network analysis: Classifying and visualizing accident narratives in construction. *Automation in Construction*, 113, 103089. <https://doi.org/10.1016/j.autcon.2020.103089>.

AN LLM AGENT-BASED APPROACH FOR CALCULATING FLOOR AREAS VIA BIM IFC FILES

Xueying Zhu¹, and Jun Ma²

1) Ph.D. Student, Department of Urban Planning and Design, The University of Hong Kong, Hong Kong, SAR. Email: xueyingz@connect.hku.hk

2) Ph.D., Asst. Prof., Department of Urban Planning and Design, The University of Hong Kong, Hong Kong, SAR. Email: junma@hku.hk

Abstract: The accurate and efficient calculation of floor areas is a critical aspect of building design processes, yet it remains a challenging task due to the intricate rules and regulations that must be adhered to. Traditional methods, which rely on manual calculations, are not only time-consuming but also prone to human error, highlighting the need for an automated solution. Building Information Modeling (BIM) can support the automated solution by providing digital representations of building designs. The openBIM workflow enables collaborative work among various stakeholders, allowing them to share and integrate project information seamlessly without being locked into proprietary systems. While Industry Foundation Classes (IFC) files serve as a robust digital representation of BIM, facilitating data exchange across different software platforms, the complexity associated with processing these files poses a significant barrier to ordinary users. To address these gaps, we propose an innovative semi-automated floor area calculation system that leverages state-of-the-art Large Language Model (LLM) agent. This system is designed to efficiently and accurately extract relevant information from IFC files applying specific filtering conditions and then perform the necessary calculations to determine floor areas according to the required standards. Due to LLMs' characteristic of interacting using natural language, our approach significantly reduces the technical barriers faced by non-expert users, making the process of floor area calculation more accessible and user-friendly. We validate the proposed method through the practice of simplified floor area calculation of an academic building, offering a promising solution to streamline BIM workflows and enhance productivity in the architecture, engineering, and construction (AEC) industries.

Keywords: Floor Area Calculation, BIM, IFC Information Extraction, LLM Agent

1. INTRODUCTION

In the architectural design process, manual area calculation has long been a fundamental but labor-intensive task. Traditional methods involve detailed measurements and adherence to complex rules and standards. The rules governing area calculations are often intricate, involving specific criteria such as the inclusion or exclusion of certain spaces based on their function, height, and other attributes. Studies have shown that architects and engineers spend a considerable amount of time on these repetitive and non-creative tasks (Sang et al., 2009), which significantly drains their energy and impacts overall work efficiency. Therefore, there is an urgent need for automated solutions to assist in these tasks and improve productivity.

Several attempts have been made to develop automated systems for area calculation. Early efforts focused on rule-based algorithms and expert systems, which could handle specific scenarios but lacked flexibility and scalability (Roanes-Lozano et al., 2000). Later research utilized Dynamo visual programming software to automate floor area calculation and compliance checking within BIM, streamlining the process and improving accuracy and efficiency (Reinhardt & Mathews, 2017). More recent approaches have explored machine learning techniques, particularly supervised learning, to improve the accuracy and adaptability of automated area calculations (Özer & Jacoby, 2022). Some researchers also use traditional programming called high-level implementable methods, such as *getTotalFloorArea()*, to accurately calculate and verify floor areas from building models for code compliance checking (Lee et al., 2023). However, these systems still face challenges in handling the variability and complexity of real-world data.

Building Information Modeling (BIM) is a digital representation of physical and functional characteristics of a building. It serves as a shared knowledge resource for information about a building, forming a reliable basis for decisions during its lifecycle. BIM models contain a wealth of information, including geometric data, material specifications, and performance metrics, which can be leveraged for various purposes, including automated area calculations. Industry Foundation Classes (IFC), a file format standard for BIM data exchange, enabling interoperability between different software platforms, can provide rich data for automated calculation of floor areas. However, the complexity of processing IFC files poses a significant barrier for non-specialist users, as existing software tools require advanced technical skills to operate effectively.

The emergence of Large Language Models (LLMs) offers a new avenue for addressing these challenges. LLMs, such as GPT series, are capable of understanding and generating human-like text, making them suitable for tasks that require natural language interaction (Achiam et al., 2023). In the AEC industry, LLMs have been applied to tasks such as document summarization, code generation, and data extraction (Zheng et al., 2023).

LLM agents as advanced AI systems capable of understanding and processing complex tasks, can handle multi-step reasoning problems and incorporate a reflection capability to ensure the accuracy of the results. They can perform complex operations and solve intricate problems by leveraging their capabilities of planning, finding information, remembering past interactions, and learning from them. This makes them particularly suitable for solving these intricate, double-hop or even multi-hop problems. Recent research has begun to investigate the use of LLM agents for some tasks. Various fields, such as the financial sector (Li et al., 2023) and chemical research (M. Bran et al., 2024), have already begun to harness these advanced technologies. However, their potential for automating area calculations in BIM has not been fully explored.

In this paper, we introduce a novel LLM agent specially designed to semi-automate floor area calculations. To enhance the problem-solving capabilities of LLM agents, techniques such as Chain-of-Thought (CoT) planning (Wei et al., 2022) and Reasoning and Acting (ReAct) frameworks (Yao et al., 2022) have been proposed. CoT planning involves breaking down complex tasks into smaller, manageable steps, allowing the LLM to think through the problem-solving process systematically. The ReAct framework further refines this process by enabling the LLM to reflect on its actions and adjust its strategy based on feedback. Our approach utilizes CoT planning, enabling the LLM to think through the problem-solving process step-by-step, like a human. This method allows the LLM to interpret and process IFC data effectively, extracting the necessary information to calculate floor areas accurately. We employ the ReAct framework to refine the agent's actions and ensure continuous improvement. The ReAct framework enables the agent to reflect on its previous actions and adjust its strategy, accordingly, enhancing the accuracy and reliability of the calculations. Besides, our system consists of two main tools: a calculator tool and a coding tool. The calculator tool performs the area calculations based on the extracted information using the coding tool to generate python codes to do the calculation. We tested our LLM agent on an academic building using a simplified method and achieved positive results. These findings provide new insights and methodologies for future work in automating building area calculations.

The main contributions of this paper are:

- **Development of an LLM Agent:** We have created a specialized LLM agent that can interpret and process IFC files to semi-automate floor area calculations. Our approach uses CoT planning to enhance the agent's problem-solving capabilities, ensuring accurate and reliable calculations. The ReAct framework allows the agent to continuously improve its performance through iterative refinement.
- **Empirical Validation:** We have demonstrated the effectiveness of our approach through testing on an academic building model in Revit, achieving positive results.

By addressing the limitations of manual area calculations and the complexities of IFC file processing, our LLM agent offers a promising solution to enhance productivity and efficiency in the architectural design process.

2. METHOD

2.1 Methods for Calculating Floor Area

According to the latest version of the "*Rules for Calculating Floor Area*" issued and implemented by the Ministry of Construction of the People's Republic of China, the main rules for calculating floor area can be divided into two categories: (a) the scope of areas that should be included in the floor area calculation, such as single-story and multi-story buildings with enclosing structures, and (b) the scope of areas that should not be included in the floor area calculation, such as outdoor ladders used for maintenance or fire safety. The scope of areas that should be included in the floor area calculation can further be divided into areas that are calculated in full, and areas that are calculated at 50% of their actual size. Detailed regulations are shown in Figure 1 below.

From Figure 1, we can see that the most significant factor affecting the area calculation coefficient is whether the floor height exceeds 2.2 meters, so in this study, we simplify the area calculation problem and focus only on the simplest scenario where the floor area is calculated in full for buildings with a story height of more than 2.2 meters. Additionally, more complex calculation scenarios, such as balconies and stairs that lie outside the main structural envelope, are not considered at this stage.

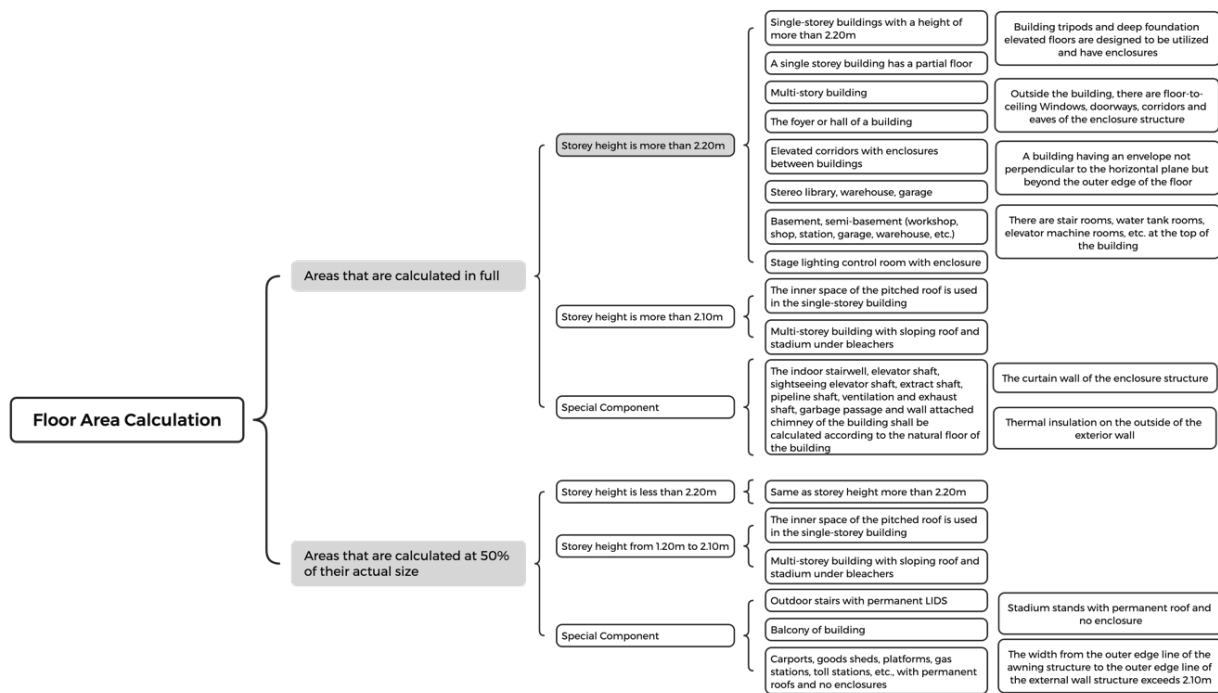


Figure 1. Detailed regulations for floor area calculation

2.2 Floor-Area-Calculation Agent

In this research we propose a specially designed LLM agent to solely solve the floor-area-calculation problems. Leveraging the power of LLMs, this agent efficiently extracts and processes data from IFC files, adhering to the latest regulations, such as the "Rules for Calculating Floor Area" issued by the Ministry of Construction of the People's Republic of China. The agent employs CoT planning and the ReAct framework to handle complex, multi-step reasoning tasks, ensuring accurate and reliable results. By simplifying the process and providing a user-friendly interface, the agent makes it accessible for non-specialist users, significantly reducing the time and effort required for manual calculations. This agent is particularly useful for scenarios where the floor height exceeds 2.2 meters, focusing on the most common and straightforward cases while laying the groundwork for more advanced applications in future developments. Figure 2 shows how the agent works.

(1) Planning

The planning component acts like the brain of the agent. It breaks down complex tasks into manageable steps and coordinates the execution of multi-step processes or queries using LLM. The planning component is essential for handling complex, multi-step user requests. At its core, CoT empowers the model to dissect complex problems into a series of more manageable, intermediate steps: First, to extract each floor's information in the given ifc file, which is commonly stored in "IFCSLAB" as this example:

```
#42252=IFCSLAB('0HupypOIXDjguK0hq5TsAf',#18,'X2\697C677F\X0\X2\5BA451855730576A\X0\120mm:310759',$, 'X2\697C677F\X0\X2\5BA451855730576A\X0\120mm',#42186,#42251,'310759',FLOOR.)
```

"0HupypOIXDjguK0hq5TsAf" in this example is the IfcGUID of the first-floor slab in the building case. We manually collect each floor's IfcGUID and input them to LLM.

Second, to extract the height of elevations, stored in ifc files as the examples:

```
#102=IFCBUILDINGSTOREY('3Zu5Bv0LOHrPC10026FoQQ',#18,'1F',$, 'X2\68079AD8\X0\X2\6B638D1F96F668079AD8\X0\#101,$,'1F',ELEMENT.,0.)
#106=IFCBUILDINGSTOREY('15Z0v90RiHrPC20026FoKR',#18,'2F',$, 'X2\68079AD8\X0\X2\4E0A68075934\X0\#105,$,'2F',ELEMENT.,4500.)
```

The numbers "0", "4500" at the end of each line are the height of elevations, which the agent can use to calculate the height of each storey. Third, to calculate the height of storeys. In this case the height of the first storey should be $4500 - 0 = 4500$ mm. Fourth, judge whether the height of storey is over 2.2 m. For example, 4.5 m is over 2.2 m. Fifth, to relate each storey and the floors they belong to. As we have already related each floor and its

IfcGUID, we can relate each storey height and the floor's IfcGUID. In this case the first floor's IfcGUID is "0HupypOIXDjguK0hq5TsAf", with the storey height of 4.5m. Sixth, to calculate the floor area whose storey is over 2.2m using Calculator tool. The detailed calculation process is explained in the next section.

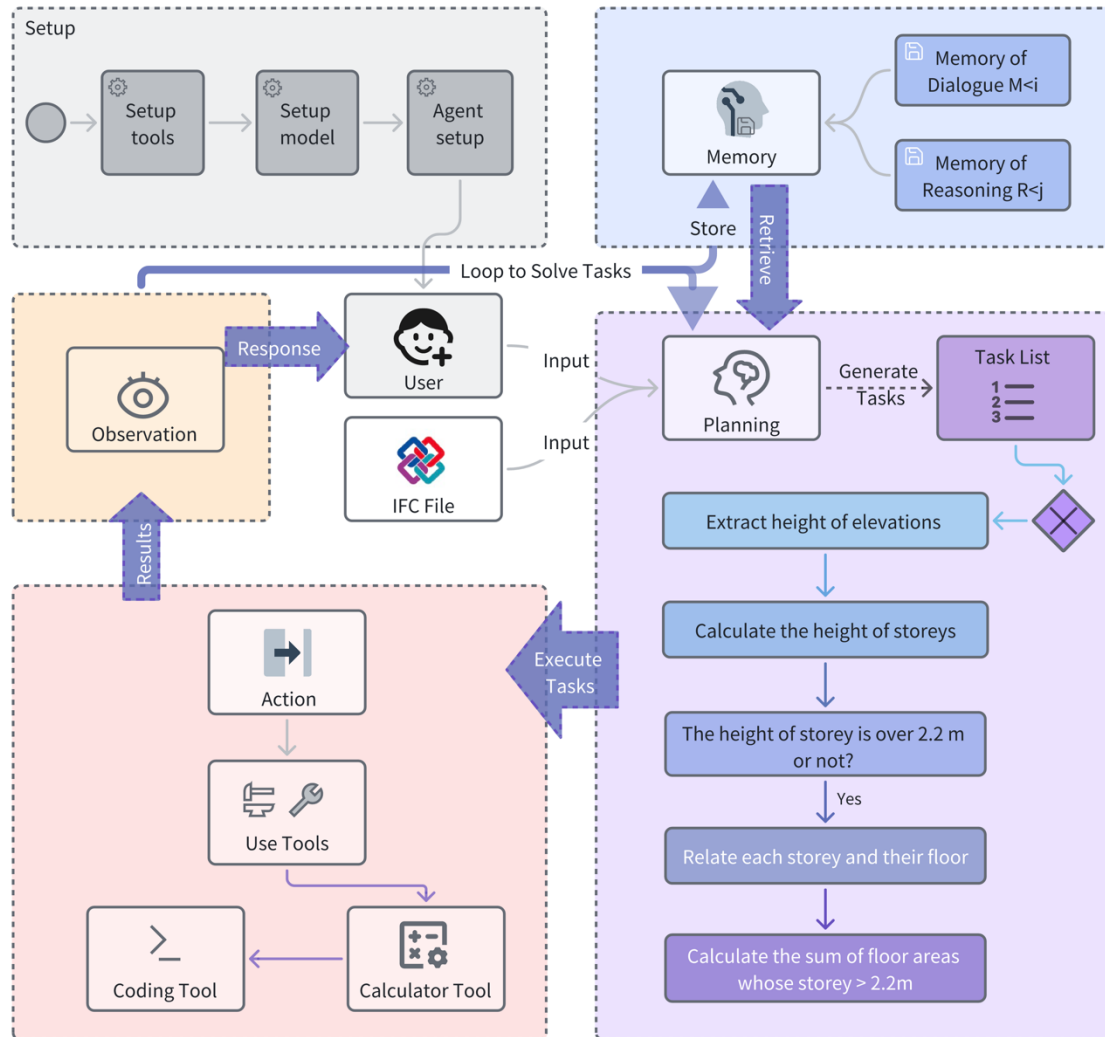


Figure 2. The framework of the proposed floor-area-calculation agent, which has four main components: planning, action, observation and memory.

(2) Action and Observation

Building upon this foundation, the Floor-Area-Calculation Agent incorporates the ReAct framework, which extends the model's capabilities beyond mere reasoning to include interactive decision-making. In the ReAct paradigm, the agent not only deliberates on the problem at hand but also takes decisive actions based on its analytical process. These actions may encompass querying additional information, making informed decisions, or engaging in clarifying dialogues to refine its understanding.

The Action component of the system incorporates two main tools to execute various tasks efficiently - Calculator Tool and Coding Tool. As illustrated in Figure 2, the system employs the Calculator Tool for performing numerical operations and the Coding Tool for generating and executing python codes using LLM to solve calculation tasks. These tools are strategically integrated to enhance the system's capabilities in processing complex queries that may require calculations or code generation. The presence of these diverse tools suggests a flexible and multifaceted approach to problem-solving, allowing the system to adapt to a wide range of user requirements and query types. The system can dynamically select and apply the appropriate tool based on the specific needs of each task, ensuring efficient and accurate execution of the planned actions.

Calculating Logic: Gauss's area formula is used to calculate the areas of floors:

$$S = 0.5 \times |\sum_{i=0}^{n-1} (x_i y_{i+1} - y_i x_{i+1})|, \quad (1)$$

where $x_n = x_0$ and $y_n = y_0$, which means the first and last points are connected. Equation (1) can be used to calculate the area of any simple polygon (one that does not intersect itself). This formula is also known as the Shoelace formula.

The geom module from the ifcopenshell library is used, which provides functionality for geometric operations and manipulations on IFC models, enabling the creation, modification, and analysis of 3D building models. Algorithm 1 shows the vectorized implementation of Gauss's area formula using the NumPy library for array operations. Here, the np.roll function cyclically shifts the elements of an array, so that the first element of the y array is multiplied by the last element of the x array, and so on, thus implementing the cross-multiplication in the aforementioned summation. Algorithm 1 is shown as follows:

Algorithm 1

```
# Set geometric parameters
settings = ifcopenshell.geom.settings()
settings.set(settings.USE_WORLD_COORDS, True)

# Get floor geometry data
shape = ifcopenshell.geom.create_shape(settings, slab)

# Extract the vertices of the geometry data
vertices = shape.geometry.verts

def calculate_polygon_area(vertices):
    """
    Calculate the area of a polygon, assuming it is a polygon in a 2D plane.
    """
    x = vertices[:,0] # Get the X coordinates (Every three vertices are X, Y, Z.)
    y = vertices[:,1] # Get the Y coordinates

    # Use the polygon area formula
    return 0.5 * np.abs(np.dot(x, np.roll(y, 1)) - np.dot(y, np.roll(x, 1)))

# Calculate floor area
area = calculate_polygon_area(vertices)
print(f" Floor area: {area} square meters ")
```

Figure 3 illustrates how a polygon floor slab is divided into triangles, with its area calculated using Gauss's area formula.

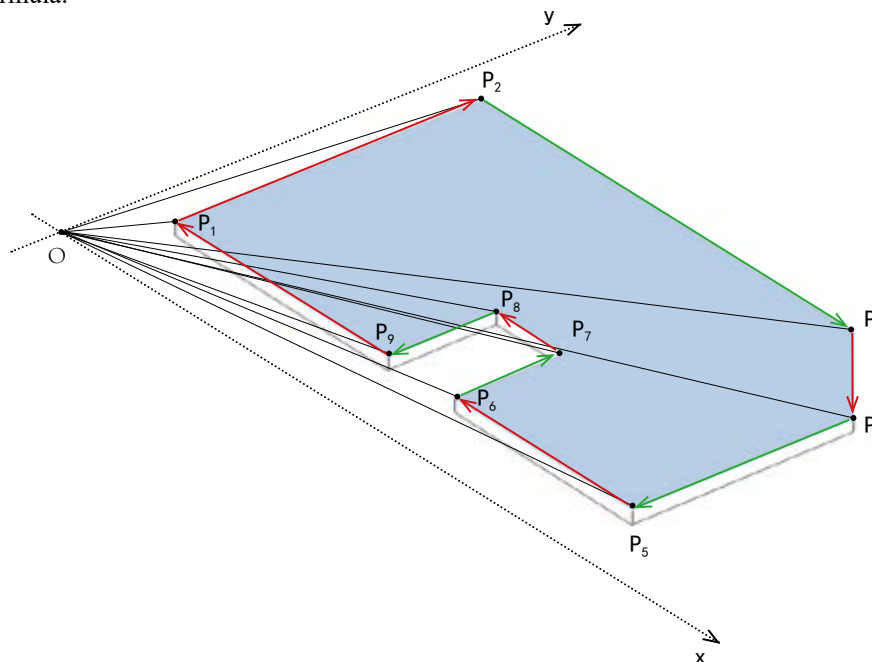


Figure 3. Triangle form of Gauss's area formula.
The color of the edges indicate, which triangle area is positive (green) and negative (red) respectively.

As for the Observation component, the primary purpose of it is to gather and analyze data from the environment or the task at hand. This data can include various types of information, such as input data, intermediate results, and feedback from previous actions. By thoroughly observing the current state, the agent can make informed decisions and adjust its strategy as needed until finally get the perfect answer.

(3) Memory

The Memory module is divided into two key sub-components: (a) Memory of Dialogue: This component stores the history of the conversation between the user and the AI system. (b) Memory of Reasoning: This sub-component retains the system's step-by-step reasoning process from previous interactions. The Memory module interacts with the Planning component through two processes: Store, which allows new information and reasoning steps to be added to the memory for future reference in the LLM's prompt input, and Retrieve, which enables the system to access stored information when planning new actions or responding to user queries. By maintaining this dual memory structure, the system can provide context-aware responses, build upon previous interactions, and apply learned reasoning strategies to new problems.

3. CASE STUDY

To further illustrate the effectiveness and versatility of our proposed framework, we conducted a case study involving the development of an academic building.

Figure 4 is the axonometric drawing of the academic building. The model consists of 3 floors. Each floor has been meticulously designed to accommodate a variety of room types with specific educational purposes. The overall height of the building stands at 12.3 meters, with the first floor measuring 4.5 meters, and both the second and third floors standing at 3.9 meters each. The skylight in the middle of the roof ensures optimal natural lighting and ventilation throughout the building. The meticulous planning of each floor's dimensions reflects a thoughtful approach to creating an environment conducive to learning and collaboration. Whether it be for lectures, seminars, or individual study sessions, this academic building is poised to serve as a hub of intellectual activity. As explained in Section 2.1, in this case study we only consider the calculations to the main floor slab areas and do not consider special areas such as balconies and stairs that lie outside the main structural envelope. Future in-depth studies will address these more complex scenarios.

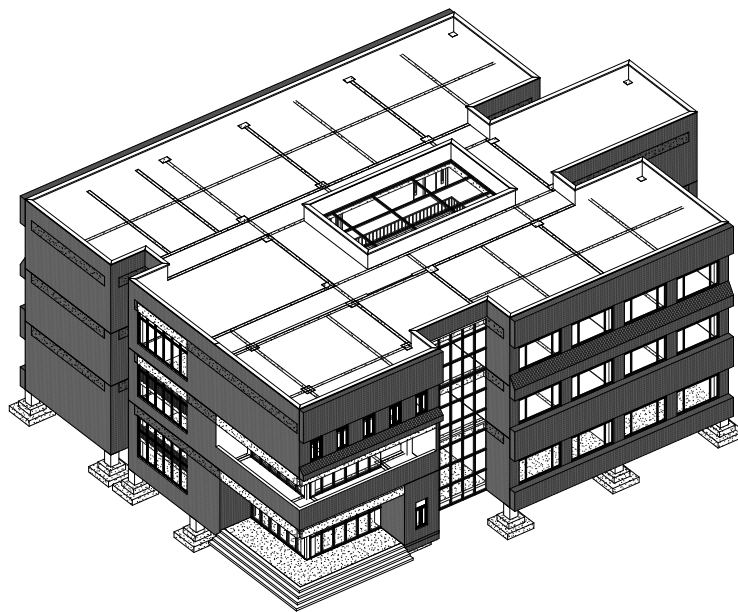


Figure 4. Axonometric drawing of the academic building, northeast view.

We asked the floor-area-calculation agent about the areas of each floor separately. The results are shown in Figure 5. For both single floor and gross area, the agent gave correct answers in the simplified case. In this case of the study, we only consider the simplest situation that is whether the storey height exceeds 2.2 meters. If it does, we calculate the full floor plate area for that level, without considering other factors that might affect the total floor area. We will delve deeper into these other factors in future research.

Question: What is the area of the second floor of the building?

Floor-Area-Calculation Agent Response

```
> Entering new AgentExecutor chain...
### Step 1: Extract the second floor's information
#42326=IFCSLAB('3$wwDIU0X6tgW97MkW7ISY',#18,'X2\697C677F
\X0\;X2\5BA45185697C677F\X0\;100mm:311742',
$, 'X2\697C677F\X0\;X2\5BA45185697C677F\X0\;100mm',#42294,
#42325,'311742',FLOOR.)

### Step 2: Extract the elevations
The elevations are:
- 4500
- 8400

### Step 3: Calculate the height of storey
- 8400 - 4500 = 3900 \text{mm}
...
### Step 4: Judge whether the height of storey is over 2.2 m
...
### Step 6: Calculate the floor area using calculator tool
...
782.77 \text{m}^2
### Conclusion
The area of the second floor is [782.77 m2].
> Finished chain.
```

Question: What is the gross area of the academic building?

Floor-Area-Calculation Agent Response

```
> Entering new AgentExecutor chain...
### Step 1: Extract the second floor's information
...
### Step 2: Extract the elevations
The elevations are:
- 0
- 4500
- 8400
- 12300
### Step 3: Calculate the height of storeys
Storey 1:
- 4500 - 0 = 4500 \text{mm}
Storey 2:
- 8400 - 4500 = 3900 \text{mm}
...
### Step 4: Judge whether the height of storey is over 2.2 m
...
### Step 6: Calculate the sum of floor areas whose storey > 2.2m
877.09 + 782.77 + 782.77 = 2442.63 \text{m}^2
### Conclusion
The total number of storey whose height is over 2.2 m is 3.
Therefore, the answer is [2442.63 m2].
> Finished chain.
```

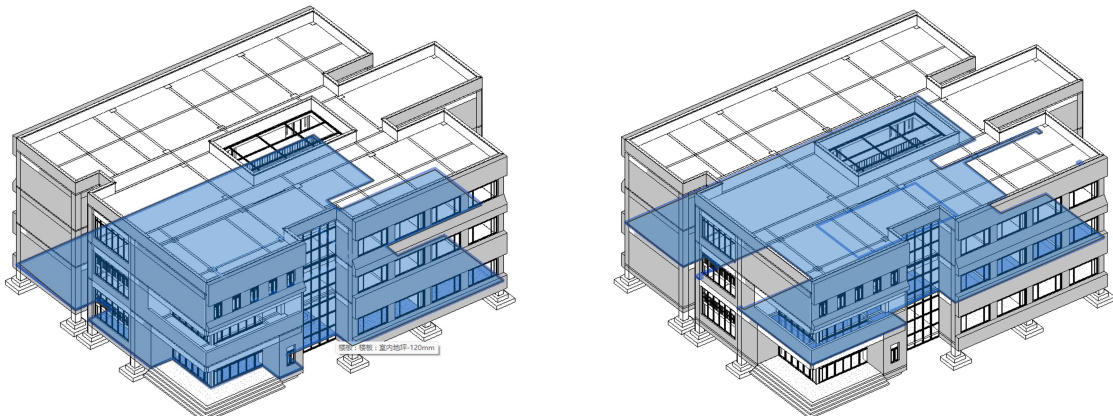
Figure 5. The results of two questions about the floor area and gross area of the academic building separately.

Table 1 presents the test results of the floor-area-calculation agent across different floors. For the first floor (1F), the actual area measured 868.73 square meters, while the calculated area was slightly higher at 877.09 square meters, resulting in an error rate of 0.96%. On the second floor (2F) and third floor (3F), both had identical actual areas of 796.19 square meters, with calculated areas of 782.77 square meters, leading to an error rate of 1.69% for both floors. Overall, when considering all floors combined, the total actual area was 2461.11 square meters, compared to a calculated area of 2442.63 square meters, yielding a total error rate of 0.75%. At this stage, the algorithm is still quite simplified, which leads to a certain degree of error in the calculation results. From the test results, the error rates are greater for the second and third floors, which may be due to the irregular shapes caused by the cutouts for the stairwells on these levels. The next steps in the research will involve a deeper investigation into the causes of these errors and further refinement of the floor area calculation method.

Table 1. Test results of the floor-area-calculation agent

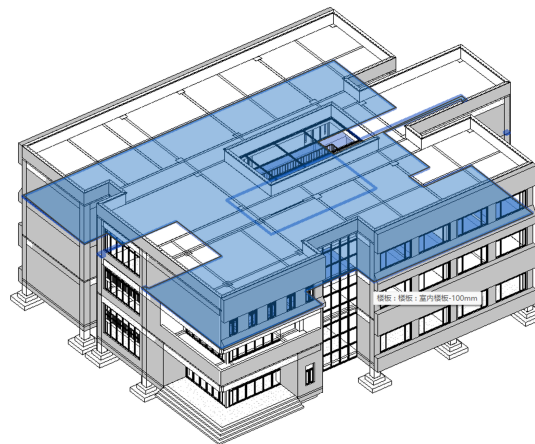
	Ground Truth Area (m ²)	Calculated Area (m ²)	Error Rate (%)
1 F	868.73	877.09	0.96
2 F	796.19	782.77	1.69
3 F	796.19	782.77	1.69
Total	2461.11	2442.63	0.75

Figure 6 (a), (b) and (c) respectively show the shapes of the first, second, and third floor slabs of this academic building, to help readers better understand the case. Areas with special calculation rules, such as balconies, are not specifically considered in half size in this study.



(a) The first-floor slab of the case.

(b) The second-floor slab of the case.



(c) The third-floor slab of the case.

Figure 6. The shapes of the first, second, and third floor slabs of this academic building in northeast view.

4. DISCUSSION

The proposed LLM agent for semi-automated floor area calculation has demonstrated significant potential in improving the efficiency and accuracy of the process. By leveraging CoT planning and the ReAct framework, the agent is capable of handling complex, multi-step reasoning tasks and ensuring the accuracy of the results. The empirical validation on the academic building showed positive outcomes, validating the effectiveness of the approach.

One of the key advantages of the LLM agent is its user-friendly interface, which makes it accessible to non-specialist users. Traditional methods and software tools often require advanced technical skills, which can be a barrier to widespread adoption. By simplifying the process and providing clear, understandable outputs, the LLM agent can be easily integrated into the workflow of architects and engineers, thereby reducing the learning curve and increasing productivity.

While the initial results are promising, there are several limitations that need to be addressed in future research. One limitation is the current focus on the simplest case, where the floor height exceeds 2.2 meters. Future work will involve expanding the agent's capabilities to handle more complex scenarios, such as the calculation of areas with varying heights, special features like stairwells, and other factors that affect the total floor area.

What's more, for slabs that have holes in them, which is a very common situation in real projects, the agent should be able to distinguish between the vertices of the holes and the vertices of the slab edges, and then calculate the correct area. This is also an issue to be addressed in the next steps of the research.

Finally, the reliance on LLMs introduces inherent biases and potential inaccuracies. The performance of the LLM agent in larger and more complex building projects needs to be evaluated to ensure its robustness and scalability. Future research should also explore the robustness and scalability of using AI in the AEC industry and develop guidelines to address these concerns.

5. CONCLUSIONS

The potential benefits of the developed tool are substantial, offering several advantages over existing software solutions. Unlike traditional methods that require specialized knowledge and can be cumbersome to use, our LLM agent simplifies the workflow, making it more accessible to non-expert users. The system's ability to interpret and process IFC files, combined with its natural language interaction capabilities, significantly lowers the technical barriers associated with BIM data manipulation.

This tool has the potential to transform the AEC industry by enabling more efficient and accurate floor area calculations, which in turn can lead to better-informed decision-making, improved project cost estimation, and enhanced productivity. In practical application, the tool can significantly reduce the time architects and engineers spend on area calculations, allowing them to focus more on creative design and other high-value activities. Additionally, for project management and cost estimation, having quick and accurate floor area data is crucial, and this tool can help enhance overall productivity and efficiency in the AEC industry.

As technology advances and research deepens, it is anticipated that the LLM agent will further optimize its performance to address a broader spectrum of scenarios and challenges. Future enhancements may extend the tool's capabilities to handle calculations involving complex features such as floor slabs with holes, curved surfaces, and distinct areas like balconies and stairs that lie outside the main structural envelope, as well as regions with height variations. Additionally, the tool could adapt to align with diverse regional or national regulations. Such advancements would render the LLM agent a more comprehensive and flexible solution, driving the

standardization and automation of BIM workflows and setting new benchmarks for efficiency and accuracy in the AEC industry.

In conclusion, this LLM-based semi-automated floor area calculation system not only improves the efficiency and accuracy of area calculations in AEC industry by reducing reliance on manual calculations, which are often time-consuming and prone to human error, but also showcases significant industry potential with its user-friendliness and flexibility. As it iterates and improves, the tool is poised to revolutionize how floor area calculations are performed in the AEC industry, promoting overall workflow efficiency.

REFERENCES

- Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., & Anadkat, S. (2023). Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.
- Lee, J.-K., Cho, K., Choi, H., Choi, S., Kim, S., & Cha, S. H. (2023). High-level implementable methods for automated building code compliance checking. *Developments in the Built Environment*, 15, 100174.
- Li, Y., Wang, S., Ding, H., & Chen, H. (2023). Large language models in finance: A survey. Proceedings of the fourth ACM international conference on AI in finance,
- M. Bran, A., Cox, S., Schilter, O., Baldassari, C., White, A. D., & Schwaller, P. (2024). Augmenting large language models with chemistry tools. *Nature Machine Intelligence*, 1-11.
- Özer, S., & Jacoby, S. (2022). Dwelling size and usability in London: a study of floor plan data using machine learning. *Building Research & Information*, 50(6), 694-708.
- Reinhardt, J., & Mathews, M. (2017). The automation of BIM for compliance checking: a visual programming approach.
- Roanes-Lozano, E., Laita, L., Roanes-Macías, E., Maojo, V., Corredor, S., De La Vega, A., & Zamora, A. (2000). A Gröbner bases-based shell for rule-based expert systems development. *Expert Systems with Applications*, 18(3), 221-230.
- Sang, K. J., Ison, S. G., & Dainty, A. R. (2009). The job satisfaction of UK architects and relationships with work - life balance and turnover intentions. *Engineering, Construction and Architectural Management*, 16(3), 288-300.
- Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q. V., & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, 35, 24824-24837.
- Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., & Cao, Y. (2022). React: Synergizing reasoning and acting in language models. *arXiv preprint arXiv:2210.03629*.
- Zheng, Z., Chen, K.-Y., Cao, X.-Y., Lu, X.-Z., & Lin, J.-R. (2023). Llm-funcmapper: Function identification for interpreting complex clauses in building codes via llm. *arXiv preprint arXiv:2308.08728*.

Heritage Building Information Management and Intelligent Querying by Multimodal Large Language Models and Knowledge Graph

Jiaying ZHANG¹, Jeff Chak Fu CHAN², Ziyu ZHAO³, Jack C P CHENG^{4*}

¹Postdoctoral Fellow, Department of Civil and Environmental Engineering, Hong Kong University of Science and Technology, Hong Kong, China (corresponding author). Email: jzhangfh@connect.ust.hk

²M.Phil. Student, Department of Civil and Environmental Engineering, Hong Kong University of Science and Technology, Hong Kong, China. Email: cfchanay@connect.ust.hk

³Ph.D, Department of computer science and software engineering, the University of Western Australia, Australia. Email: zhaoziyu929@126.com

* Corresponding author. Professor, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, China. 999077. Email: cejcheng@ust.hk.

Abstract: Heritage buildings face challenges in documentation due to inconsistent records and complex data from historical documents, archaeological surveys, and materials. Traditionally, converting unstructured data into structured formats required significant expert effort. The advent of large language models (LLMs) has transformed heritage research by enabling the creation and maintenance of knowledge graphs. These graphs integrate diverse data sources, facilitating the preservation and study of heritage buildings. LLMs help extract and organize unstructured data, improving knowledge graph accuracy and consistency. This research proposes a comprehensive framework that integrates multimodal data, including text, images, and videos, into a unified knowledge graph. The framework employs LLMs for extracting information from textual data, the CLIP model for aligning images with corresponding text, and keyword searches for processing video content. The resulting knowledge graph is stored in a Neo4j graph database, providing an interactive platform for users to query and explore detailed information about heritage buildings. This approach not only supports academic research but also contributes to practical applications in cultural heritage conservation, enabling more efficient access to valuable information and enhancing preservation efforts. The proposed method was validated in European 'Gothic' and 'Gothic Revival' architecture by comparing the relationships between components.

Key words: Heritage Building, Information Management, Large Language Models, Knowledge Graph

1. Introduction

Heritage buildings have gone through different uses and eras, as well as multiple periods of changes and advancements in technologies and regulations [1]. Information on these buildings comes from historical documents, archaeological surveys, building materials, academic research, and more, often involving inconsistent and complex records. For example, structural design, construction techniques, and materials recorded in ancient architectural documents may differ from the current state of the buildings. Traditionally, converting unstructured data into structured formats required significant human effort, particularly for heritage buildings, where domain experts had to process the data [2]. Previously, acquiring knowledge about heritage buildings required extensive learning and literature review [2]. Khalil et al [3] indicates that the documentation of cultural heritage faces additional challenges, primarily due to the varying sizes of the documented objects and the differing requirements for quality and resolution. The purpose of documentation is to make information accessible to other users. Challenges include stability, surface damage, variations in scale and precision requirements, time constraints, and the involvement of

multi-sensor and multi-resolution data. The documentation of cultural heritage is not an end but aims to provide information to others. Patias et al [4] demonstrated that main difficulties in generating traditional knowledge graphs include the complexity of data acquisition and integration, challenges in information extraction and semantic understanding, technical issues in knowledge representation and storage, and the accuracy of relationship extraction and entity linking. Additionally, maintaining the dynamic updating and reasoning capabilities of knowledge graphs poses demands on their application effectiveness and reliability.

The rise of large language models (LLMs) has brought significant changes to how we study and disseminate knowledge of ancient architecture. Previously, learning about heritage buildings required extensive literature reviews. Documentation of cultural heritage faces challenges, including varying object sizes, damage, time constraints, and multi-sensor data. The purpose of documentation is to provide accessible information to users, though challenges remain in maintaining accuracy and consistency. LLMs, through deep learning and pre-training, have addressed many natural languages processing challenges, including language understanding, multitasking, and cross-domain applications. They offer new opportunities for constructing, updating, and reasoning through knowledge graphs, which can enhance the management and application of historical building data. A knowledge graph integrates various data sources such as historical literature and multimedia resources, creating a comprehensive database for preservation and restoration efforts. Using LLMs to build a knowledge graph for historical building components allows for effective extraction and organization of information from unstructured data. This process involves data collection, preprocessing, information extraction, and continuous updating. The resulting knowledge graph, stored in a graph database, enables efficient querying and reasoning, supporting the research and preservation of historical buildings. The graph also aids cultural heritage protection, providing valuable insights for tourism, smart guide systems, and personalized travel experiences. By integrating various data sources and maintaining consistency, the system ensures high-quality information that supports both academic and practical applications in building restoration.

2. Proposed method

In this research, we present a framework, which integrates multimodal data to enhance structured representations. Firstly, we use LLM via prompt engineering to extract informative triplets. Meanwhile, the CLIP model is applied to align images with given text input, as well as video data via key word search. It constructs a knowledge graph which is stored in neo4j graph database and can be interacted with via an interactive LLM-based chat agent. Ultimately, the goal is to develop an LLM-based chatbot interface that leverages this enriched knowledge graph to provide users with sophisticated query capabilities. The proposed method is shown in Figure 1.

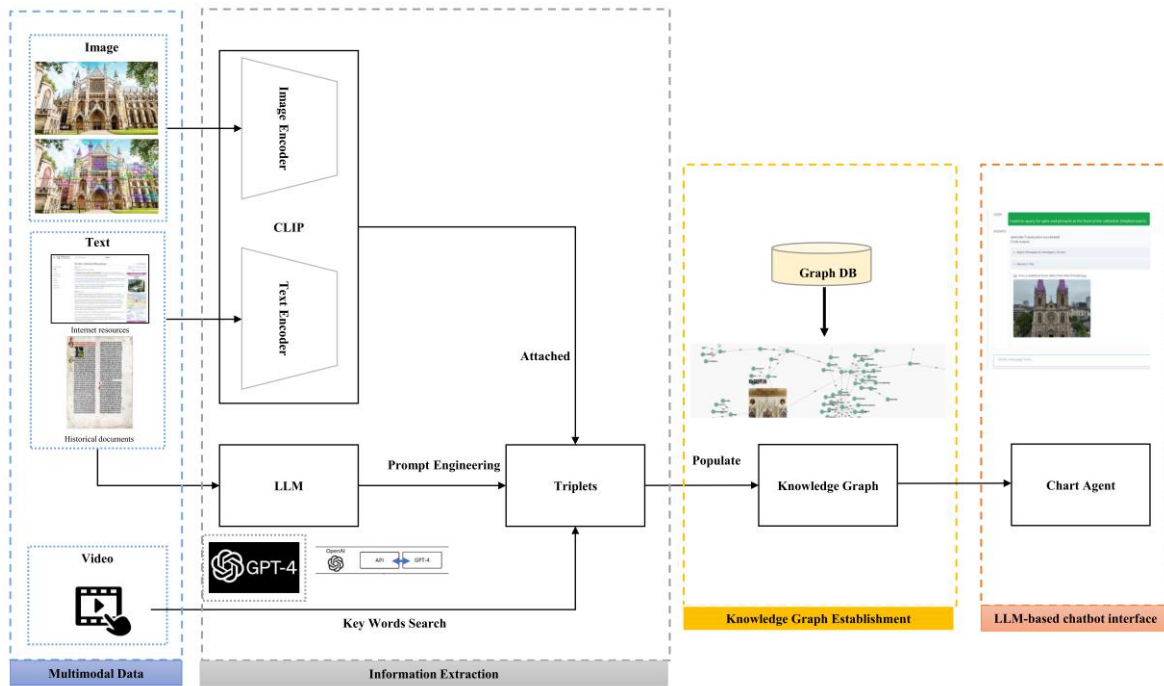


Figure 1. The proposed method

1.1 Database Generation and knowledge graph

A knowledge graph can clearly express the hierarchical relationship of architectural heritage knowledge [5, 6]. Building a knowledge graph of Gothic architecture using large language models starts with collecting textual data about Gothic architecture, including its stylistic features, origins, representative buildings, development history, and especially semantic information about architectural structures. Secondly, the collected textual data is processed and analyzed to extract key information and knowledge points. Entities and relationships are established in the knowledge graph, organizing the various aspects of Gothic architecture into structured knowledge graph data and identifying their intrinsic connections. Conventional knowledge graph construction requires a large amount of textual data for training. This study proposes a large language self-learning model that can accurately present the target building with just a small set of data samples. Thirdly, as the cultural significance and symbolic features behind ancient architecture have a long history and evolve over time, with the emergence of new content and the construction of new buildings imitating ancient styles, it is necessary to continue updating and optimizing the knowledge graph data to ensure timeliness and accuracy. Combining user feedback and requirements, continuously improving the retrieval efficiency of the knowledge graph, and enhancing user experience.

1.2 Querying ability of the LLM-based assistant

Semantic recognition of heritage building components can be very challenging for non-experts, but AI large language models (LLMs) provide a dialogue interaction agent for building digital models of heritage structures. The digital twin agent can engage in conversation with users, who only need to input vague query instructions or capture images of building components, and the agent will provide accurate and relevant information, as well as query the building's background and historical context. Through interaction with the AI large language model, users can quickly retrieve information about the historical background, architectural style, and cultural significance of heritage buildings, greatly improving the documentation and querying capabilities of components in heritage

digital twins. The query capability of LLMs is a core advantage in data management, information retrieval, and decision support, and it manifests in several key ways: first, LLMs support natural language queries, allowing users to express their queries in everyday language, thus lowering the technical barrier and accommodating various query expressions. Second, LLMs can process complex multi-condition queries and fuzzy queries, helping users find data that is not an exact match. Third, LLMs provide intelligent recommendations and auto-completion features, offering real-time query suggestions and automatically completing queries to improve efficiency. Fourth, LLMs possess strong contextual understanding, comprehending the semantics of queries rather than just keyword matching, and can retain context throughout a conversation to support continuous queries. Finally, LLMs support multilingual queries, making them useful in cross-lingual environments for data query and management. In practical applications, users can query the historical background, restoration records, and current status of heritage buildings using natural language, for example, "query the restoration history of Notre-Dame Cathedral" or "find all 18th-century wooden structures," which greatly facilitates the querying and management of heritage building data.

3. Illustrative example

Using the extremely famous Gothic architectural style as the training theme, this research has sorted out the names, characteristics, and interrelationships of components in European 'Gothic' and 'Gothic Revival' architecture. We have organized 20 illustrated documents as training materials for entity extraction, highlighting the differences between the two styles. The comparison of Gothic and Gothic Revival are summarized in Table, examples of the relationships between components of Gothic architecture extracted from the text are listed in Table.

Table 2: The comparison of Gothic and Gothic Revival

Name of Components	Gothic	Gothic Revival
Pointed Arch	Basic load-bearing element, sharp and towering	Serve as a decorative element, with more varied forms
Ribbed Vault	Light and stable, supporting large spaces	Be simplified or use modern materials and techniques
Flying Buttress	Important load-bearing structure, enhancing stability	Be smaller or absent, depending on design requirements
Rose Window	Iconic decorative element, intricate and ornate	Be more simplified or have more complex decorations
Building Materials	Primarily stone and glass	Incorporate a variety of materials
Overall Style	Mysterious, solemn, and soaring	Retro, diversified, and possibly incorporating innovative designs

Table 3: Examples of the relationships between components of Gothic architecture extracted from the text

Entity 1	Entity Type 1	relationship	Entity 2	Entity Type 2
Pointed arches	component	Bear	Buttresses	component
Buttresses	component	prop up	Fly buttresses	component
Stained glass	component	Installed on	walls	Architectural parts
Beam columns	component	prop up	vault	component
Spire	component	connect	The main body of the building	building
Sharp coupons	component	cooperate	Clover arch	Decorative components
Clover arch	Decorative components	ornament	Four-leaf windows	Decorative components
Pointed arches	component	belong	Gothic architecture	Construction period
Spire	component	belong	Gothic Revival	Construction period
Fly buttresses	component	prop up	Nave vault	Architectural parts
Ribbed vaults	component	prop up	Chapel space	Architectural parts
Small minarets	Decorative components	ornament	Fly buttresses	component

3.1 Database Generation and knowledge graph

Gothic architectural elements not only have specific names with literary connotations but also interact with one another and serve distinct functions. For example, the flying buttress is an external structure used in Gothic architecture to support the vaulted ceilings. Through a series of diagonal supports, it transfers the weight of the roof to the foundation, reducing the pressure on the interior columns. The introduction of flying buttresses allowed Gothic cathedrals to have taller and more open spaces, while also giving architects more freedom to incorporate elements like pointed arches and rose windows into their designs. Ribs are one of the most iconic features of Gothic architecture. They consist of a series of arched structures that distribute the weight to the walls and columns, allowing the building to have higher spaces and lighter structures. The ribs are typically in the shape of pointed arches, which are not only aesthetically pleasing but also capable of supporting larger spans, making the interior of the cathedral more spacious and well-lit. The relationship of building components of heritage building with Gothic and Gothic Revival style are investigated in the knowledge graph (Figure 2a-c).

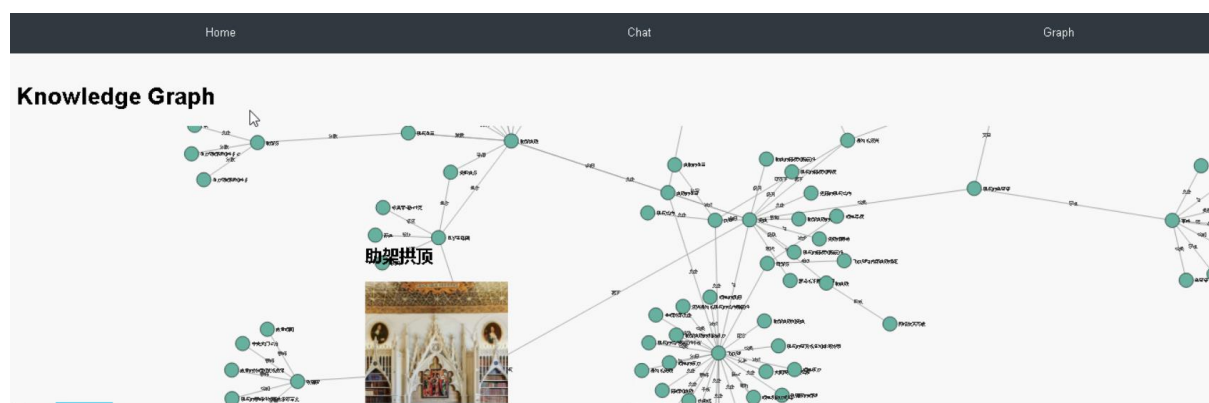


Figure 2a. Knowledge graph of building components

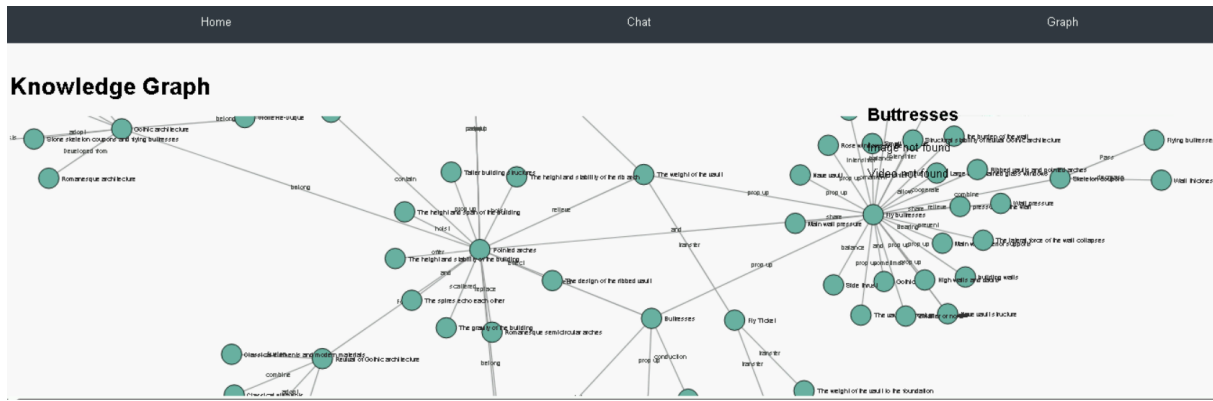


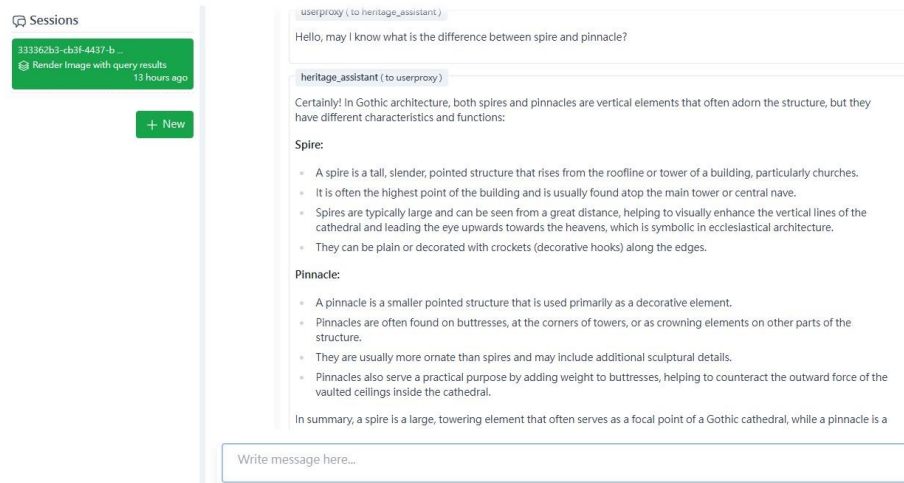
Figure 2b. Knowledge graph of building components



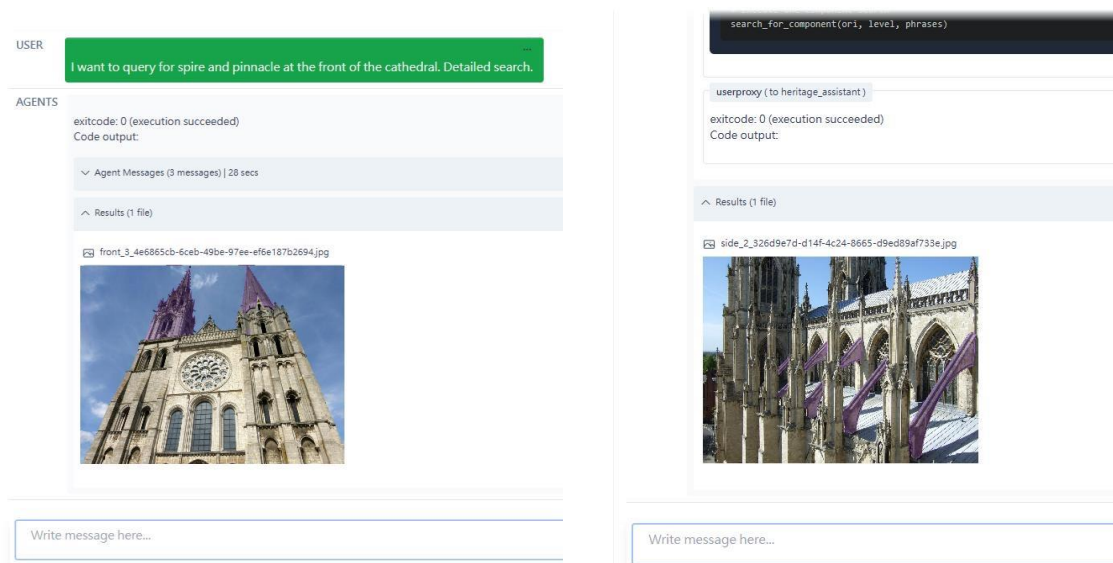
Figure 2c. Knowledge graph of building components

3.2 LLM-based Intelligent Querying

The visualization of querying related to components within the LLM-based chatbot interface of the application was demonstrated in Figure 3. The interface was built via Autogen Studio and the LLM heritage assistant was based on GPT-4 with specific system message for instruction. We first asked the heritage assistant a basic question about the difference between “spire” and “flying buttress”, and the assistant could answer correctly as shown in Figure 3 (a). It showed that the assistant can correctly explain the meaning of both components and differentiate their characteristics and functions. The LLM assistant was provided with the details of the cosine similarity search function on language features in the backend, so that it could provide suitable input parameters to the system given the user’s query. It can identify and extract the wanted view and semantic level of the model and the search keywords from user’s natural language query, and pass them as input parameters to the function so that the system can execute the function correctly and return an annotated image with query result, as shown in Figure 3 (b). Not only can the LLM assistant extract clear search keywords from the user’s query, but it can also transform vague words in the query like “big round window” and “flying arched column” into standard component names like “rose window” and “flying buttress” to conduct the similarity search. Through a simple demonstration application, it is shown that an LLM-based assistant could improve the documentation and querying ability of components in the heritage, facilitating user-friendly navigation and information retrieval of the heritage.



(a). The basic question about the difference between “spire” and “pinnacle”



(b). Query results of the “spire” and “flying buttress” at the front and side views

Figure 3. Question and answering in the chatbot interface regarding the previous image

4 Conclusions

This research proposes a framework that integrates multimodal data to enhance the structured representation of heritage building knowledge. The framework leverages Large Language Models (LLMs) with prompt engineering to extract informative triplets from textual data. In parallel, the CLIP model is employed to align images with corresponding text inputs, and video data is processed using keyword searches to further enrich the dataset. This multimodal data is then used to construct a comprehensive knowledge graph, which is stored in a Neo4j graph database. The knowledge graph is designed to be interactive, with users able to query and explore the data through an LLM-based chatbot interface. This chatbot allows users to make sophisticated queries and retrieve detailed information about historical buildings, supporting both academic research and practical applications in cultural heritage conservation. The goal of this framework is to enable seamless access to enriched knowledge about heritage buildings, facilitating advanced data exploration and enhancing user experience.

5 References

- [1] J. C. Cheng, J. Zhang, H. H. Kwok, J. C. Tong, Thermal performance improvement for residential heritage building preservation based on digital twins, *Journal of Building Engineering* 82 (2024) 108283.
- [2] A. P. News, Next-Gen Digital Twin and AI Breathe New Life into Architectural Heritage Preservation: Innovative Research Released, 2024. <https://apnews.com/press-release/ein-presswire-newsmatics/cultural-preservation-hong-kong-68b0a82c70d9103ff8d563dfc6fd7471>.
- [3] A. Khalil, S. Stravoravdis, D. Backes, Categorisation of building data in the digital documentation of heritage buildings, *Applied Geomatics* 13 (2021) 29-54.
- [4] P. Patias, Cultural heritage documentation, International Summer School Digital Recording and 3D Modeling, Aghios Nikolaos, Creta, Grecia, Abril (2006) 24-29.
- [5] Q. Fan, Z. Shi, G. On, Construction of knowledge graph of Intangible Cultural Heritage, *Libr. Forum*, 2021, pp. 100-109.
- [6] X. Zhang, Y. Zhi, J. Xu, L. Han, Digital protection and utilization of architectural heritage using knowledge visualization, *Buildings* 12(10) (2022) 1604.

BRIDGE INSPECTION USING A MULTI-MODAL VISION LANGUAGE MODEL

Zhengxing Chen¹, Yang Zou², Vicente A. González³, Jason Ingham⁴, Liam M. Wotherspoon⁵

1) Ph.D. Candidate, Department of Civil and Environmental Engineering, University of Auckland, Auckland, New Zealand. Email: zche570@aucklanduni.ac.nz

2) Ph.D., Senior Lecturer, Department of Civil and Environmental Engineering, University of Auckland, Auckland, New Zealand. Email: yang.zou@auckland.ac.nz

3) Ph.D., Professor, Department of Civil and Environmental Engineering, University of Alberta, Edmonton, Canada. Email: vagonzal@ualberta.ca

4) Ph.D., Professor, Department of Civil and Environmental Engineering, University of Auckland, Auckland, New Zealand. Email: j.ingham@auckland.ac.nz

5) Ph.D., Professor, Department of Civil and Environmental Engineering, University of Auckland, Auckland, New Zealand. Email: l.wotherspoon@auckland.ac.nz

Abstract: Using an Unmanned Aerial Vehicle (UAV) in bridge inspections can reduce human involvement in complex and hazardous inspection environments and automate the inspection process. Current practices require human operators to define task objectives, oversee safe flight operations, and evaluate bridge conditions. There is a growing demand for improving the seamless collaboration between UAVs and human inspectors to complete the inspection task efficiently and more safely, especially in post-disaster scenarios where critical bridges and other infrastructure facilities need to be inspected within hours or days. A significant gap exists in enabling UAVs to intelligently perceive and understand the bridge inspection scene according to human instructions. An intuitive human-UAV collaboration system using a multi-modal Vision Language Model (VLM) was proposed to partially fill this gap. This system leverages a few-shot Contrastive Language–Image Pretraining (CLIP)-based model to enable UAVs to visually and semantically understand the bridge inspection environment based on human commands. By incorporating text prompt learning with a cache adapter, the proposed model enhances the ability of CLIP to interpret both textual and visual inputs in the context of bridge inspection. The model was trained and evaluated in a bridge inspection image dataset and achieved an accuracy of 83.33%, outperforming other few-shot image classification methods, demonstrating its effectiveness in the bridge inspection domain. This approach is expected to improve collaboration between AI-empowered UAVs, inspectors, and bridge environments, thereby enhancing the overall efficiency of bridge inspections.

Keywords: Bridge inspection, Unmanned Aerial Vehicle (UAV), Human-robot collaboration (HRC), Natural language interaction, Vision language model (VLM)

1. INTRODUCTION

Bridges constitute essential elements of the transportation infrastructure, and the maintenance of bridges is imperative for safeguarding public safety. Assessing the structural health of bridges is particularly problematic because the inspection processes are mainly conducted manually (Zhang et al., 2022). The traditional inspection of bridges is particularly challenging because many bridge structures are either too costly or too dangerous for human inspectors to access directly (Dorafshan & Maguire, 2018). Unmanned aerial vehicles (UAVs) have been introduced in recent years for bridge inspection, aiming to enhance efficiency, reduce labour demand, and minimize human exposure to hazardous environments (Bolourian & Hammad, 2020). Current practices still require significant manual effort because UAVs cannot perform inspections independently without human input and skilled pilots. There is a growing demand for a collaborative environment where human operators and UAV robotic systems work side by side during bridge inspections. When faced with emergencies such as post-disaster bridges, emergency inspection and repair are essential ways to restore bridge transportation capacity quickly. There is a growing need of the collaboration between inspectors and UAVs to enhance the efficiency, safety, and effectiveness of bridge inspections in complex and challenging conditions.

Human-robot collaboration (HRC) is defined as the integration of human adaptability and decision-making capabilities with the physical precision, strength, and repeatability of robotic assistants to achieve common goals efficiently within shared workspaces (Ajoudani et al., 2018; Michalos et al., 2014). Current methods for controlling and interacting with UAVs in the physical world have been dominated by complex teleoperation controllers (Seo et al., 2018), hand gestures (Naseer et al., 2022), and rigid command protocols (Contreras et al., 2020), where the robots execute predefined tasks based on specialised programming languages. Among these methods, natural language-based HRC stands out for its intuitive and accessible nature. Natural language-based HRC allows non-experts in robot programming to intuitively communicate with robot assistants, making the interaction efficient and accessible (Park et al., 2024). Traditional natural language processing (NLP) models are usually trained on a limited dataset that shows limited adaptability in diverse working environments and with different inspectors. The application of HRC to bridge inspections comes with unique challenges and specifications. The inherent uncertainties and complexities in structural deterioration and failure contribute to varying

probabilities and consequences of failure across different bridges. The advent of Large Language Models (LLMs), such as ChatGPT (OpenAI, 2020), offers the potential to develop an interactive and communicative approach to HRC. LLMs trained on extensive and diverse datasets bring a deep understanding of natural language and human intentions, especially in the context of HRC tasks.

While LLMs excel in understanding human interaction commands, they lack the visual and semantic comprehension needed to interpret the environment around robots. Previous research has primarily focused on using deep learning-based object detection algorithms, such as You Only Look Once (YOLO) (Redmon et al., 2016), Faster R-CNN (Ren et al., 2015), and Single Shot MultiBox Detector (SSD) (Liu et al., 2016), to enhance the scene understanding of robots through image processing. Applying these algorithms in bridge inspections presents challenges, including the high costs of data annotation, significant computational demands, and the scarcity of large-scale training datasets (Liang et al., 2024). There is a need for computer vision algorithms that can operate effectively with minimal training samples in bridge inspection tasks. With the further development of LLMs and the increasing demand for integrating multiple modalities such as language and vision, Vision-Language Models (VLMs) have emerged (Zhou et al., 2022b). VLMs facilitate open-vocabulary visual recognition and make complex inferences about interactions between objects and agents within images (Kirillov et al., 2023). VLMs provide “eyes” to explore and find arbitrary objects described by humans, understand the environments, and provide context for decision-making. VLMs such as Contrastive Language–Image Pretraining (CLIP) (Pan et al., 2022), DALL-E (Ramesh et al., 2021), and Vision-and-Language BERT (ViLBERT) (Lu et al., 2019) have demonstrated strong zero-shot image classification performance on public datasets as a result of extensive training on large-scale image-text pairs. VLMs allow robots to navigate to and identify objects they have not been explicitly trained on, enhancing their adaptability and performance in unstructured and unforeseen scenarios (Gadre et al., 2022). Although the robotic research community has integrated VLMs for robotic control, these systems are primarily used for industrial applications (Shibata et al., 2024) or household services (Brohan et al., 2023). There is a need to develop a low-cost, highly efficient human-UAV collaboration method that enables UAVs to perform semantic scene understanding based on human instructions for UAV-assisted bridge inspections.

In early 2021, OpenAI released a large-scale multi-modal model for aligning images and texts called CLIP (Pan et al., 2022), which was trained on over 400 million image-text pairs. CLIP exhibits powerful zero-shot inference capabilities that can recognise and understand unseen images without being explicitly trained on specific tasks or datasets. The core concept of the CLIP model is to embed both text and images into a shared semantic space, where related text descriptions and image representations are positioned closely together. In contrast unrelated image-text pairs are placed farther apart. The CLIP model comprises two main components: an image encoder and a text encoder. The image encoder converts images into feature vectors, while the text encoder, typically a Transformer model, converts text into feature vectors. These two encoders operate within the same vector space, enabling cross-modal information interaction and fusion. CLIP has demonstrated excellent zero-shot performance on public datasets and performs effectively in many everyday tasks. Due to regulatory constraints, bridge inspection datasets are highly domain-specific and often unavailable on the Internet. The zero-shot capabilities of CLIP may be limited because it has not been trained in the context of bridge inspection. This highlights the need to adapt CLIP for this domain through transfer learning. Completely retraining CLIP for bridge inspection poses several challenges: 1) Due to the large number of parameters in CLIP, fine-tuning the entire network requires substantial computing resources. 2) UAV-assisted bridge inspection is an emerging technology that has not yet been widely adopted, resulting in a limited dataset that cannot cover all inspection object categories. Therefore the data distribution in bridge inspection often diverges from the pre-training data of CLIP. 3) Labelling bridge inspection data is time-consuming and labour-intensive, requiring domain expertise, further complicating the process. To address these challenges, it is essential to leverage few-shot learning techniques. By fine-tuning the CLIP model with a limited number of labelled samples, domain-specific knowledge can be transferred, enabling CLIP to perform well in bridge inspection tasks despite the scarcity of training data.

The authors propose a few-shot CLIP-based model to enable UAVs to visually and semantically understand the bridge inspection environment based on human commands. First a few-shot CLIP model integrated with text prompt learning and a cache adapter was proposed to enable UAVs to visually and semantically interpret the bridge inspection environment based on human commands. Second a bridge inspection dataset, collected by UAVs, was developed to test the proposed model. This dataset includes four bridge components: pier, girder, railing, and pavement; four structural details: bearing, cover plate termination, gusset plate connection, and out-of-plane stiffener; and two types of damage: cracks and corrosion. To evaluate the performance of the proposed method in bridge inspection tasks, the accuracy of the proposed model was compared with baseline CLIP and other fine-tuned CLIP-based models.

2. METHOD

The core objective of this research was to develop an algorithm that enables UAVs to navigate toward specific target objects within an unknown bridge environment. Figure 1 shows the framework of the proposed human-UAV collaboration method. Successful navigation requires UAVs to possess semantic scene understanding

and natural language processing capabilities. The capabilities allow UAVs to identify objects in the environment based on task goals defined by human inspectors and translate these goals into a semantic context in textual form. To achieve this goal the authors leverage the CLIP model, which is a VLM that facilitates cross-modal understanding by learning to compare text and images. CLIP captures the semantic relationships between text and images through contrastive learning without supervision labels.

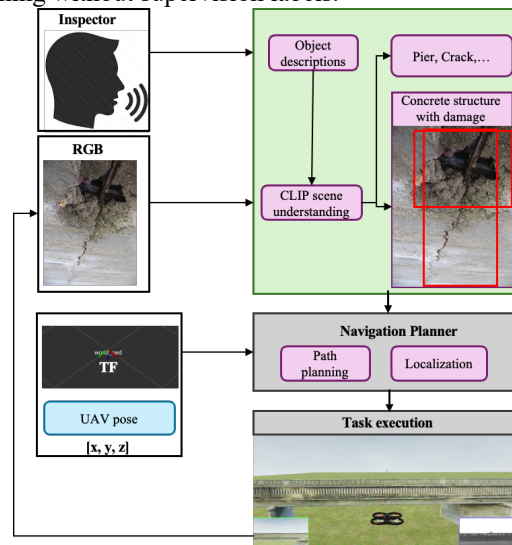


Figure 1 The framework of human-UAV collaboration in bridge inspections

The task is formulated as follows: A UAV is randomly placed within an unseen environment E , with a sequence of predefined navigation goals $G = \{g_1, g_2, \dots, g_n\}$, decoded from natural language inputs (e.g., “pier,” “girder,” or “crack”). The objective of the UAV is to navigate to any specified goal object. At time t , the UAV receives an observation in the form of an RGB image I_t from its onboard camera and must select an action from the action space A . Navigation is considered successful if the UAV stops within a safe distance of the object, and the object is visible without further movement. In this study, the authors focused on scene object recognition using CLIP. Given a set of navigation goals G and the collected image I_t at time t , the objective is to find if the image I_t contains the target object $g_m \in G$. If the target is present, a semantic scene understanding prompt $P_{g_m,t}$ is generated to describe the goal object. The task can be formulated as the following equation:

$$P_{g_m,t} = \text{CLIP}(I_t, G) \quad (1)$$

2.1 Few-shot CLIP for UAV-assisted Bridge Inspections

Few-shot image classification offers a solution for fine-tuning CLIP with fewer training datasets, reduced computational resources, and shorter training time. Authors have explored few-shot adaptation techniques, leading to two primary strategies: prompt-based and adapter-based approaches (Liu et al., 2024). Prompt-based fine-tuning methods, such as CoOp (Zhou et al., 2022b) and CoCoOp (Zhou et al., 2022a), transform the fixed textual prompts of CLIP’s text encoder into learnable vectors. These vectors are then fine-tuned using a small dataset to improve the performance of CLIP on domain-specific tasks. Adapter-based methods, such as CLIP-Adapter (Gao et al., 2021) and Tip-Adapter (Zhang et al., 2022), introduce lightweight adapter modules into the pre-trained model. These adapter parameters are fine-tuned with a small dataset, allowing the foundation model to address domain-specific tasks with minimal resource requirements effectively.

CoOp and Tip-Adapter are representative works in prompt-based and adapter-based fine-tuning methods, respectively. CoOp is designed to optimise the prompt’s context to enhance the image recognition performance of CLIP. The goal is to iteratively refine the prompt based on the performance of CLIP on domain-specific tasks, ultimately finding the optimal prompt for classification. In the conventional zero-shot method, CLIP uses prompt templates such as “A photo of a {label},” where class labels are inserted into predefined text prompts, and image-text similarity is calculated for classification. CoOp improves this process by introducing a set of learnable vectors $\{[V]_1, [V]_2, \dots, [V]_M\}$ that model contextual text alongside class labels within the prompt. Both text and image features are computed during forward propagation, and cross-entropy loss is calculated concerning the labels. The learnable vectors are updated during backpropagation to minimise the loss, while the weights of the pre-trained CLIP model remain fixed, with only the learnable tokens being fine-tuned. CoOp focuses solely on prompt optimisation of the text. CoOp does not incorporate domain knowledge transfer on the image side, which may limit its effectiveness in domain-specific tasks.

The Tip-Adapter method enhances few-shot classification by utilising a pre-trained CLIP model and constructing a key-value cache model. In a typical few-shot setting, there are N classes, each with K samples (K -shot), resulting in NK images in the training set. Visual features are extracted from these NK images using

the visual encoder of CLIP, which serves as the keys, while the corresponding one-hot encoded labels act as the values in the cache model. This cache model is integrated with the pre-trained CLIP classifier without requiring additional parameter tuning. During inference, the affinity between the test image features and the cache keys is calculated, and the corresponding values are aggregated to form the prediction of the adapter. Tip-Adapter effectively combines the zero-shot prediction capabilities of CLIP with domain-specific visual knowledge from few-shot tasks. Tip-Adapter still relies on manually designed prompts for image classification, limiting its ability to fully harness the extensive knowledge embedded in the text encoder of CLIP. This dependency on manual prompts reduces its potential to leverage the complete semantic understanding of the model across modalities.

This study incorporated text prompt learning by CoOp with the cache model of Tip-Adapter, as shown in Figure 2. Given the new bridge inspection dataset (K samples and N classes), image dataset I_K , text labels L_N , the process of text prompt learning with Tip-Adapter is as follows. A context vector L_{initial} was randomly initialised by drawing from a zero-mean Gaussian distribution. L_N was converted into N -dimensional one-hot vectors $\text{OneHot}(L_N)$. The learnable vector L_{learn} is trained through a tunable *TextEncoder*, with the input as $\text{OneHot}(L_N)$ and L_{initial} to get the optimised prompt L_{learn} in the training phrase:

$$L_{\text{learn}} = \text{TextEncoder}(L_N, L_{\text{initial}}) \quad (2)$$

According to CoOp, the prompt can be designed to put class labels at the end of L_{learn} : $[V]_1[V]_2 \dots [V]_M[\text{Class}]$ or in the middle of L_{learn} : $[V]_1 \dots V_{\frac{M}{2}}[\text{Class}][V]_{\frac{M}{2}+1} \dots [V]_M$.

Then the parameters of the text prompt were frozen, and the pre-trained CLIP *VisualEncoder* was used to extract the L2-normalized C -dimensional visual features of each image I_K in the training set:

$$F_{\text{train}}^T \in \mathbb{R}^{N \times C} = \text{VisualEncoder}(I_K) \quad (3)$$

During inference the visual features f_{test} from the test image I_{test} are extracted using *VisualEncoder*:

$$f_{\text{test}} \in \mathbb{R}^{1 \times C} = \text{VisualEncoder}(I_{\text{test}}) \quad (4)$$

Then the affinities A between F_{train}^T and f_{test} are computed as follows:

$$A = \exp(-\beta(1 - f_{\text{test}}F_{\text{train}}^T)) \quad (5)$$

Where β is used to control the sharpness of the similarity distribution, which ensures better classification performance, the ache model AL_{learn} is built as the dot product of the affinities A and the optimised prompt L_{learn} . The final logits are computed as the summation of cache model AL_{learn} and prior knowledge $f_{\text{test}}W_c^T$ from the pre-trained CLIP:

$$\text{logits} = \alpha \cdot AL_{\text{learn}} + f_{\text{test}}W_c^T \quad (6)$$

where α is a weighting factor, W_c^T is the weight *TextEncoder* generates.

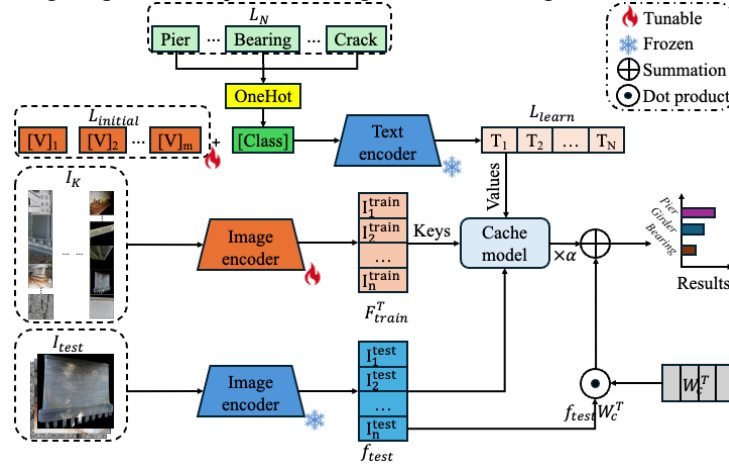


Figure 2. Few-shot CLIP incorporated text prompt learning with a cache model

3. EXPERIMENT AND RESULTS

To evaluate the performance of the proposed few-shot CLIP algorithm for object recognition in bridge inspections, experiments were conducted using both public datasets and a newly developed domain-specific dataset. The evaluation involved 1-shot, 2-shot, 4-shot, 8-shot, and 16-shot image classification tasks, where a limited number of labelled examples were available for each class. The accuracy of the proposed method was compared with that of the baseline CLIP and other fine-tuned CLIP-based models.

3.1 Dataset

To train and evaluate the performance of few-shot image classification algorithms for bridge inspection tasks, experiments were conducted on three public datasets, including Common Objects in Context Dataset for Structural Detail Detection of Bridges (COCO-Bridge) (Bianchi et al., 2021), Labelled Cracks in the Wild (LCW) (Bianchi & Hebdon, 2022), and Corrosion Condition State Semantic Segmentation Dataset (CCSSSD) (Bianchi & Hebdon, 2022), along with a newly developed dataset, the Bridge Member Dataset (BMD). COCO-Bridge

comprises 774 images and over 2,500 object instances collected by UAV, targeting the detection of four key structural bridge details: bearing, cover plate termination, gusset plate connection, and out-of-plane stiffener. The dataset provides a broad range of structural features essential for bridge inspection. LCW contains 3,817 finely annotated images of segmented cracks gathered from structural inspection reports provided by the Virginia Department of Transportation (VDOT). This dataset is focused on identifying and segmenting cracks, which are critical for assessing structural integrity.

Similarly CCSSSD includes 440 finely annotated images of segmented corrosion sourced from VDOT Bridge Inspection Reports. Corrosion is another primary concern in bridge inspections, and this dataset enables detailed analysis and classification of such defects. In addition to these public datasets, a new small-scale dataset, BMD, was developed specifically for this study. BMD comprises 150 images and includes four segmented bridge components: pier, girder, railing, and pavement. Including this dataset allows for further testing on distinct bridge elements often inspected during UAV-assisted bridge inspection tasks. Together, these datasets provide a comprehensive testing environment for assessing the effectiveness of few-shot image classification models in real-world bridge inspection scenarios.

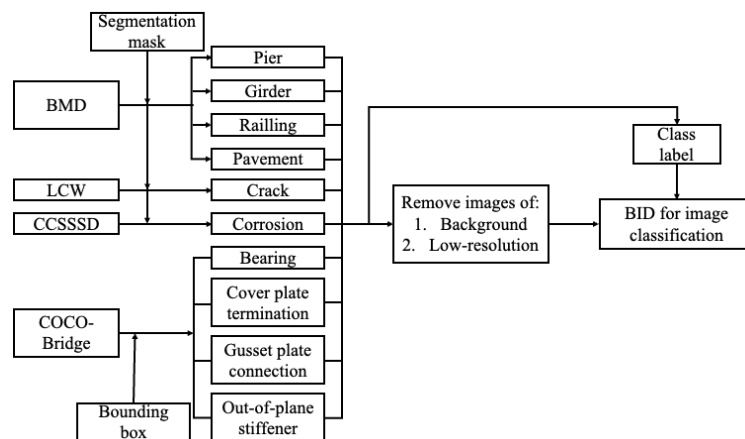


Figure 3. Preprocessing of datasets

The authors implemented a customisation process to adapt the datasets for classification, as shown in Figure 3. First the segmentation areas for each object in the BMD, LCW, and CCSSSD datasets were extracted using the annotated segmentation masks and class labels. In contrast the remaining areas were filled with black. Regions of Interest (ROIs) were cropped for the COCO-Bridge dataset based on the annotated bounding boxes and class labels. Next task-unrelated background areas in the cropped images were removed, and any segmented or cropped areas with low image quality were discarded to ensure the overall quality of the dataset. Finally the pre-processed datasets were split into 80% for training, 10% for testing, and 10% for evaluation. Sample images and their corresponding descriptions are shown in Figure 4.

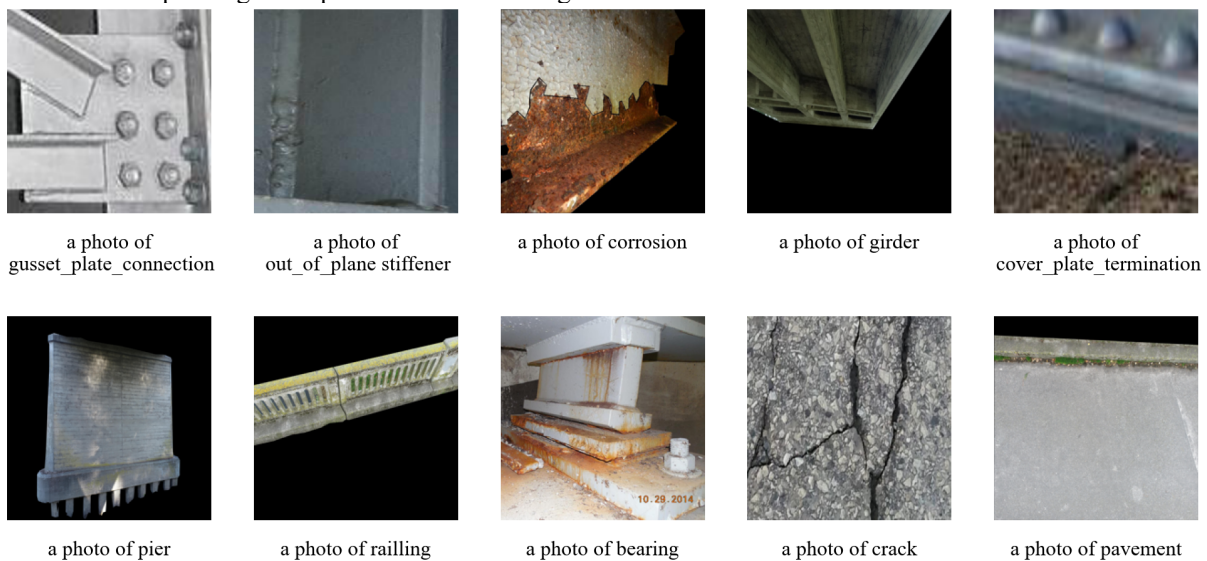


Figure 4. Sample images and corresponding descriptions

3.2 Evaluation of the proposed Few-shot CLIP model

The authors conducted experiments using the proposed few-shot CLIP model incorporating text prompt

learning with a cache model on the bridge inspection dataset. To evaluate few-shot learning, performances were compared across 1, 2, 4, 8, and 16-shot training sets, with testing conducted on the whole test set. ViT-B/16 (Dosovitskiy et al., 2021) was used as the visual encoder, and the training epoch for prompt tuning via CoOp was set to 50. Once the text prompt was trained, the optimised text weights were frozen and extracted as pre-trained text features, which were then used as input for the text encoder in Tip-Adapter. During Tip-Adapter training, the parameters were configured with 50 training epochs, a batch size 256, and a learning rate of 0.001.

Performance comparison was conducted between the model proposed by authors and Zero-shot CLIP (Pan et al., 2022), CoOp (Zhou et al., 2022b), COCoOp (Zhou et al., 2022a), MaPLe (Khattak et al., 2023), and Tip-Adapter (Zhang et al., 2022). All experiments were performed on 1, 2, 4, 8, and 16-shot training sets and evaluated on the complete test sets. For a fair comparison, the visual encoder backbone for all models was standardised to ViT-B/16. The training epoch was all set as 50. Following the framework of CoOp, four variants were tested: class token placed at the end or middle of the prompt, unified context (UC) versus class-specific context (CSC). The number of context tokens for CoOp was set to 16.

Figure 5 illustrates the performance of various models across different shot settings, ranging from 1 to 16 shots per class. At 16 shots, CoOp variants exhibit strong performance, with accuracy ranging from 79.43% to 80.33%, while Tip-Adapter, COCoOP, and MaPLe achieve 76.58%, 27.90%, and 30.50%, respectively. The proposed model surpasses all these methods, achieving the highest accuracy of 83.33%, highlighting its effectiveness in improving accuracy for bridge inspection tasks. The proposed model consistently outperforms the other methods in lower-shot settings (1, 2, 4, and 8 shots). For instance, in the 8-shot setting, it achieves 75.89%, outperforming CoOp (68.40% to 71.67%) and Tip-Adapter (71.64%). This trend continues in the 1, 2, and 4-shot settings, where the proposed model remains competitive or superior, underscoring its robustness and adaptability in few-shot bridge inspection image classification tasks.

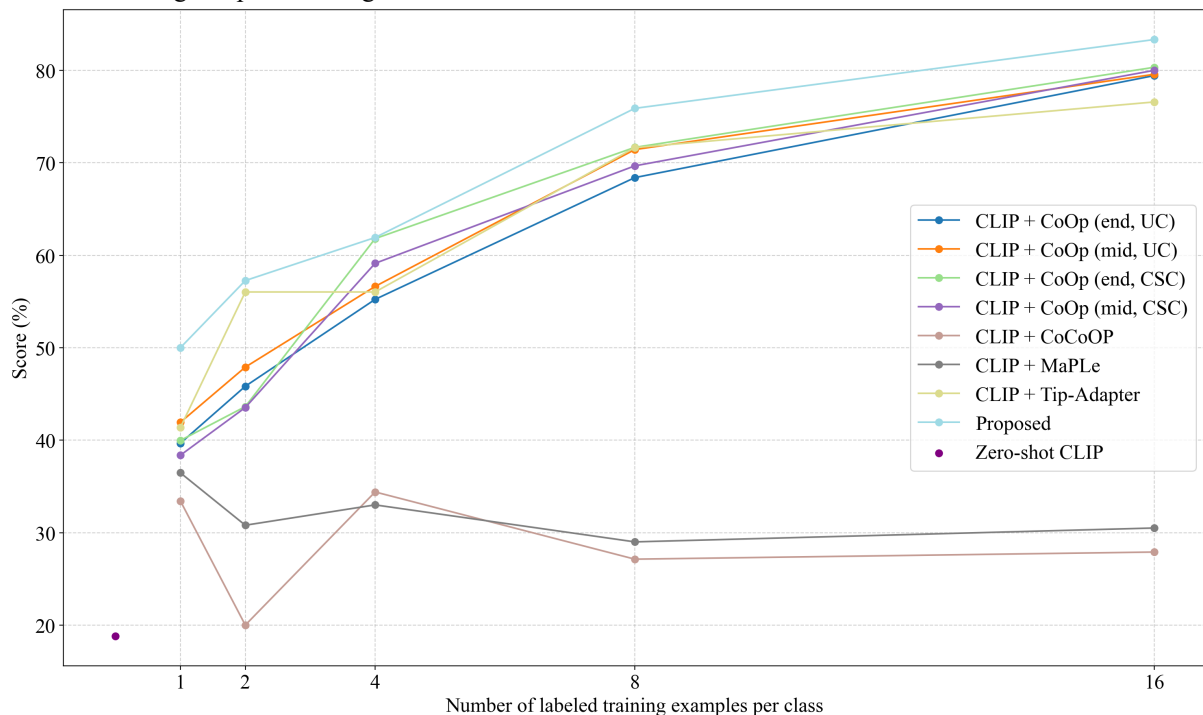


Figure 5. Few-shot classification accuracy of different models using 1, 2, 4, 8, 16 shots

4. DISCUSSION AND CONCLUSIONS

In this study the authors proposed a few-shot CLIP model incorporating text prompt learning with a cache model to enhance HRC in UAV-assisted bridge inspection and scene understanding tasks. The proposed model achieved the highest accuracy of 83.33%, outperforming other few-shot image classification methods, demonstrating its effectiveness in leveraging few-shot learning for bridge inspection tasks. By enabling UAVs to interpret the bridge environment based on human instructions, the model shows promise in improving the accuracy and efficiency of visual bridge inspections.

The contributions of this paper are twofold. First the proposed approach combines a text prompt learning method introduced by CoOP with a cache model developed by Tip-Adapter within a unified CLIP-based framework to capture both language and visual knowledge for bridge inspection tasks. Second the newly developed BMD dataset is introduced along with three modified datasets: COCO Bridge, LCW, and CCSSSD, to train and evaluate the proposed algorithm. This work demonstrates the potential of vision language models to enhance HRC in UAV-assisted bridge inspections.

Although the proposed model with pre-trained prompts outperforms manually designed prompts, the text prompts trained by CoOp are a string of vectors that are relatively difficult to interpret. Recent research has begun exploring external knowledge, such as knowledge graphs (e.g., CuPL by Pratt et al. (2022)), to help models better handle unseen samples, enhance semantic comprehension and robustness, improve interpretability, and specialise in specific domains. Incorporating domain knowledge from bridge inspection into both prompt engineering and visual adapter design presents a promising direction for improving performance in future bridge inspection scene understanding tasks. Future work will also focus on refining and integrating the model into the Robotic Operating System (ROS) for deployment on real UAVs to perform bridge inspection tasks. This study underscores the potential for AI-empowered UAVs to revolutionise bridge inspection processes, making UAVs more efficient and reliable in inspection tasks.

ACKNOWLEDGMENTS

This research is supported financially by the University of Auckland and the China Scholarship Council (CSC) (Grant number: 202207000017).

REFERENCES

- Ajoudani, A., Zanchettin, A. M., Ivaldi, S., Albu-Schäffer, A., Kosuge, K., & Khatib, O. (2018). Progress and prospects of the human–robot collaboration. *Autonomous Robots*, 42(5), 957–975. <https://doi.org/10.1007/s10514-017-9677-2>
- Bianchi, E., Abbott, A. L., Tokekar, P., & Hebdon, M. (2021). COCO-Bridge: Structural Detail Data Set for Bridge Inspections. *Journal of Computing in Civil Engineering*, 35(3), 04021003. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000949](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000949)
- Bianchi, E., & Hebdon, M. (2022). Development of Extendable Open-Source Structural Inspection Datasets. *Journal of Computing in Civil Engineering*, 36(6), 04022039. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0001045](https://doi.org/10.1061/(ASCE)CP.1943-5487.0001045)
- Bolourian, N., & Hammad, A. (2020). LiDAR-equipped UAV path planning considering potential locations of defects for bridge inspection. *Automation in Construction*, 117, 103250. <https://doi.org/10.1016/j.autcon.2020.103250>
- Brohan, A., Brown, N., Carbajal, J., Chebotar, Y., Chen, X., Choromanski, K., Ding, T., Driess, D., Dubey, A., Finn, C., Florence, P., Fu, C., Arenas, M. G., Gopalakrishnan, K., Han, K., Hausman, K., Herzog, A., Hsu, J., Ichter, B., ... Zitkovich, B. (2023). *RT-2: Vision-Language-Action Models Transfer Web Knowledge to Robotic Control* (arXiv:2307.15818). arXiv. <https://doi.org/10.48550/arXiv.2307.15818>
- Contreras, R., Ayala, A., & Cruz, F. (2020). Unmanned Aerial Vehicle Control through Domain-Based Automatic Speech Recognition. *Computers*, 9(3), Article 3. <https://doi.org/10.3390/computers9030075>
- Dorafshan, S., & Maguire, M. (2018). Bridge inspection: Human performance, unmanned aerial systems and automation. *Journal of Civil Structural Health Monitoring*, 8(3), 443–476. <https://doi.org/10.1007/s13349-018-0285-4>
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale* (arXiv:2010.11929). arXiv. <https://doi.org/10.48550/arXiv.2010.11929>
- Gadre, S. Y., Wortsman, M., Ilharco, G., Schmidt, L., & Song, S. (2022). *CoWs on Pasture: Baselines and Benchmarks for Language-Driven Zero-Shot Object Navigation* (arXiv:2203.10421). arXiv. <https://doi.org/10.48550/arXiv.2203.10421>
- Gao, P., Geng, S., Zhang, R., Ma, T., Fang, R., Zhang, Y., Li, H., & Qiao, Y. (2021). *CLIP-Adapter: Better Vision-Language Models with Feature Adapters* (arXiv:2110.04544). arXiv. <https://doi.org/10.48550/arXiv.2110.04544>
- Khattak, M. U., Rasheed, H., Maaz, M., Khan, S., & Khan, F. S. (2023). *MaPLe: Multi-modal Prompt Learning* (arXiv:2210.03117). arXiv. <https://doi.org/10.48550/arXiv.2210.03117>
- Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollár, P., & Girshick, R. (2023). *Segment Anything* (arXiv:2304.02643). arXiv. <https://doi.org/10.48550/arXiv.2304.02643>
- Liu, F., Zhang, T., Dai, W., Zhang, C., Cai, W., Zhou, X., & Chen, D. (2024). Few-shot adaptation of multi-modal foundation models: A survey. *Artificial Intelligence Review*, 57(10), 268. <https://doi.org/10.1007/s10462-024-10915-y>
- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. (2016). SSD: Single Shot MultiBox Detector. In B. Leibe, J. Matas, N. Sebe, & M. Welling (Eds.), *Computer Vision – ECCV 2016* (pp. 21–37). Springer International Publishing. https://doi.org/10.1007/978-3-319-46448-0_2
- Lu, J., Batra, D., Parikh, D., & Lee, S. (2019). ViLBERT: Pretraining task-agnostic visiolinguistic representations

- for vision-and-language tasks. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems* (pp. 13–23). Curran Associates Inc.
- Michalos, G., Makris, S., Spiliotopoulos, J., Misios, I., Tsarouchi, P., & Chrysosolouris, G. (2014). ROBO-PARTNER: Seamless Human-Robot Cooperation for Intelligent, Flexible and Safe Operations in the Assembly Factories of the Future. *Procedia CIRP*, 23, 71–76. <https://doi.org/10.1016/j.procir.2014.10.079>
- Naseer, F., Ullah, G., Siddiqui, M. A., Jawad Khan, M., Hong, K.-S., & Naseer, N. (2022). Deep Learning-Based Unmanned Aerial Vehicle Control with Hand Gesture and Computer Vision. *2022 13th Asian Control Conference (ASCC)*, 1–6. <https://doi.org/10.23919/ASCC56756.2022.9828347>
- OpenAI. (2020). *ChatGPT* [Computer software]. <https://chat.openai.com>
- Pan, X., Ye, T., Han, D., Song, S., & Huang, G. (2022). *Contrastive Language-Image Pre-Training with Knowledge Graphs* (arXiv:2210.08901). arXiv. <https://doi.org/10.48550/arXiv.2210.08901>
- Park, S., Wang, X., Menassa, C. C., Kamat, V. R., & Chai, J. Y. (2024). Natural language instructions for intuitive human interaction with robotic assistants in field construction work. *Automation in Construction*, 161, 105345. <https://doi.org/10.1016/j.autcon.2024.105345>
- Pratt, S., Covert, I., Liu, R., & Farhadi, A. (2022, September 7). *What does a platypus look like? Generating customized prompts for zero-shot image classification*. arXiv.Org. <https://arxiv.org/abs/2209.03320v3>
- Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M., & Sutskever, I. (2021). *Zero-Shot Text-to-Image Generation* (arXiv:2102.12092). arXiv. <https://doi.org/10.48550/arXiv.2102.12092>
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 779–788. <https://doi.org/10.1109/CVPR.2016.91>
- Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1*, 91–99.
- Seo, J., Duque, L., & Wacker, J. (2018). Drone-enabled bridge inspection methodology and application. *Automation in Construction*, 94, 112–126. <https://doi.org/10.1016/j.autcon.2018.06.006>
- Shibata, K., Deguchi, H., & Taguchi, S. (2024). *CLIP feature-based randomized control using images and text for multiple tasks and robots* (arXiv:2401.10085). arXiv. <https://doi.org/10.48550/arXiv.2401.10085>
- Zhang, C., Zou, Y., Wang, F., del Rey Castillo, E., Dimyadi, J., & Chen, L. (2022). Towards fully automated unmanned aerial vehicle-enabled bridge inspection: Where are we at? *Construction and Building Materials*, 347, 128543. <https://doi.org/10.1016/j.conbuildmat.2022.128543>
- Zhang, R., Zhang, W., Fang, R., Gao, P., Li, K., Dai, J., Qiao, Y., & Li, H. (2022). Tip-Adapter: Training-Free Adaption of CLIP for Few-Shot Classification. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, & T. Hassner (Eds.), *Computer Vision – ECCV 2022* (pp. 493–510). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-19833-5_29
- Zhou, K., Yang, J., Loy, C. C., & Liu, Z. (2022a). *Conditional Prompt Learning for Vision-Language Models* (arXiv:2203.05557). arXiv. <https://doi.org/10.48550/arXiv.2203.05557>
- Zhou, K., Yang, J., Loy, C. C., & Liu, Z. (2022b). Learning to Prompt for Vision-Language Models. *International Journal of Computer Vision*, 130(9), 2337–2348. <https://doi.org/10.1007/s11263-022-01653-1>

AN INTERACTIVE LLM-BASED FRAMEWORK FOR SPATIAL RELATIONSHIP QUERY ON OPENBIM ELEMENTS WITH IFC4 SCHEMA

Ang Li¹, Peter Kok-Yiu Wong², Xingyu Tao³, Jack C.P. Cheng⁴

1) Ph.D. Student, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: lionel_a.li@connect.ust.hk

2) Postdoctoral Fellow, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: cekywong@ust.hk

3) Research Assistant Professor, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. E-mail: xtaoab@connect.ust.hk

4) Professor, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. E-mail: cejcheng@ust.hk

Abstract: Spatial relationships between BIM elements are crucial for various BIM-based analysis (e.g., identifying external building envelope components for energy simulation), however current BIM information retrieval researches are mainly in the field of extracting attributes from BIM element, a solution for obtaining implicit spatial information from BIM model is needed. Thus, this research focus on how to accurately and effectively obtaining spatial relationship between BIM elements and reliable answering spatial-related BIM query. Addressing this issue, we proposed a IFC-based spatial relation calculation and query-answer framework, which is summarized as follows: (1) extract geometric information of the IFC entities with openBIM standards and obtain the triangulated boundary data of each entities; (2) generate AABB tree of the entities using triangulated boundary data to index the entities and improve the search efficiency in spatial calculation; (3) bSDD-aided LLM workflow to align natural-language queries with corresponding IFC entities to answer spatial relationship queries. We use several building cases to verify our proposed method, the results indicate that our method can accurately understand the natural language query (92.1% correct rate on query understanding tasks) and efficiently determine elements' spatial relationship (saving 61.4% average query time) to answer the original query. With our query method, users with minimal BIM experience (e.g., construction site workers) can still easily query the spatial relationship in a user-friendly way, improving the applicability of the BIM technique.

Keywords: Large Language Model (LLM), Spatial Relationship Query, Prompt Pattern, AABB Tree

1. INTRODUCTION

Identifying the spatial relationship of building elements or building element sets is essential for a wide range of BIM applications across the life-cycle (Tao et al., 2022), e.g., extracting external building envelope components for energy simulation, recognizing the touching elements for laser scanning planning (Chen et al., 2024). The spatial relationship can be referred to topological relationship of spatial elements, which can basically be classified into six type i.e., disjoint, equal, touch, contain, overlap and cover (Borrmann and Rank, 2009). However, the spatial relationship is difficult to directly be obtained from BIM due to the heterogeneous and unstructured BIM data (Wang et al., 2024). The industry foundation classes (IFC) file format, which is the uniform standard for BIM data storage and exchange, didn't define spatial relationship as a general property or attributes of BIM entities. IFC schema provides rich approach to represent geometric and spatial data, for example, swept area representation (IfcExtrudedAreaSolid), boundary representation (IfcFacetedBrep), constructive solid geometry (IfcCsgSolid), each representation approach uses different IFC data structure to record the geometric data (Solihin et al., 2017), but it is recorded with complexity multi-layer attributes inheritance and various attribute formats. Thus, it is valuable to design a framework to automatically extract relevant spatial-geometric information of BIM elements from complex IFC data and calculate the spatial relationship between them.

Usually, a BIM model contains a large number of building elements (e.g., wall, beam column) and various of geometric attribute (e.g., shape, boundary), but they are not indexed based on their spatial position. So, searching an element that satisfy the given relationship (e.g., find out all the touching element of a wall) requires to traverse through all the elements and calculating their spatial relationship (Ying et al., 2022), which is low efficiency since the time cost exponentially increase as the number of elements increase ($t_0 \cdot 2^n$, t_0 is the calculation time of intersection test for two elements). Although several spatial indexes (e.g., Octree, Quadtree) to decompose the whole BIM space and improve the efficiency (Borrmann and Rank, 2009), but an BIM object might be divided into two sub-space which may result to computing error. Thus, a powerful spatial index could accelerate the calculation and reduce computational cost is important for the spatial relationship query.

In addition, most of BIM query languages are structured query language-based, users need to learn query grammar and query function to access to the required data or information (Wang et al., 2022). However, not only the BIM professionals will use the BIM query, people who are unfamiliar with BIM also need to query information from BIM, like construction worker need to query about the equipment contained in the room for better installation. Traditional BIM query method might not be user-friendly for all the user. The emerging large language model (LLM) like chatGPT shows excellent performance on context understanding and extracting specific information

(Zheng and Fischer, 2023), which imply the potential on semantic processing of spatial query in natural language and interacting with spatial information. Therefore, a methodology is possible to engage LLM to facilitate spatial query to enhance its applicability, so that user even without any background on IFC or SQL can also finish the query.

Against this contextual backdrop, this paper intends to propose an interactive query method for spatial relationship. With such query method, users can use natural language to query the desired spatial information and effectively obtain the correct answer of the original query, which can promote the application of BIM for unexperienced users.

2. RELATED WORKS

2.1 BIM spatial relationship

The spatial relationship of BIM elements is a type of implicitly semantic information in BIM models, inferred from the geometric data. In engineering practice, the BIM data including geometric data (e.g., object shape) and non-geometric data (e.g., materials) (Wang et al., 2022), are usually stored in stakeholder-neutral IFC files for the sake of effective information exchange. To mine the spatial relationship in BIM, current approaches usually comprise two steps: (1) Extract geometric data from BIM/IFC file; (2) Determine the spatial relationship using computational geometry (CG) algorithms.

Among all geometric representation, the boundary representation (B-rep) is the most popular one for analysing topological and spatial relationship using CG algorithms (Solihin et al., 2017). The mathematical model for such geometric representations is robust since it can be validated rather easily and consistently, due to obeying the Euler operators. With the normalized B-rep (e.g., triangulated B-rep, using minimal triangles to represent the shape boundary), other B-rep like Axis-aligned bounding box (AABB), Oriented bounding box (OBB) can also be easily obtained (Ying and Lee, 2020). Thus, we use triangulated B-rep to represent the extracted geometric information for facilitating further calculating. Additionally, buildingSMART proposed Model View Design (MVD) to standardize the information exchange requirements (e.g., for energy simulation or quantity take-off (Jeon et al., 2021)), but few MVDs is designed for spatial relationship determination, so we develop an MVD to standardize the necessary geometric information unit and support information extraction algorithm.

The second step for spatial relationships is determining spatial relationships with CG algorithms. Borrmann and Rank (2009) examined the use of the 9-IM as a formal basis for defining topological operators in an BIM environment, which depicts 8 typical topological relationships (i.e., Disjoint, Inside, Equal, Touching, Containing, Overlapping, Covering and CoveredBy) using the 9-IM matrix. The 9-IM approach can cover most of practical spatial relationship, thus our proposed method use it as theoretical basis to determine the spatial relationship.

2.2 BIM query language

Previous BIM query languages (e.g., BIMQL, PMQL) have been developed for filtering objects according to their attribute and type values, manipulating the result set like union, complementary, and get the value of the attributes (Daum et al., 2014). However, none of the BIM query take the geometric representation of IFC objects for the filtering of data into account, so it can't directly execute spatial query or provide information for spatial calculation. Thus, our research integrates the geometric information retrieval and spatial relationship determination into spatial query method to enable the reliable answer can be obtained.

To make the query system more acceptable for nonexpert user, Natural language processing (NLP) is used to enable query system to understand query in natural human language and build a natural language-based virtual assistant to structuralize the original query and support BIM activities (Wang et al., 2022). However, traditional NLP methods (e.g., Word2Vec, BERT) have restricted performance on semantic similarity comparison (Devlin et al., 2019), for example, the "measured net distance" and "NetPerimeter" are different concept (distance and perimeter are disjoint) but are likely to be assigned a high similarity score. The LLM like chatGPT has excellent semantic and context understanding performance (Zheng and Fischer, 2023), but it hasn't been applied to BIM query. Thus, we develop an LLM workflow to understand the NL spatial query and convert it into structured command, which is subsequently utilized to retrieve relevant information and answer the original query.

3. RESEARCH METHOD

Addressing key challenges (complex data extraction, low search efficiency, user-unfriendly interaction) in spatial query of BIM, we use openBIM MVD standard to support IFC data extraction, generate spatial index of BIM elements to improve efficiency, develop LLM-based workflow to understand natural language query of user and answer the query with determined spatial relationship.

As Figure 1 shows, the proposed methodology includes three parts as follows: 1) extract IFC information. In this part, the geometric information contained in IFC files is extracted and the triangulated B-rep are utilized to represent corresponding geometry. 2) generate AABB tree. In this part, the AABB of IFC entities are calculated and the AABB tree structure is engaged to index the IFC entities and improve the efficiency in calculation. 3) understand and answer query. In this part, the keyword in natural language query is identified and align to its

corresponding IFC entities to structuralized the query, then relevant information is retrieved to determine relationship, subsequently answer the query in natural language.

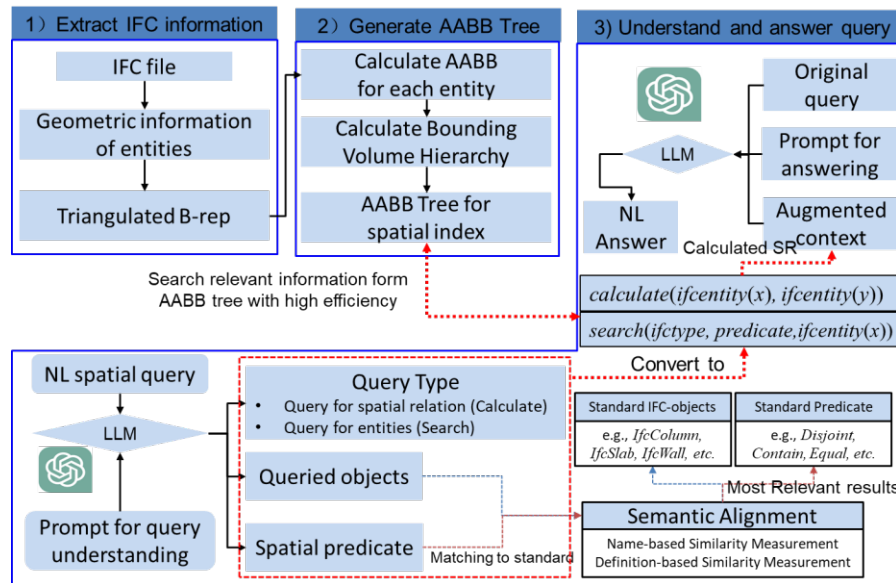


Figure 1. Overall Methodology of the spatial query system

Part 1: Extract IFC information. Extract geometric information of the IFC entities based on openBIM IDM-MVD method. The objective of this part is to convert implicit geometric data (in IFC) into explicit geometric representation. The Model View Definition (MVD) is utilized to standardize the necessary information for determining spatial relationship, where 4 spatial containments (Site, Building, Building Storey, Space) and 6 building elements (Beam, Column, Door, Slab, Wall, Window) as well as their corresponding IFC geometric data structures are defined in the MVD concepts for IFC data reduction and facilitate the spatial calculation. The defined MVD is presented in Figure 2. Besides, the Information Delivery Manual (IDM) can be utilized to extend the MVD to specify the information exchange requirement between different BIM users across the life-cycle. Under the support of these openBIM standards, we use IfcOpenShell to extract relevant geometric information of IFC objects and uniform their coordinate systems, then the obtained spatial data are imported to FreeCAD to generate explicit geometric representation and calculate triangulated B-rep, which can normalize further calculation process for determining spatial relationship (i.e., triangle intersection algorithms) and guarantee computational accuracy.

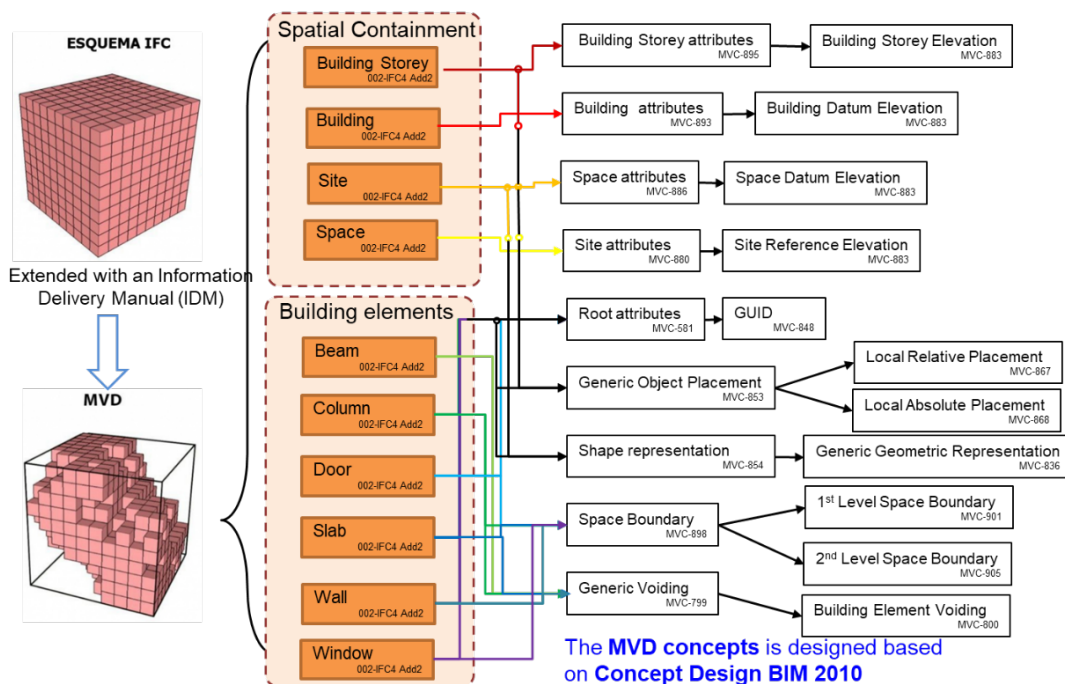


Figure 2. MVD for the spatial relationship

Part 2: Generate AABB tree. Spatial index of the entities is generated using triangulated boundary data to index the entities and improve the search efficiency in spatial calculation. Although the triangulated B-rep can be directly utilized to determine the spatial relationship of entities, the numerous objects in large-scale building models may cause the processing time become extremely high. In this step, the AABB tree indexing algorithm is established to index objects in the entire set of extracted IFC entities and accelerate the search efficiency by removing irrelevant candidates, as presented in Figure 3. An AABB tree is a space partitioning tree that indexes geometric primitives (e.g., points, triangles, and polyhedrons) based on their AABBs with $O(n \log(n))$ complexity (better than $O(n^2)$ complexity of without any index), which can be utilized to organize all the extracted building objects into tree structure using a top-down recursive method.

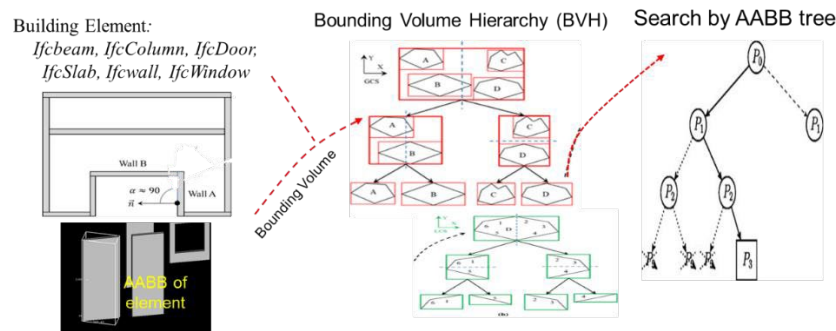


Figure 3. Spatial index by AABB tree

Part 3: Understand and answer query. The objective of this step is understanding the natural language spatial query and answering the query with relevant information to make the BIM query more user-friendly. In this step, we developed a prompt pattern for LLM (based on GPT-4) to understand query and translate it into a spatial query function command that can execute by the computer. The prompt pattern includes five components (i.e., system, input query, rules, instruction, output format) as figure 4 shows, which can help the LLM classify query type (i.e., query the relationship of two entities or query the entities with given relationships), identify the query keywords (i.e., relevant building element and spatial predicate), align the identified keywords to IFC entities (e.g., IfcBeam, IfcDoor) and standard spatial predicate (e.g., disjoint, overlap, touch), so that the LLM GPT can understand the original query. The buildingSMART bSDD is utilized to design the rules component, where the name and definition of the IFC entities and spatial predicate provide to help LLM GPT to align the keyword and avoid mismatching. The original spatial query can be converted into command format by LLM GPT, i.e., *calculate(ifcentity(x), ifcentity(y))* for query the relationship of two entities and *search(ifcentity, predicate, ifcentity(x))* for query the entities with given relationships.

Then, the query command is executed to retrieve the relevant geometric information from efficient spatial index, then the spatial relationship of objects is determined using spatial calculation (i.e., triangle intersection test, 9-Intersection Model) according to their geometric information. With the obtained spatial relationship as augmented context in answer prompt, the LLM generate reliable and easy understanding answer for different parties.

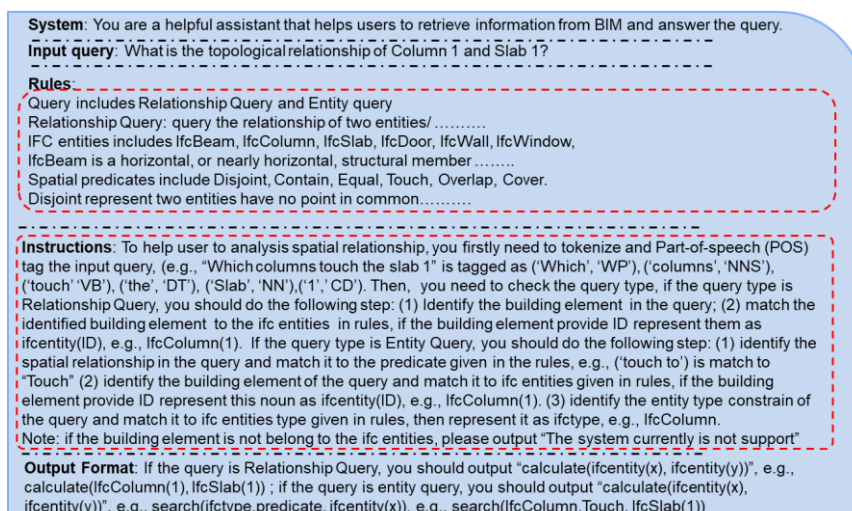


Figure 4. Prompt pattern incorporating bSDD

4. IMPLEMENTATION AND RESULTS

We use three practical BIM models (IFC files) in Hong Kong to verify our proposed method as shown in Figure 5, i.e., the Cheng Yu Tung (CYT) Building, Shaw Auditorium, Domain Mall. Each selected project's BIM model contains a large number of BIM elements (i.e., Cheng Yu Tung Building (contain 7004 IFC entitles), Shaw Auditorium (contain 33959 IFC entitles), Domain Mall (contain 22177 IFC entitles)). The results of the cases validation indicate that our proposed method can accurately extract geometric information from complex IFC data, generate two-level spatial index for efficient search (average 61.4% search time saving), precisely understand (92.1% correct rate on relationship query understanding and 90.3% correct rate on entity query understanding) and answer the natural language query by LLM to support user-friendly query system of BIM spatial relationship.

4.1 IFC-based geometric information extraction

To demonstrate whether our algorithms can correctly extract information from IFC file and index the elements, we apply our proposed method to the selected cases as Figure 5 shown. The extracting algorithm is developed in Python using IfcOpenShell and FreeCAD toolkit to parse the complex IFC schema and process the B-rep of the entities. The case study indicate that the proposed algorithm can automatically and accurately extract the geometric information of basic IFC entities (e.g., IfcBeam, IfcDoor, IfcColumn) and represent the entities in explicit geometry (i.e., triangulated B-rep) using global coordinate system, which is transformed from the local coordinate system based on IFC schema. In addition, the spatial index of the numerous elements can also be successfully generated based on the AABB tree, which utilize the well-structured tree-based approach to index the unstructured and numerous elements in complex BIM models.

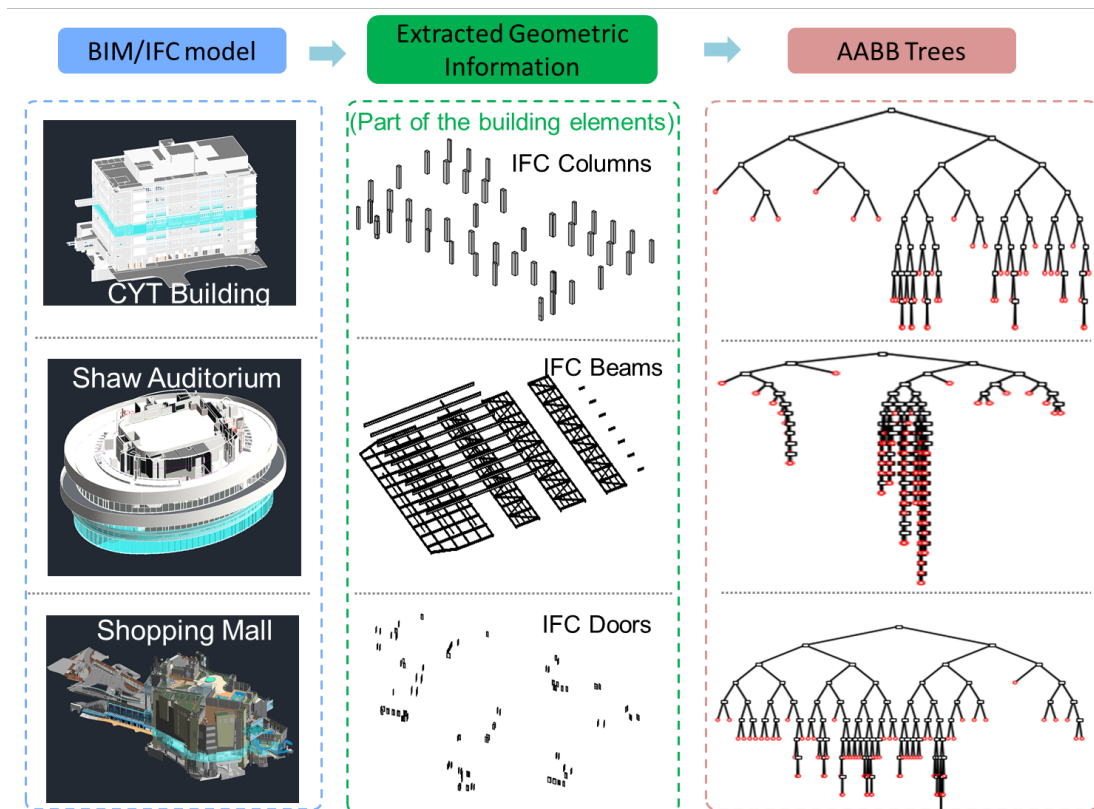


Figure 5. Generating spatial indexes based on geometric information

4.2 Efficiency improvement of spatial index

To verify the efficiency improvement of the spatial index, different index strategies (i.e., the proposed AABB tree and brute force implementation without any spatial indexing) and their average query time on the spatial search tasks were compared, as presented in Table 1. To handle with the high computational cost in traversing through all the relationship of elements, the AABB tree of all the entities is calculated by a top-down recursive method, which use hierarchical AABB to cascadingly partition the space. The results of case study indicate that the proposed AABB tree can significantly improve the search efficiency in different spatial relationship search task (e.g., search touch elements, disjointed elements), the query time is reduced by averagely 61.4% than not using any spatial index. For the BIM model with irregular shape (i.e., Shaw and CYT) and extremely complex layouts, our proposed spatial index still shows excellent performance, which imply the applicability of the AABB tree.

Table 1. Overall query time to obtain spatial relationship by different spatial index

Query Time (s)	Touch			Contain			Cover			Overlap		
	CYT	Shaw	Mall	CYT	Shaw	Mall	CYT	Shaw	Mall	CYT	Shaw	Mall
No index	106.7	346.8	572.2	82.4	302.5	472.5	80.6	358.3	549.0	79.0	371.1	554.2
AABB tree	49.8	131.4	259.6	28.1	98.5	206.8	35.7	116.0	173.2	33.9	119.5	221.2
Improvement	53.3%	62.1%	54.6%	65.9%	67.5%	56.2%	55.7%	67.6%	68.4%	57.1%	67.8%	60.1%

4.3 LLM-based spatial query understanding and answering

To validate the proposed LLM workflow, we use the actual natural language query to check the output of LLM. Figure 6 (a) shows the original spatial relationship query and Figure 6 (b) shows its output answer of LLM (including the converted command of the original query). The outputs shows that the LLM can understand the given query and convert it to structured command format for spatial calculation. In addition, the LLM can engage the determined spatial relationship form spatial calculation as augmented context to generate easy-understanding natural language answer of the original spatial query, which makes the query process of the spatial relationship more user-friendly.

To evaluate the understanding performance of the LLM on spatial query, we compared the understanding correctness (on query data set) of different LLM prompt patterns. The query data set contains 180 element queries and 180 relationship queries by random combinations of patterns (i.e., relationship query and element search query) and elements (e.g., beam, wall, column) to generate representative spatial queries which can cover different query styles. The difference prompt is designed based on the ablation study, which is presented in Table 2.

Table 2. Summary of different prompt strategies

Prompt	Description
bSDD-aided Prompt (Our)	System+ Input Query+ Rules (provide the name and definition of IFC Element for semantic mapping) + Instruction+ Output Format
Ablation Prompt-1	System+ Input Query+ Rules (only provide IFC Element name for name-based similarity measurement) + Instruction+ Output Format
Ablation Prompt-2	System+ Input Query + Instruction+ Output Format
Common Prompt	The prompt may include content of original query, the instruction and the format, but it is not clearly divided into different components, the task are directly given to LLM. No example shows to LLM.

The results of ablation study indicate that our proposed prompt has better semantic understanding capability than common prompt (i.e., only give instruction component) in different type of spatial query task since the correctness rate of LLM output for our prompt is 92% while that for common prompt is about 70%. The understanding performance significantly decrease after removing the definition of IFC entities, implying that the providing bSDD concepts can the enhance semantic understanding capability of LLM.

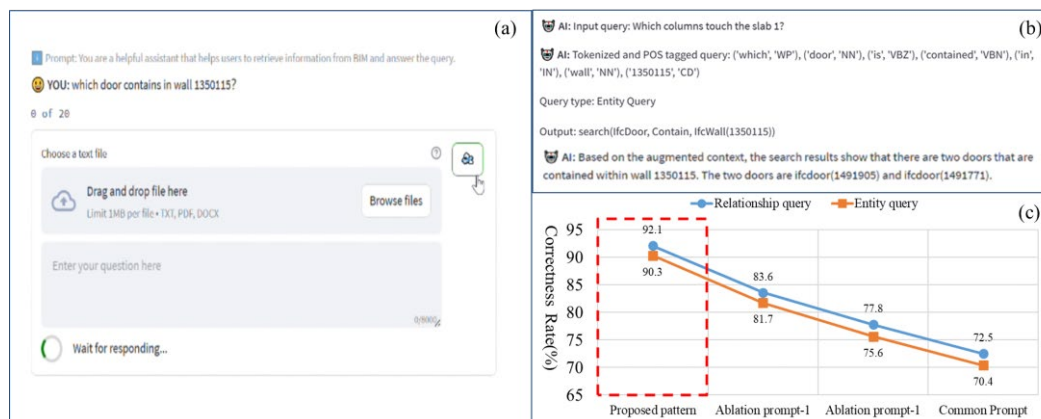


Figure 6. Understand & answer query: (a) Input query, (b) LLM outputs, (c) Performance of different prompt

5. DISCUSSIONS

5.1 Contributions

This research developed an algorithm for extracting geometric information of building elements (IFC entities) based on IFC schema, a spatial index for generating AABB tree of IFC entities to improve spatial calculation efficiency, an LLM prompt-based workflow to understand spatial query in natural language and answer the query with obtained spatial relationship as augmented context. The contributions of this research can be summarized as follows:

(1) Developed an algorithm for extracting geometric information of building elements based on MVD standard. The buildingSMART MVD can standardize the necessary geometric information for determining spatial relationship, where 4 fundamental spatial containments and 6 basic building elements as well as their IFC data structure are included in the MVD concepts. Under the requirement of MVD, we can develop a Python script to achieve automatic and standardized IFC information extraction, where the IfcOpenShell is utilized to parse the IFC schema and extract geometric information in python script and the FreeCAD is incorporated to such python script to process the extracted geometric information.

(2) Developed a spatial indexing for BIM elements based on their AABB to improve the search efficiency in spatial query. The AABB tree can convert the massive unstructured building elements into a well-structured tree structure based on the elements' geometric information (i.e., the triangulated B-rep). The proposed indexing system enables search function to rapidly find candidates and remove irrelevant elements by pruning subtree/node, which can significantly improve search efficiency in spatial query (save 61.4% average query time) and avoid exponential runtime complexity.

(3) Proposed an LLM workflow based on bSDD-aided prompt pattern to understand natural language spatial query and answer the query with relevant information. The proposed prompt pattern for understanding including System, Input Query, Rules, Instruction and Output Format components, where the bSDD concept can be added into rule component to provide name and definition of IFC entities for LLM. Such prompt pattern has shown 92.1% correct rate on relationship query understanding and 90.3% correct rate on entity query understanding, which has 20% higher correctness rate than common prompt. With the understanding prompt pattern, the LLM can identify the keyword in natural language query and align them to standard IFC entity, so that the LLM can convert the original query into command format to retrieve relevant geometric information of the queried object from spatial dataset. The relevant geometric information is subsequently retrieved and utilized to determine spatial relationship and added to answer prompt as augmented context to generate reliable and easy-understanding answer, which make the whole query process more user-friendly than traditional SQL-based query.

5.2 Limitations

While this research contributes to an efficient and user-friendly BIM spatial query system that can use NL to query the spatial relationship, it has several limitations that we need to acknowledge and address.

(a) This paper concentrates on the building elements in structural system (e.g., wall, beam), but the elements in different building system (e.g., MEP system) might not be successfully queried since they have different geometric data structures. Querying spatial relationship between different type of elements is meaningful in engineering practices (e.g., relationship of wall and MEP pipe), so the future research is suggested to consider different types of elements in the building.

(b) Another limitation is that we currently consider several general spatial queries pattern and the workflow of determining spatial relationship cannot be completely demonstrated during the interaction with the user. So, the future research could address the complex and combinatorial spatial query to facilitate users work, and provide a more interactive query system to help the user understand the spatial relationship.

6. CONCLUSIONS

Addressing the key challenges in spatial query of BIM model (i.e., complex BIM data structure, low-efficient spatial calculation and high threshold query language), This research use MVD to standardize the necessary information for determining spatial relationship of elements, which helps to develop an automatically algorithm for extracting IFC geometric information. To decompose the BIM model and index the elements for higher efficiency, this research engage AABB tree for generating spatial index of building elements based on their geometric information extracted from IFC data, which can convert the unstructured elements into a well-structured tree based structure. To enhance the user-friendliness of the query system, this research use bSDD-aided prompt engineering to develop LLM workflow, so that the LLM can understand the BIM-related concepts and answer the query with relevant information.

Three practical projects in Hong Kong are selected to verify our proposed method, the results of case studies indicate that our proposed framework can automatically obtain explicit geometric representation for spatial calculation, significantly improve the efficiency in search relevant spatial data, accurately understand the query in natural language to support user-friendly interaction. Specifically, for the search efficiency improvement, our developed AABB tree spatial index can save about 61.4% average query time for different element search query

task. For the understanding performance, the LLM using the bSDD-aided prompt pattern can achieve 92.1% correct rate on relationship query understanding and 90.3% correct rate on entity query understanding.

With our proposed method, non-expert (e.g., construction workers and end user) can still user-friendly query and get relevant information from complex BIM model. Due to the efficient spatial index, users can quickly query and retrieve BIM data, so different parties can spend less time to get the relevant information and respond to their collaborator with lower waiting time, the collaboration can be smoother, thus increasing work efficiency. To achieve higher applicability on different building system and various users, further research is required to consider various elements and equipment (e.g., plumbing) in the building and to support complex and combinational query pattern

ACKNOWLEDGMENTS

The authors would like to thank the Innovation and Technology Fund, HKSAR (No. PRP/009/23LI) for partially supporting this research. The authors would like to thank Foshan HKUST Project (FSUST21-HKUST05C) for providing partial support to this research. The authors declare no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

- Borrmann, A., & Rank, E. (2009). Topological analysis of 3D building models using a spatial query language. *Advanced Engineering Informatics*, 23(4), 370-385.
- Chen, Z., Wang, H., Chen, K., Song, C., Zhang, X., Wang, B., & Cheng, J. C. (2024). Improved coverage path planning for indoor robots based on BIM and robotic configurations. *Automation in Construction*, 158, 105160.
- Daum, S., & Borrmann, A. (2014). Processing of topological BIM queries using boundary representation based methods. *Advanced Engineering Informatics*, 28(4), 272-286.
- Devlin J., Chang, M., Lee, K. & Toutanova, K. (2019, June). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of naacL-HLT*, Vol. 1, p. 2.
- Solihin, W., Eastman, C., & Lee, Y. C. (2017). Multiple representation approach to achieve high-performance spatial queries of 3D BIM data using a relational database. *Automation in Construction*, 81, 369-388.
- Tao, X., Liu, Y., Wong, P. K. Y., Chen, K., Das, M., & Cheng, J. C. (2022). Confidentiality-minded framework for blockchain-based BIM design collaboration. *Automation in Construction*, 136, 104172.
- Wang, N., Issa, R. R., & Anumba, C. J. (2022). NLP-based query-answering system for information extraction from building information models. *Journal of computing in civil engineering*, 36(3), 04022004.
- Wang, Z., Sacks, R., Ouyang, B., Ying, H., & Borrmann, A. (2024). A Framework for Generic Semantic Enrichment of BIM Models. *Journal of Computing in Civil Engineering*, 38(1), 04023038.
- Ying, H., & Lee, S. (2020). Automatic detection of geometric errors in space boundaries of IFC-BIM models using monte carlo ray tracing approach. *Journal of Computing in Civil Engineering*, 34(2), 04019056.
- Ying, H., Zhou, H., Degani, A., & Sacks, R. (2022). A two - stage recursive ray tracing algorithm to automatically identify external building objects in building information models. *Computer - Aided Civil and Infrastructure Engineering*, 37(8), 991-1009.
- Jeon, K., Lee, G., Kang, S., Roh, H., Jung, J., Lee, K., & Baldwin, M. (2021). A relational framework for smart information delivery manual (IDM) specifications. *Advanced Engineering Informatics*, 49, 101319.
- Zheng, J., & Fischer, M. (2023). Dynamic prompt-based virtual assistant framework for BIM information search. *Automation in Construction*, 155, 105067.

OPTIMIZED LANGUAGE-EMBEDDED 3DGS FOR REALISTIC MODELING AND INFORMATION STORAGE OF HISTORICAL BUILDINGS

Zhenyu LIANG¹, Jeff Chak Fu CHAN², Jiaying ZHANG³, Zhaolun LIANG⁴, Boyu WANG⁵,
Mingzhu WANG⁶ and Jack C.P. CHENG⁷

- 1) Ph.D. Candidate, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: zhenyu.liang@connect.ust.hk
- 2) M.Phil. Candidate, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: cfchanay@connect.ust.hk
- 3) Ph.D., Post-doc. Fellow, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: jzhangfh@connect.ust.hk
- 4) Ph.D. Candidate, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: zliangaq@connect.ust.hk
- 5) Ph.D., Post-doc. Fellow, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: bwangbb@connect.ust.hk
- 6) Ph.D., Asst. Prof., Department of Architecture and Civil Engineering, City University of Hong Kong, Hong Kong, SAR. Email: m.wang@cityu.edu.hk
- 7) Ph.D., Prof., Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: cejcheng@ust.hk

Abstract: A realistic and informative 3D digital model of historical buildings holds significant value for heritage preservation, public education, and cultural dissemination. Traditional digital representations, such as Heritage Building Information Modeling, panoramic images, LiDAR point clouds, photogrammetric mesh models, face limitations in user interaction and engagement. The automatic generation of a semantically enriched 3D model requires advanced scene-understanding capabilities. Pre-trained zero-shot methods struggle with domain-specific knowledge in heritage component semantics, while CNN-based approaches demand extensive manual effort for dataset preparation and model training. Therefore, this study proposes an optimized language-embedded 3DGS framework for the digitalization of historical buildings. It involves three steps: (1) data preparation of on-site images and relevant text; (2) component segmentation by the integration of SAM and MLLM; (3) scene reconstruction using the language-embedded 3DGS. The combination of SAM's localization ability and MLLM's in-context learning achieves 95.6% accuracy in the semantic segmentation of historical building components, requiring only a single annotated sample for each component category. Compared with previous methods, our language-embedded 3DGS model accurately captures complex semantics while providing realistic appearance and convenient navigation. The generated 3D model can be further integrated with an LLM-based chatbot assistant to achieve open-vocabulary and vague searches. This framework was validated on the Shishi Sacred Heart Cathedral in Guangzhou, China, offering a novel digital solution for the protection and sustainment of historical buildings.

Keywords: Language-embedded 3D Gaussian Splatting, Multimodal Large Language Models, Historical Buildings, 3D Reconstruction, Segment Anything Model, Scene Understanding

1. INTRODUCTION

Historical buildings record the history and social transformations of a region. Those with religious significance, such as cathedrals, often serve as spiritual homes for communities. Effective documentation, preservation, and dissemination of these heritages are essential for cultural protection and sustainability. To capture and record the current conditions of historical buildings, advanced techniques are employed to create digital models that reflect geometric shapes, textures, and attribute information, including component semantics and defects. A 3D heritage model with precise geometry and comprehensive information is valuable for various applications such as archiving, maintenance, navigation, and education (Yang, 2020).

Several types of digital models are employed for the preservation and dissemination of historical buildings. Heritage Building Information Modeling (HBIM) is widely used due to its superior capability to manage and document historic structures with parametric objects (Cheng, 2024). HBIM can be further transformed into mechanical models, such as Finite Element Analysis models, to comprehensively analyze structural performance and proactively prevent defect development (Ursini, 2022). However, HBIM struggles to accurately reflect the realistic appearance and aesthetics of historical buildings, which limits its applications in areas involving public interaction, such as remote viewing and on-site navigation. Besides, panoramic images with informative labeling are popular for effectively displaying scene textures while storing a certain amount of information. Nevertheless, users cannot move continuously within these panoramic scenes, and texture distortion may occur, diminishing the user experience. Additionally, reality-capturing techniques, such as LiDAR point cloud scanning and photogrammetric mesh 3D reconstruction, can accurately capture geometrical shapes and textures (Croce, 2021; Pritchard, 2017). However, they do not provide a first-person perspective for user visualization and interaction,

which is inconvenient for navigation and education in VR environments or on mobile devices. To enhance user interaction and broaden public engagement with digital models of historical buildings for effective cultural dissemination, it is crucial to propose a novel digital representation that overcomes these existing limitations.

3D Gaussian Splatting (3DGS) is a recently developed 3D reconstruction technique based on generative artificial intelligence (Kerbl, 2023). This method takes images or videos as input and employs an iterative learning process to train an explicit model composed of 3D Gaussian ellipsoids for displaying the scenes. This technology can render realistic first-person view images, supporting fast training and rendering speeds, allowing users to navigate within the scene. It is an ideal 3D digital representation for user interaction with historical buildings.

However, the basic 3DGS model is primarily used for rendering colorful scenes, lacking object-based information. Consequently, several studies have focused on the semantic enrichment of the 3DGS model (Qin, 2024; Shi, 2024). Rather than merely adding semantic labels to the scene, language features are preferred, as they enable users to interact with the 3DGS model through natural language. The automatic generation of the semantically enriched 3DGS model directly from the input images or videos is the chasing of the previous studies, as well as this paper. This necessitates an accurate semantic understanding of the input images, followed by the segmentation of target objects. Previous studies, such as LangSplat (Qin, 2024), employ pre-trained zero-shot models like Contrastive Language-Image Pre-training (CLIP) for automatic scene understanding. Although these zero-shot models enhance the implementation convenience, their capabilities are constrained by the pre-training dataset. In the context of historical buildings, the target objects for segmentation include components such as flying buttresses, arched windows, and spires. These semantics require strongly domain-specific knowledge and may not be presented in the pre-training dataset. Consequently, directly applying pre-trained zero-shot models to historical buildings may result in incorrect semantic segmentation of input images, leading to erroneous information storage in the 3DGS model. Conversely, the CNN-based models can effectively address domain-specific segmentation problems, but they demand significant manual effort for label annotations and involve a time-consuming training process. To ensure accurate information storage of historical buildings in the 3DGS model while balancing manual effort, it is crucial to develop a component semantic segmentation method using a few-shot dataset.

Multimodal Large Language Models (MLLMs) are renowned for their reasoning and in-context learning abilities (Zhang, 2024). These models can learn from a few input samples, then achieve impressive recognition performance for similar objects. However, they face challenges with weak localization ability (Yin, 2023), which can result in inferior segmentation results.

To fulfill the aforementioned requirements and address the existing problems, this paper proposes an optimized language-embedded 3DGS framework for realistic modeling and accurate information storage of historical buildings. The framework consists of three steps: (1) data preparation of on-site images and relevant text; (2) component segmentation by the integration of SAM and MLLM; (3) scene reconstruction using the language-embedded 3DGS. Our method was validated on the Shishi Sacred Heart Cathedral in Guangzhou, China. Results show that our method can achieve 95.6% accuracy of component segmentation using few manual-annotated samples. The generated 3DGS model accurately represents a realistic and informative scene, which can be further integrated with LLM assistants for open-vocabulary and vague searches. This innovative digital model of historical buildings enhances user interaction and is valuable for cultural protection.

2. 3D GAUSSIAN SPLATTING TECHNIQUE

3DGS can be conceptualized as a point cloud, where each point is associated with a Gaussian ellipsoid. From user-specified viewpoints, it renders and displays the information encoded in the Gaussians onto images. The basic RGB-based 3DGS model includes the following parameters for each 3D Gaussian: position x , covariance matrix Σ (defining its shape), opacity α , and spherical harmonics (SH) coefficients c (defining its color). The expression of 3D Gaussian characterized by a mean μ is shown in Equation 1.

$$G(x) = e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)} \quad (1)$$

3D Gaussians are the ellipsoids in 3D space. They are then projected into the 2D image space as ellipses for rasterization-based rendering, which is called splatting. Given the viewing transformation W and Jacobian of the affine approximation of the projective transformation J , the splatting is to project the 3D covariance matrix Σ into 2D covariance matrix Σ' using Equation 2.

$$\Sigma' = JW\Sigma W^T J^T \quad (2)$$

The 2D ellipses are then sorted by depths of Gaussians and cumulated to render images. Given the position of a pixel $v \in \{1, \dots, H\} \times \{1, \dots, W\}$ and a sorted list of Gaussians N , the final color $c(v)$ of this pixel is calculated by the α -blending as shown in Equation 3.

$$c(v) = \sum_{i \in N} c_i a'_i \prod_{j=1}^{i-1} (1 - a'_j) \quad (3)$$

where c_i is learned color in 3D Gaussian ellipsoids, and final opacity a'_i is the multiplication result of learned opacity α_i in 3D Gaussian ellipsoids and 2D covariance.

In the training process of the 3DGS, input images are used to optimize parameters within each 3D Gaussian and to adjust the number of Gaussians. These input images act as ground truth, compared with rendered images in each iteration to calculate the loss for optimization. For further details, refer to (Kerbl, 2023).

3. METHODOLOGY

Our proposed framework is illustrated in Figure 1. Initially, on-site images and relevant textual materials are collected as the overall input. Then, the SAM and MLLM are integrated for the semantic segmentation of target components in the images, requiring only one annotation per component. Subsequently, the semantic segmentation masks are utilized to generate 3DGS scenes, embedding language features to facilitate open-vocabulary and vague searches. Two datasets, front view and side view, are collected for Shishi Sacred Heart Cathedral, as different perspectives may contain different components. The operational environment is the computer equipped with an NVIDIA A40 GPU and an Intel Xeon Platinum 8358P CPU.

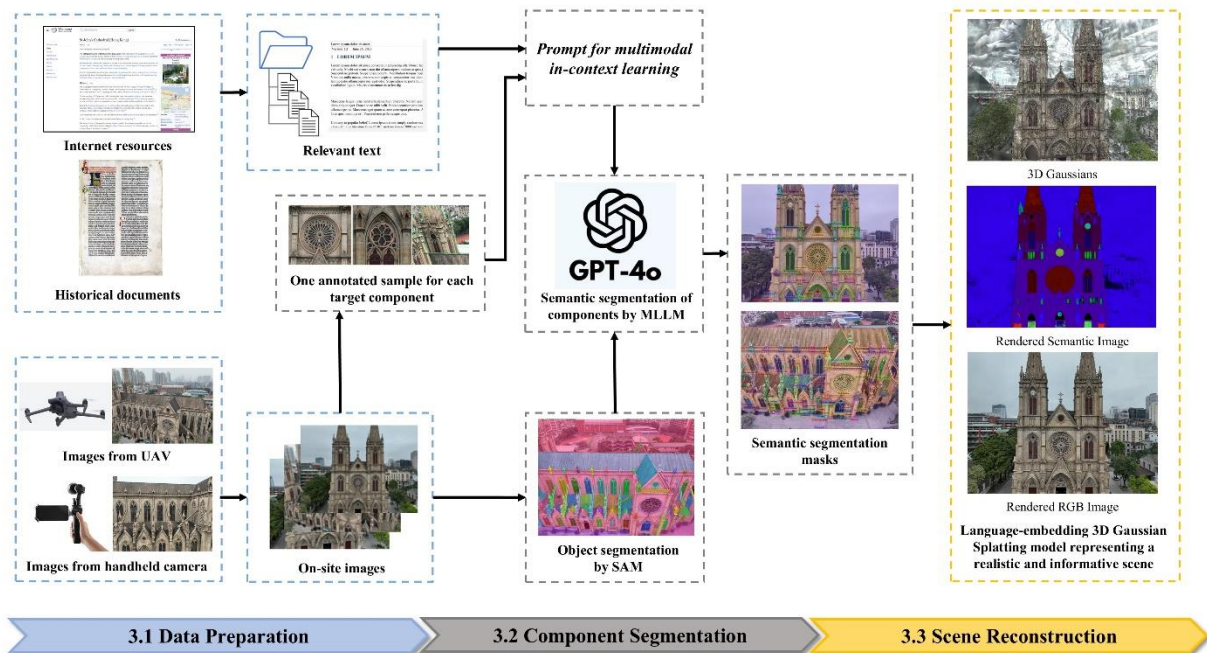


Figure 1. Overall framework of the optimized language-embedded 3DGS

3.1 Data Preparation of On-site Images and Relevant Text

In the first step, on-site images are captured for 3DGS generation, and textual materials related to component characteristics are prepared for subsequent MLLM-assisted semantic segmentation.

To ensure high-quality 3DGS reconstruction, multi-angle images must be captured to fully cover the target scene, providing a faithful viewing experience from various perspectives of the 3DGS model. As illustrated in Figure 2, images of the Cathedral's front façade were taken from multiple angles and distances. A total of 1,371 images at a resolution of 2K pixels were captured for the Shishi Sacred Heart Cathedral using UAV and handheld cameras.



Figure 2. Multi-angle images captured for the frontier façade

The exploration of textual materials related to the Cathedral was conducted through internet sources and historical documents. The relevant texts primarily include descriptions of the Cathedral's most distinctive components, which serve as a part of input for the MLLM to enhance component understanding and segmentation. Examples of these textual descriptions include:

- **Rose window:** A large, circular window with intricate stone tracery radiating from the center,

resembling the petals of a rose.

- **Flying buttress:** A semi-circular or arched exterior support that projects from the upper portion of a church wall over the roof of the church's aisle or chapel below.
- **Spire:** A tall, slender, and pointed structure rising from the top of a tower or the roof of a cathedral, typically made of stone or wood, adding height and verticality to the overall design.

3.3 Scene Reconstruction using the Language-embedded 3DGS

Once accurate semantic segmentation masks of the components are generated, they are employed in developing the language-embedded 3DGS model. Compared with the original 3DGS, this model extends a semantic attribute to store language features as a vector. These features enable natural language search through cosine similarity calculations.

The CLIP text encoding module (Radford, 2021) E_{text} is utilized to generate the language feature L of pixel v with the semantic label S , as shown in Equation 4.

$$L(v) = E_{text}(S(v)) \quad (4)$$

However, these language features are 512-dimensional vectors, which can easily cause "out of memory" issues if directly added to Gaussians. Inspired by (Qin, 2024), a scene-wise autoencoder is developed to reduce the dimensionality by mapping the 512-dimensional vector to a 3-dimensional scene-specific vector. These 3-dimensional latent features are then used to supervise the learning of the 3D language field and facilitate rendering to visualize semantic embedding conditions in 3D scenes. The rendering of the latent features f on pixel v also employs the α -blending, similar to color rendering, as shown in Equation 5.

$$f(v) = \sum_{i \in N} f_i a'_i \prod_{j=1}^{i-1} (1 - a'_j) \quad (5)$$

where f_i is learned latent feature in 3D Gaussian ellipsoids.

The training process of the language-embedded 3DGS involves two steps. First, a colorful 3DGS model is developed based on the original 3DGS pipeline. Following this, the number and learned attributes of Gaussians are fixed in the colorful 3DGS model, and latent feature maps are input to enable the model to learn component semantics through the same rasterization process. In this experiment, the first step involves 100,000 training epochs, and the second step involves 30,000 epochs.

Figure 3 demonstrates the realistic scene rendering capabilities of 3DGS technology. Notably, the scenes in 3DGS closely resemble real-life environments. Additionally, users can freely navigate within 3DGS, making it highly suitable for applications with significant user interaction.

Figure 4 illustrates the rendered semantic masks, also known as the 3-dimensional latent features. The edges of different objects are clear and distinct, indicating that the semantic information has been accurately learned and stored in the 3DGS model. However, the lower roof in the side view appears somewhat blurred, likely due to occlusion by the flying buttress, preventing complete segmentation. Additional views from different angles are needed to clearly capture the roof and enhance the scene information through improved segmentation.

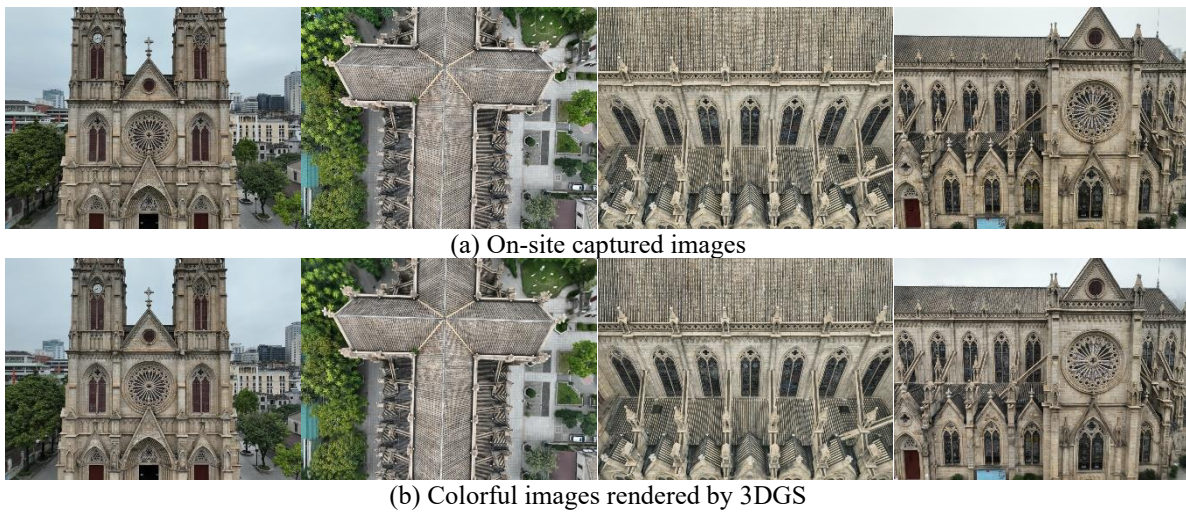


Figure 3. Comparison of on-site captured images and 3DGS-rendered images

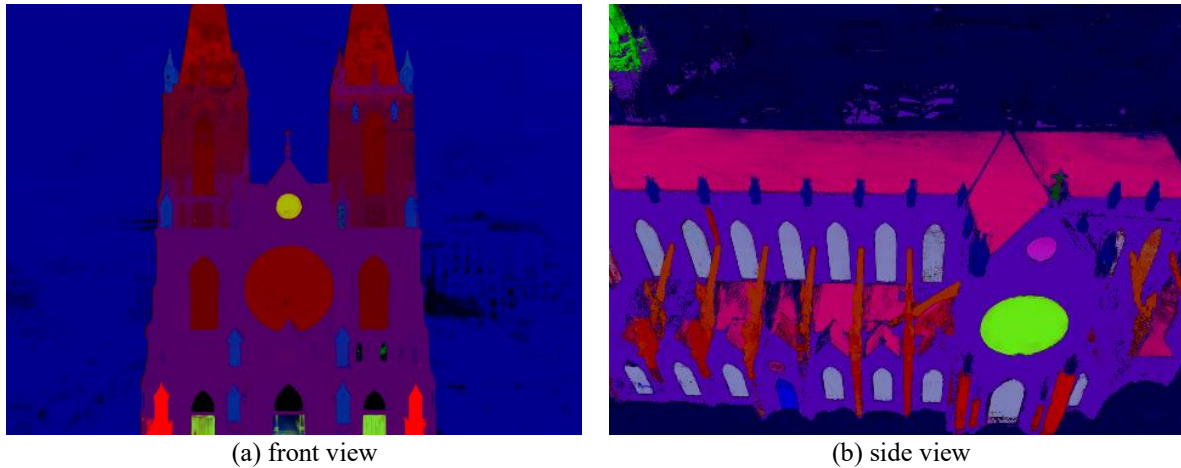
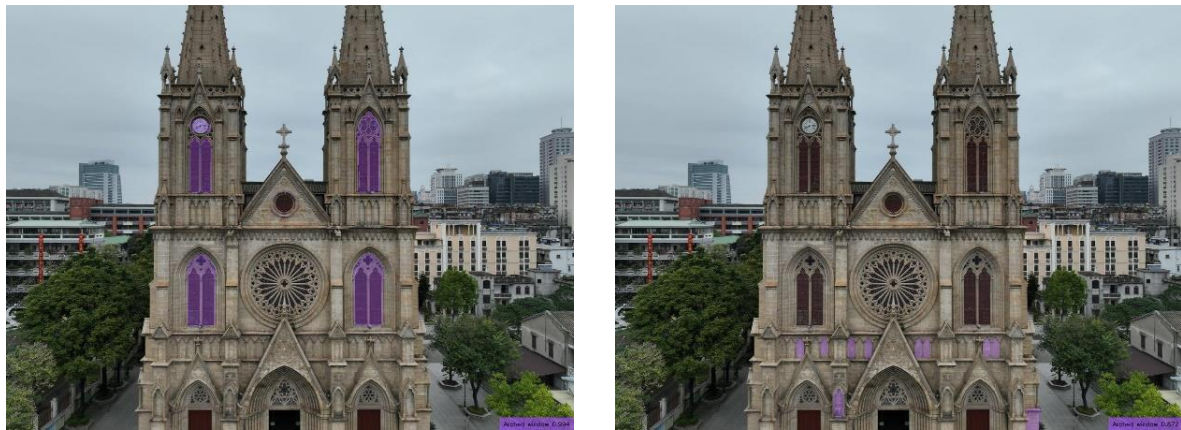


Figure 4. 3DGS-rendered semantic masks

4. VALIDATION

The validation analyses focus on two aspects to demonstrate the superiority of the proposed pipeline. First, our optimized language-embedding 3DGS is compared with the language-embedding 3DGS using the pre-trained model for scene understanding (LangSplat). Second, a quantitative analysis is conducted on the accuracy of MLLM-assisted semantic segmentation of components.

The comparison between our method and LangSplat focuses on the ability to effectively handle domain-specific knowledge and accurately store complex component semantics. Since both methods store language features in the 3DGS models, a cosine similarity search is conducted on every pixel using the target component name. As shown in Figure 5, the search for the specific terms "Arched window" and "Spire" was conducted on two 3DGS models. Our method successfully highlights the appropriate locations on the images, indicating that the semantics stored in the 3DGS are both accurate and meet the specialized requirements of historical buildings. In contrast, LangSplat is unable to identify these terms because it relies on CLIP for scene understanding, which is constrained by pre-trained data and struggles to expand its knowledge domain. Consequently, it cannot accurately identify complex objects or concepts in specific fields. This result demonstrates that our proposed optimized language-embedding 3DGS effectively addresses automatic 3D scene reconstruction with semantic complexity.



(a) Arched window



(b) Spire

Figure 5. Comparison of our method (left) with LangSplat (right)

Our semantic segmentation method employs MLLM to recognize SAM-segmented objects, ensuring that every object in the images is labeled. The accuracy of these predicted labels is crucial. Therefore, validation is performed to check the correctness of semantic labels for each component.

Both front and side view images were validated. In the front view sample shown in Figure 6(a), 41 out of 42 components were accurately recognized. In the side view sample shown in Figure 6(b), 60 out of 66 components were accurately identified. The overall recognition accuracy for all components is 95.6%, demonstrating that our SAM-MLLM integrated method achieves high-quality semantic segmentation with only a few annotated samples as input. The highly accurate semantic segmentation with less manual effort is also valuable for other scenes understanding tasks.


 (a) Front view, (accuracy = $41/42 = 97.6\%$)

 (b) Side view (accuracy = $60/66 = 90.9\%$)

Figure 6. Quantitative analysis of the accuracy of MLLM-assisted semantic segmentation. The left shows the generated segmented masks, and the right shows the ground truth.

5. DISCUSSION

The generated language-embedding 3DGS model can be further integrated with an LLM-based chatbot interface to facilitate user-friendly searching, as shown in Figure 7 and 8. This interface, developed via Autogen Studio, utilizes the LLM heritage assistant powered by GPT-4o, configured with a specific system message for instruction. With component semantics embedded as language features, searching operations are performed using cosine similarity calculations, which are implemented as a backend function in the LLM heritage assistant.

In Figure 7, open-vocabulary searches were initially conducted for testing. User input prompts included specific terms like "spire" and "pinnacle". Following the search, the corresponding components were successfully selected and highlighted. However, these terms represent specialized knowledge, which may be challenging for non-expert users to articulate accurately. Therefore, in Figure 8, a vague search is performed to test the generalizability of our method. Users employed descriptive phrases like "big round window" instead of standard terms like "rose window." The results demonstrate that the correct components can still be identified and highlighted, indicating that our method is highly user-friendly. Additionally, the LLM heritage assistant, capable of accessing historical and contextual information from the web, allows for the effective integration of online knowledge with reconstructed 3D models. This integration holds significant value for the education and cultural dissemination of historical buildings.

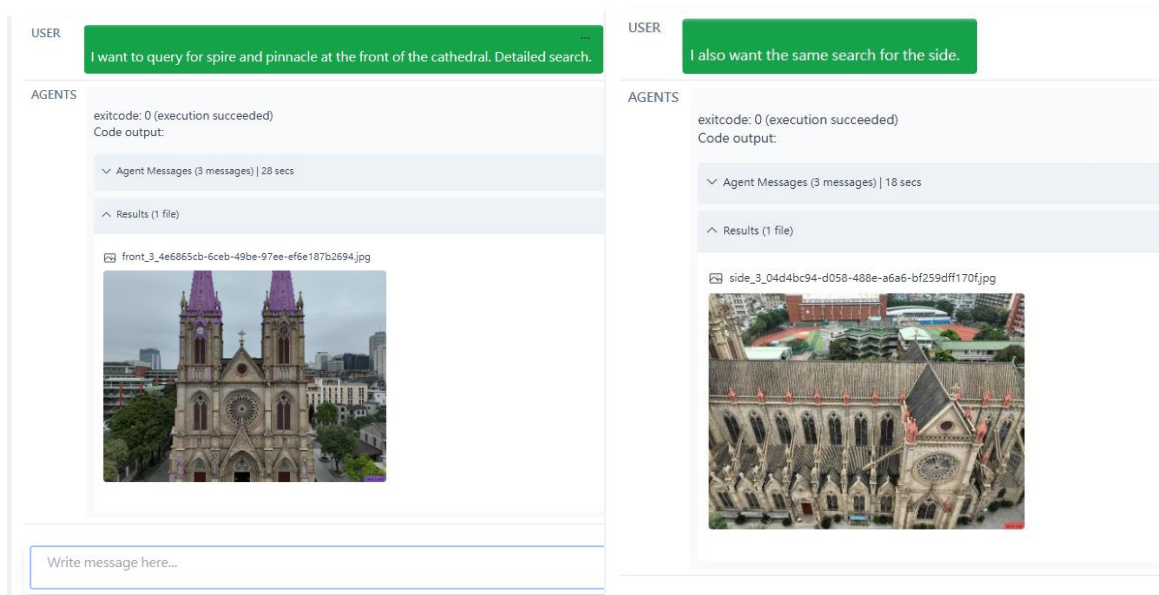


Figure 7. Results of open-vocabulary searches

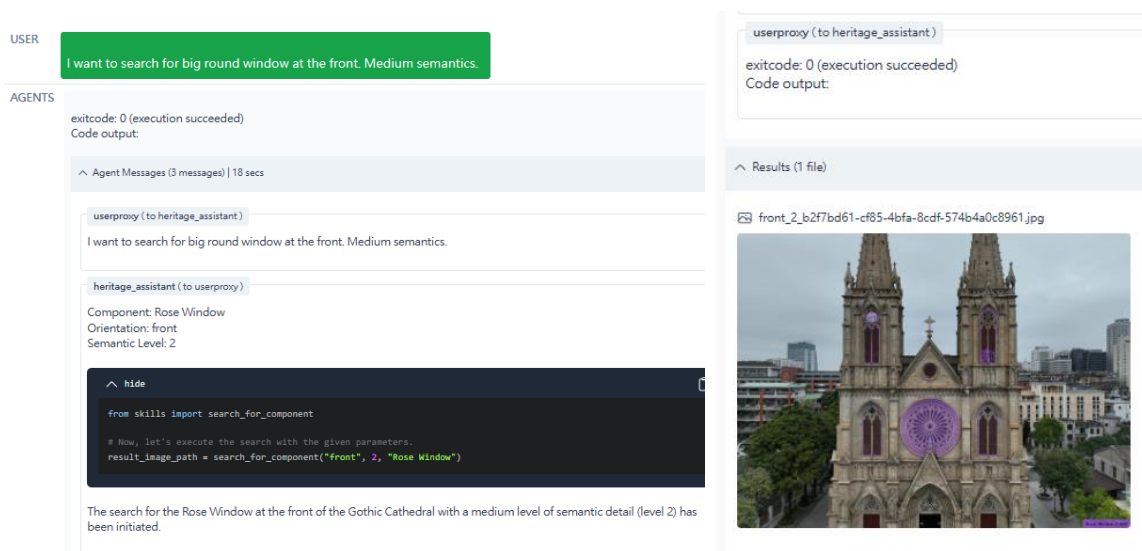


Figure 8. Results of vague search

6. CONCLUSIONS

This paper proposes an optimized language-embedded 3DGS framework for the automatic and accurate modeling of realistic and informative 3D scenes of historical buildings. The framework includes: (1) preparation of on-site images and relevant textual material; (2) semantic segmentation of historical building components using SAM-MLLM integrated method; (3) development of a language-embedded 3DGS model for scene representation. Our semantic segmentation method achieves 95.6% accuracy for component segmentation using only one annotated sample per component category. The optimized language-embedded 3DGS method outperforms the previous model in accurate domain-specific semantics recognition and storage. The 3DGS-based digital representation of historical buildings also facilitates user interaction through an LLM-based chatbot assistant for open-vocabulary and vague searches. This scene-realistic, semantic-enriching, convenient-navigating and easy-interacting 3DGS model brings significant value for the heritage preservation, public education and cultural dissemination.

Future work will extend our study to include the reconstruction of indoor scenes of historical buildings, aiming to better document and disseminate the intricate interior decorations. Currently, the MLLM used is GPT-4o, which requires the API. Future efforts will also focus on developing the open-source MLLM-assisted method, which will be beneficial for private deployment.

ACKNOWLEDGMENTS

The authors would like to express their gratitude to the management team of Shishi Sacred Heart Cathedral (Guangzhou, China) for providing access to the on-site investigation. The authors would like to acknowledge the support by RGC Theme-based Research Scheme 2023/24 T22-606/23-R.

REFERENCES

- Cheng, J. C., Zhang, J., Kwok, H. H., & Tong, J. C. (2024). Thermal performance improvement for residential heritage building preservation based on digital twins. *Journal of Building Engineering*, 82, 108283.
- Croce, V., Caroti, G., De Luca, L., Jacquot, K., Piemonte, A., & Véron, P. (2021). From the semantic point cloud to heritage-building information modeling: A semiautomatic approach exploiting machine learning. *Remote Sensing*, 13(3), 461.
- Kerbl, B., Kopanas, G., Leimkühler, T., & Drettakis, G. (2023). 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Trans. Graph.*, 42(4), 139-1.
- Pritchard, D., Sperner, J., Hoepner, S., & Tenschert, R. (2017). Terrestrial laser scanning for heritage conservation: The Cologne Cathedral documentation project. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4, 213-220.
- Qin, M., Li, W., Zhou, J., Wang, H., & Pfister, H. (2024). Langsplat: 3d language gaussian splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 20051-20060).
- Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. (2021, July). Learning transferable visual models from natural language supervision. In *International conference on machine learning* (pp. 8748-8763). PMLR.
- Shi, J. C., Wang, M., Duan, H. B., & Guan, S. H. (2024). Language embedded 3d gaussians for open-vocabulary scene understanding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 5333-5343).
- Ursini, A., Grazzini, A., Matrone, F., & Zerbinatti, M. (2022). From scan-to-BIM to a structural finite elements model of built heritage for dynamic simulation. *Automation in Construction*, 142, 104518.
- Yang, X., Grussenmeyer, P., Koehl, M., Macher, H., Murtiyoso, A., & Landes, T. (2020). Review of built heritage modelling: Integration of HBIM and other information techniques. *Journal of Cultural Heritage*, 46, 350-360.
- Yin, S., Fu, C., Zhao, S., Li, K., Sun, X., Xu, T., & Chen, E. (2023). A survey on multimodal large language models. *arXiv preprint arXiv:2306.13549*.
- Zhang, D., Yu, Y., Dong, J., Li, C., Su, D., Chu, C., & Yu, D. (2024). Mm-llms: Recent advances in multimodal large language models. *arXiv preprint arXiv:2401.13601*.

Sustainable Development Management Method for Construction Projects Using Large Language Model

Xingbo Gong¹, Yuqing Xu², Helen H.L. Kwok³, Xingyu Tao⁴, and Jack C.P. Cheng⁵

1) Ph.D. Candidate, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: xgongai@connect.ust.hk

2) Ph.D. Candidate, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: yuqing.xu@connect.ust.hk

3) Ph.D., Postdoc., Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: hlkwokab@connect.ust.hk

4) Ph.D., Research Assistant Professor., Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: turbetao@ust.hk

5) Ph.D., Professor., Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: cejcheng@ust.hk

Abstract: As sustainable development is gaining more and more attention, the construction industry continues to explore this aspect. Both Sustainable Development Goals (SDGs) and Environmental, Social and Governance (ESG) provide management goals for corporate sustainable development and assessment. However, due to the complexity of construction events and multiple data sources, sustainable development management in the construction industry is still hindered by the need for a large amount of labor costs. Therefore, this paper proposes an LLM-based sustainable development data processing framework for construction, which achieves three goals: (1) identifying indicators of SDGs and ESG assessment frameworks for construction projects, (2) mapping sustainable development indicators to construction events and data, and (3) developing an LLM-based localized data processing framework for construction sustainability. The proposed method has been tested in a design scenario and can realize the rapid processing of construction project data and provide information and information sources related to sustainable development goals. It provides efficient and manageable technical method suggestions for realizing automated report generation or associated traceability of sustainable development of construction projects.

Keywords: Construction, ESG (Environmental, Social and Governance), Large Language Model (LLM), Sustainability, Sustainable Development Goals (SDGs).

1. INTRODUCTION

As a major consumer of resources and energy, the construction industry has always been one of the core industries for sustainable development. Since 2015, when governments around the world agreed to implement a set of sustainable development goals (SDGs), it has provided guidance for the construction industry to promote sustainable development from environmental, social and economic perspectives (Pradhan et al., n.d.). The United Nations 2030 Agenda covers many environmental, economic and social development issues, such as health and well-being, poverty, hunger, quality education, gender equality, climate action, water, sanitation, energy and environment, and peace and social justice. The 17 Sustainable Development Goals and 169 related targets contained in the SDGs proposed on the agenda provide guiding recommendations for the management of construction industry (Fei et al., 2021). The 2030 Agenda and its goals and targets also represent a long-term political framework for companies to contribute to sustainable development. However, although many companies have proposed their own sustainable development goals, the design of meaningful goals, sustainable strategies or key performance indicators has not been widely proposed. Corporate participation in the SDGs is limited and largely symbolic. This is partly due to the market's vague setting of performance evaluation indicators and scope related to corporate sustainability or SDGs (Johnsson et al., 2020).

Now, all large companies should understand their environmental and social footprint and disclose, and report related activities. This is achieved through a series of standards, frameworks and indicators related to the so-called ESG (Environmental, society and governance). Although development and environmental protection may be considered a task for the government, business plays an increasingly active role in achieving sustainable development goals (Chaitanya et al., 2016). These efforts and recognition come from legal and regulatory measures and market behaviors such as ESG investment. Sustainable business models have been proven to bring advantages to companies in the market and resist possible risks in the future (Banihashemi et al., 2017; De Franco, 2020). As an evaluation framework for corporate behavior, ESG can provide detailed indicators for corporate managers in a more detailed manner, so that companies can provide a basis for considering relevant factors such as the environment in management and operations. By exploring the correspondence between SDGs and ESG, it will help companies collect and share relevant data and provide clear quantitative indicators for qualitative SDGs goals.

In the assessment of the level of sustainable development of enterprises, ESG reports require a lot of data support. In the construction industry, these data come from the company's construction projects, the company's

own operations and the environment of the construction area (Gong et al., 2024). At the same time, a large amount of information throughout the life cycle of each project needs to be exhaustively counted and processed to meet the requirements of ESG evaluation and sustainable development reporting (Mishra, 2023). The huge workload requires a lot of labor to process this information. This hinders the enthusiasm of enterprises to evaluate and manage the ESG performance of construction projects. So far, large language models (LLMs) have made significant breakthroughs in many automated learning tasks by using generative pre-trained transformers (GPTs) to perform unsupervised training on large corpora and datasets (Pu et al., 2024). These tasks include machine translation, text classification, intelligent question answering, etc. LLMs can accurately extract entities and entity relationships through the ability of language understanding. These characteristics have also led to some expansions and attempts in the construction industry, such as building energy analysis and urban information management.

However, the current LLM can only process limited information. A large amount of information on construction projects has little to do with sustainable development, and this additional information will affect the operation of LLM (Giudici et al., 2023). Which data and construction events are related to SDGs and ESG, and which sustainable development goals are related to these construction events need to be further considered and analyzed. Therefore, to overcome the challenges of processing sustainable development-related data in construction projects and improve its automation in project management, this paper proposes a data processing framework for sustainable development goals of construction projects based on LLM. Through automated text information processing, LLM demonstrates efficiency advantages in corporate SDGs and ESG management.

2. METHOD

Figure 1 shows an overview of the proposed LLM-based approach for assessing the level of sustainable development in construction projects. This approach has three modules. The first step is the analysis of SDGs and ESG evaluation indicators. This will screen out sustainable development goals and ESG indicators related to the construction industry and construction projects. The second step is to establish the correspondence between events and data in construction projects and these indicators. The established data map will describe the relationship between sustainable development goals and construction processes and provide a basis for further data processing. The third step is to establish a localized database so that LLM can obtain construction data. And further structure the text information so that the automatically processed content can provide a reference for corporate sustainable development.

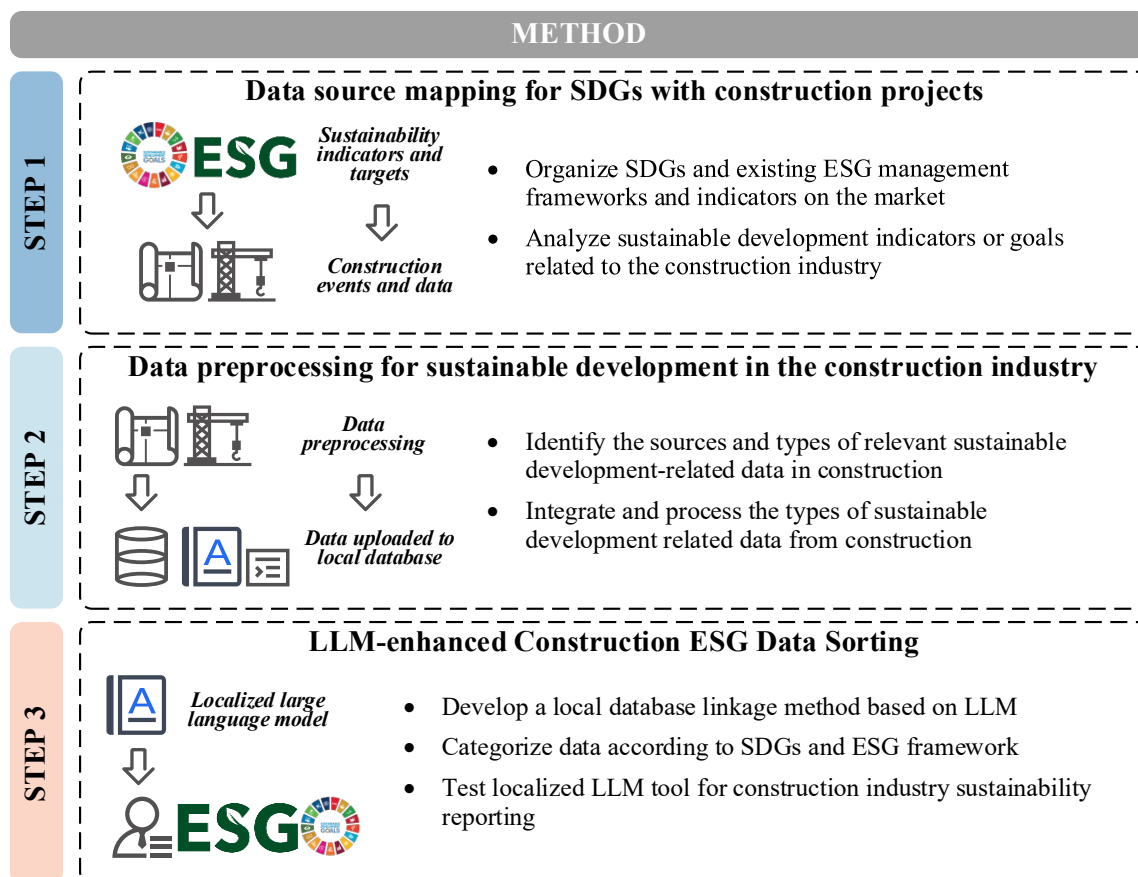


Figure 1. Method overview

2.1 Construction Sustainable Development Index

SDGs provide a new opportunity for the construction industry to expand its focus from the environmental aspect of sustainability. The behavior of enterprises and society is closely related to the SDGs during the design, construction, delivery and operation of construction projects. For example, resource utilization during construction, low-carbon model design during the operation phase, and pollution and recycling during building demolition activities. According to the BDG, the construction industry is related to each of the 17 SDGs. Among them, the more representative is that the building materials used in construction projects can directly affect the achievement of the SDGs, especially SDG 3, SDG 7, SDG 9, SDG 11, SDG 12, SDG 13 and SDG 15, and indirectly affect SDG 2, SDG 5, SDG 10 and SDG 16 (Fei et al., 2021). The relevance of SDGs to construction projects becomes more complex for the broader concept of construction projects. For example, offshore structures and bridges are relevant to SGD14. These structures consider material and environmental impacts, as underwater life may be affected by factors such as corrosion of metal materials. At the same time, ESG management of construction projects has also shown its importance today. The correlation between innovation and sustainable practices in construction projects and corporate performance has been widely studied. It can measure the performance of enterprises in these areas and serve investors and social participants. Both SDGs and ESG provide goals and foundations for the sustainable management of construction projects (Litvinenko et al., 2022). However, the general sustainable development goals and indicators are not accurate for the construction industry. Specific indicators or goals still need to be accurately identified to meet the management of relevant indicators in complex construction affairs.



Figure 2. SDGs and ESG indicators related to Construction Projects

As shown in Figure 2, this paper proposes a correspondence table between SDGs and ESG (MSCI) indicators. The selected ESG indicators are similar to the scope of SDGs, both of which are general and industry wide. However, they have different focuses in the construction industry. In this correspondence, only the most direct sustainable development goals, the relationship between ESG impacts and construction activities are considered. In terms of environmental indicators, SDG 6, SDG 7, SDG 9, SDG 11, SDG 12, SDG 13, SDG 14 and SDG 15 are all relevant. In terms of ESG indicators, they include carbon emissions, carbon footprint, water resources and waste. These goals and indicators are more relevant to construction projects. Similarly, in terms of social indicators, human resources, worker health and data security are more important in construction projects. In governance indicators, ESG management often follows the company's financial statements or annual accounting information, which is related to construction projects, but this information cannot be directly obtained in a single project. Therefore, these goals are not considered as the main consideration in the relevant affairs of construction projects.

2.2 Data Map for Sustainable Construction

After determining the sustainable development indicators related to construction projects, Figure 3 shows the correspondence between these indicators and construction-related data and data sources. In the scope of construction projects, four data sources are identified, including pre-construction preparation documents, supply chain information of construction projects, construction site data and corporate reports (Lou et al., 2021; Yan et al., 2020). The sources of this information are complex and require a lot of manual processing. For example, carbon emission information has a high correlation with SDG 13, and this data comes not only from energy use (such as operational carbon), but also from the use of materials in construction projects (such as embodied carbon) (Xu et al., 2024). For other aspects such as biodiversity, land use and green technology data at the construction site, they often come from the environmental impact assessment and reports of the construction. By corresponding SDGs, ESG and construction project data, sustainable development assessment or environmental management personnel can more conveniently collect and manage relevant content.

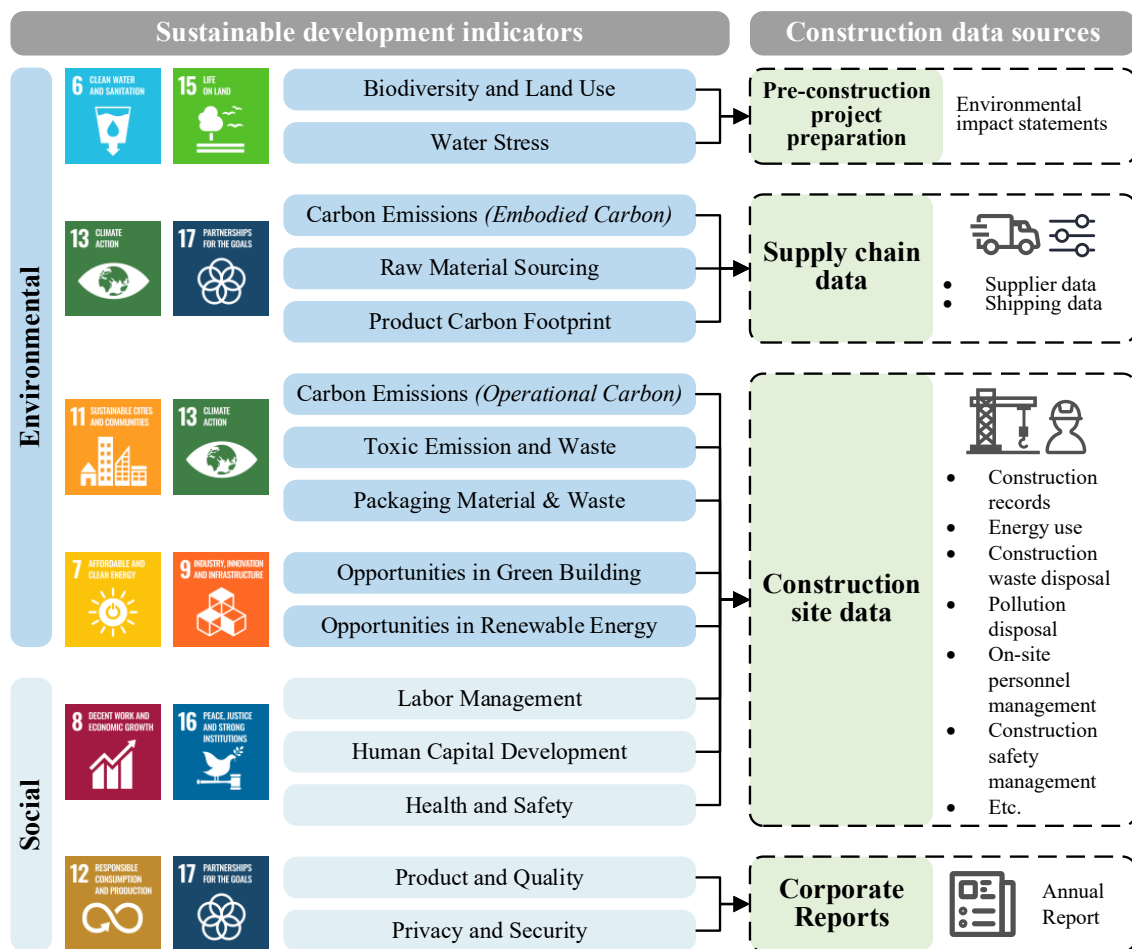


Figure 3. Data Mapping between SDGs and ESG indicators to Construction Data

2.3 LLM-based Data Processing of Sustainable Development of Construction

Environmental impact assessment, construction records and reports are directly related to the sustainable development management of construction projects. However, large amounts of text content require high labor costs to process. LLMs can process these contents quickly. Therefore, this paper links a local database with a localized large prediction model to achieve rapid organization and management of sustainable development affairs in construction projects. In the local database, reports and records in the construction project will be uploaded. At the same time, according to the correspondence between the goals and indicators and the construction data in Figure 3, the uploaded files will be attached with one or more indicators at the same time. Based on these data, an information collection and output program is developed for LLM. The pseudo code of the program is shown in Table 1.

Table 1. Pseudo Code for LLM and Sustainable Data Local Database Connection

Step 1: Connect to MySQL database
MySQLTool.connection = pymysql.connect(→ host='localhost', database='Sustainable data for Construction'
Step 2: Execute query statement
cursor = connection.cursor()
Input: Sustainable data indicator or File name or Upload time
sql = f"SELECT * FROM 'Sustainable data' where indicator = {indicator} or name = '{query}' or filetime = '{query}'" → print("sql: ", sql)
cursor.execute(sql)
Step 3: Close the cursor object and database connection
cursor.close()
Step 4: Call langchain.tools import BaseTool
Call the above query statement from the Sustainable database
Step 5: Structured output
model_with_structure = model.with_structured_output(indicator name + file source)
structured_llm = llm.with_structured_output(Sustainable)
structured_llm.invoke("indicators related data")
Output: ResponseFormatter(answer="indicator name" + "file source".)
End

It can link the local MySQL database and LLM. This process is based on the linking mode of Langchain (Asyrofi et al., 2023). At the same time, during the LLM output process, the relevant sustainable development indicators and the data location in the construction project can be directly obtained through specific prompt words. This can help managers check the original information to ensure and correct the accuracy of the LLM output information.

3. PROOF OF CONCEPT

After completing the above development, this paper verifies the designed construction project sustainable development data processing framework. The verification environment of the framework is to verify through a design scenario, in which the data in the project comes from an infrastructure construction process in Hong Kong. In this verification scenario, the environmental impact assessment report of the construction project is disassembled so that it can be directly applied to the designed construction sustainable development indicators. At the same time, the supply chain data in the construction project (including material suppliers, construction project location, transportation records and energy consumption), construction site data (human resource records, construction safety reports, material usage, energy and water usage, etc.) and the company's current operation report and strategic development report are also submitted, see Figure 4 (a). After the sustainable development indicators and documents and data are matched one by one, they are uploaded to the local database, see Figure 4 (b). After completing the above operations, the managers or evaluators of the sustainable development of the construction project can briefly obtain this information through the local LLM, as shown in Figure 4 (c). Similar to the pre-trained large language model connected to the normal network, the LLM linked to the local database can interact. At the same time, if the pre-designed prompt words are used, the model will obtain the text information of the construction project through the local database and output the content related to the sustainable development of the construction project. At the same time, the output of information sources can also be realized.

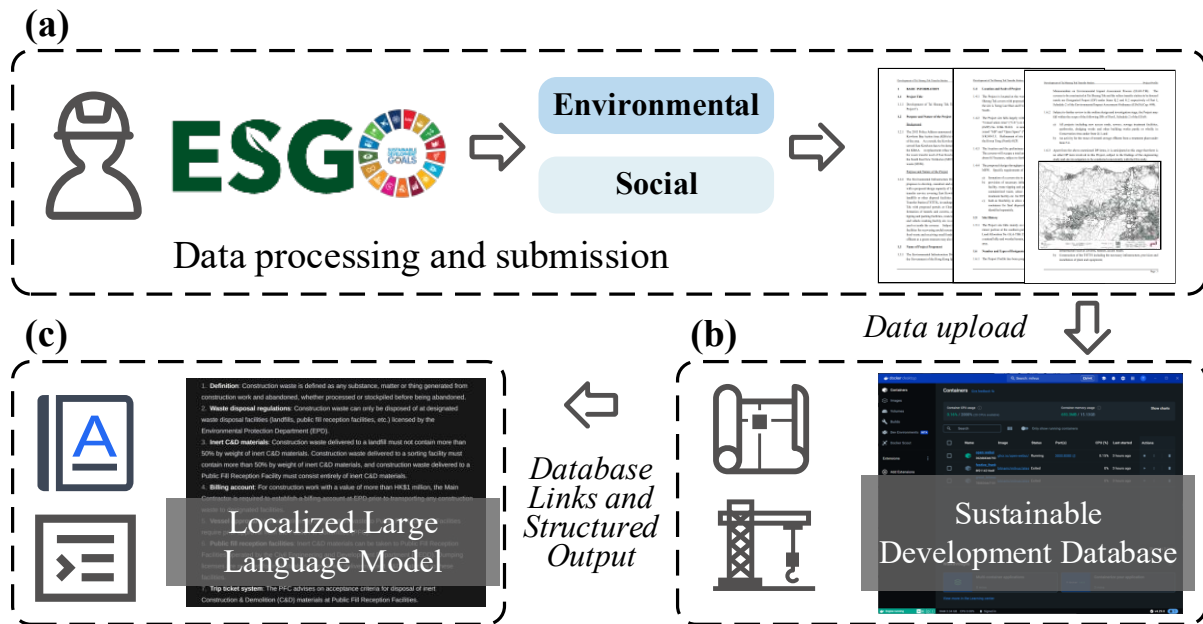


Figure 4. Verification scenarios: (a) Sustainable data collation of construction projects; (b) Data upload; (c) Data acquisition based on LLM.

Through the above validation steps, the framework has been proven to be able to process and manage sustainable development related content in construction projects. For some sustainable development indicators, LLM can also organize and classify information based on the local construction database. For example, when asking about waste-related content in construction projects, the output content will include construction events of different types of waste, such as C&D, solid waste, and building waste. Through pre-designed prompt words, the model can also locate and output the document sources related to the indicators in the database. The efficiency of each output of the framework is also counted, as shown in Figure 5. The output efficiency is maintained within an acceptable range.

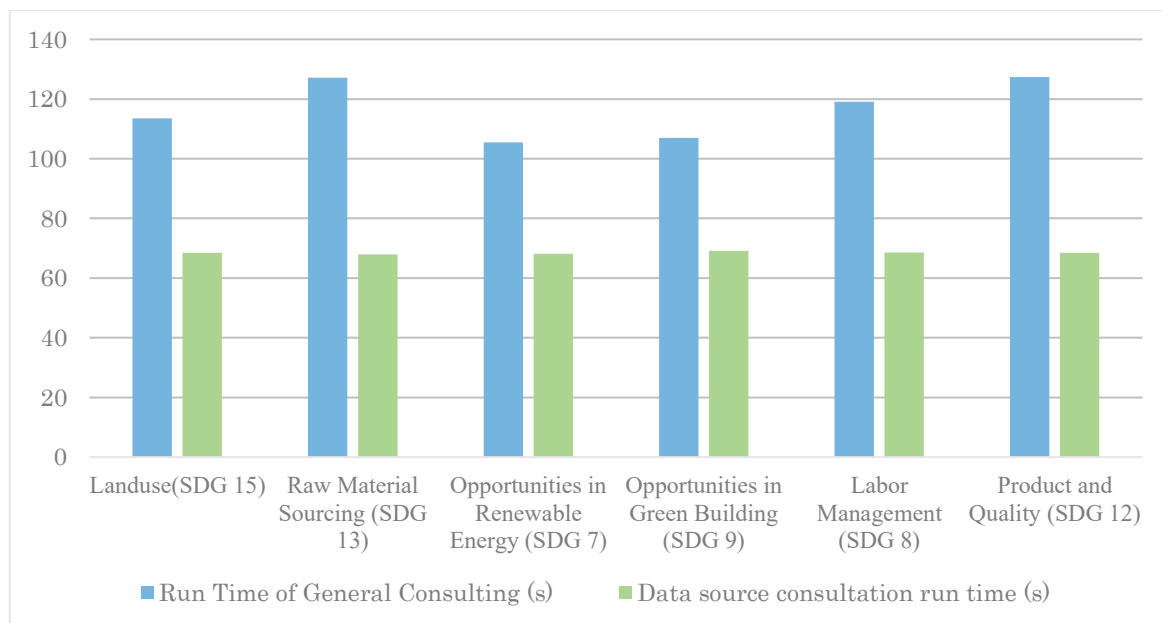


Figure 5. Average Runtime of Framework Validation

The framework verification step was run on Windows 11 23H2, with i7-12700K, RTX 4060Ti and 32GB memory. Each call for sustainable development indicators in the construction project took about two minutes to respond. The call for confirmation of the data source took about one minute. In terms of the accuracy of the

response content, the accuracy of TP/(TP+FP) 88.3%. The accuracy rate comes from human expert verification of the answers output by the LLM. Error outputs from the model are annotated, which often come from similar keywords in random construction project documents, but these documents themselves are not directly related to the SDGs or ESG indicators being consulted.

4. DISCUSSION

In recent years, various industries have realized the importance of sustainable development on their impact, including the construction industry. This shift comes from the attention of investors, regulators and governments to related matters. The Sustainable Development Goals provide a new perspective for the construction industry. From this perspective, the global vision for sustainable development can be transformed into business solutions. In building management and operations, construction companies should emphasize collaborative ESG activities to gain a good public business reputation, such as reducing carbon dioxide emissions and construction waste and improving safety and health (Jain et al., 2020). The refinement of SDG goals and ESG assessment frameworks can help the construction industry deal with these issues in a more accurate way. At the same time, accurate management of corporate development also requires the promotion of technological development to reduce management costs and improve efficiency. LLM can help managers achieve this goal in the face of large and complex construction data.

The sustainable development data management framework for construction projects developed in this paper based on the above background can achieve rapid processing of complex text data to a certain extent. This relies on the ability of the pre-trained large language model to understand and deconstruct the text. At the same time, it provides a mitigation method for the situation that sustainable development management relies on a large number of professional people to carry out work. It effectively reduces the management cost of relevant data in construction projects. At the same time, the localized database and locally deployed LLM can avoid the risk of commercial data leakage. It provides support for the secure data management of construction projects.

However, the existing framework has only been validated in designed scenarios. For more complex and multi-party actual engineering projects and business operations, the method is not capable enough. It is necessary to develop a specialized training LLM model for the construction industry. This will not only enhance the efficiency of data exchange in the construction industry, but also help the industry achieve sustainable development goals. The accuracy tested in this paper is obtained by querying each SDGs-related content of the project construction itself through LLM, and then through manual expert review. It is somewhat subjective. The existing LLM is pre-trained based on a large amount of text and data, and sometimes cannot meet professional requirements in answering professional questions (Liu et al., 2024). For construction activities, environmental management such as SDGs and ESG, specific training of the LLMs used may be a suitable solution.

At the same time, for actual engineering projects, multimodal data is related to sustainable development assessments. For example, building information modeling (BIM) data, engineering drawing data, and even sensor data from construction sites and building operations are related to comprehensive sustainable development assessments of projects. The completer and more comprehensive the data, the more realistic it can show the situation and expected development of the construction project. This is helpful for dynamic management of sustainable development. Current methods often target a single data source. Further management models, such as the development of management frameworks and methods based on digital twins and machine learning, are necessary.

5. CONCLUSIONS

The SDGs proposed in the United Nations 2030 Agenda for Sustainable Development and the ESG assessment methods widely used in the market have put forward new goals and directions for environmental protection, social support and economic development of enterprises. The construction industry is one of the key industries for sustainable development because of its own pollution and high social relevance. To solve the difficulties caused by the large amount of complex data in construction projects for sustainable development management, this paper proposes a sustainable development data management framework for construction projects using LLM. It achieves three goals. First, this paper identified sustainable development goals for construction projects. This is conducive to accurately grasping the project's development focus on environmental protection and social development for the construction industry. At the same time, it mapped the relationship between sustainable development indicators and data sources in the project to help managers more quickly identify the relationship between different indicators and construction events. Finally, a data management framework based on LLM was developed. After verification, the developed framework can realize the extraction and organization of sustainable development information of construction projects based on local databases. It provides a feasible technical solution for rapid data processing and efficient management of sustainable development of construction projects.

However, the current methods still have limitations. First, for the existing data management model, the complex format in the construction project is not fully considered. In addition to text, structured data such as

building information models, image data, etc. are also related to sustainable development assessment. How this multimodal data can be linked and interacted with LLM is still difficult. In addition, the current LLM model is usually universal across the industry, and it is often difficult to meet the specialized needs of the construction industry. Therefore, for sustainable development-related matters, the development and adjustment of LLM is in demand for the construction industry. In the future, some specific optimization methods such as vector database design and specialized report output will be studied.

REFERENCES

- Asyrofi, R., Dewi, M.R., Lutfhi, M.I., Wibowo, P., 2023. Systematic Literature Review Langchain Proposed, in: 2023 International Electronics Symposium (IES). Presented at the 2023 International Electronics Symposium (IES), pp. 533–537. <https://doi.org/10.1109/IES59143.2023.10242497>
- Banihashemi, S., Hosseini, M.R., Golizadeh, H., Sankaran, S., 2017. Critical success factors (CSFs) for integration of sustainability into construction project management practices in developing countries. *International Journal of Project Management* 35, 1103–1119. <https://doi.org/10.1016/j.ijproman.2017.01.014>
- Chaitanya, K., Aromar, R., Jessica, E., Holger, K., 2016. Getting Started with the SDGs in Cities | Environment & Urbanization. Sustainable Development Solutions Network 110.
- De Franco, C., 2020. ESG Controversies and Their Impact on Performance. *JOI* 29, 33–45. <https://doi.org/10.3905/joi.2019.1.106>
- Fei, W., Opoku, A., Agyekum, K., Oppon, J.A., Ahmed, V., Chen, C., Lok, K.L., 2021. The Critical Role of the Construction Industry in Achieving the Sustainable Development Goals (SDGs): Delivering Projects for the Common Good. *Sustainability* 13, 9112. <https://doi.org/10.3390/su13169112>
- Giudici, M., Abbo, G.A., Belotti, O., Braccini, A., Dubini, F., Izzo, R.A., Crovari, P., Garzotto, F., 2023. Assessing LLMs Responses in the Field of Domestic Sustainability: An Exploratory Study, in: 2023 Third International Conference on Digital Data Processing (DDP). Presented at the 2023 Third International Conference on Digital Data Processing (DDP), pp. 42–48. <https://doi.org/10.1109/DDP60485.2023.00019>
- Gong, X., Tao, X., Zhang, M., Xu, Y., Kwok, H.H.L., Dai, J., Cheng, J.C.P., 2024. Secure Environmental, Social, and Governance (ESG) Data Management for Construction Projects Using Blockchain. *Sustainable Cities and Society* 105582. <https://doi.org/10.1016/j.scs.2024.105582>
- Jain, S., Singhal, S., Jain, N.K., Bhaskar, K., 2020. Construction and demolition waste recycling: Investigating the role of theory of planned behavior, institutional pressures and environmental consciousness. *Journal of Cleaner Production* 263, 121405. <https://doi.org/10.1016/j.jclepro.2020.121405>
- Johnsson, F., Karlsson, I., Rootzén, J., Ahlbäck, A., Gustavsson, M., 2020. The framing of a sustainable development goals assessment in decarbonizing the construction industry – Avoiding “Greenwashing.” *Renewable and Sustainable Energy Reviews* 131, 110029. <https://doi.org/10.1016/j.rser.2020.110029>
- Litvinenko, V., Bowbrick, I., Naumov, I., Zaitseva, Z., 2022. Global guidelines and requirements for professional competencies of natural resource extraction engineers: Implications for ESG principles and sustainable development goals. *Journal of Cleaner Production* 338, 130530. <https://doi.org/10.1016/j.jclepro.2022.130530>
- Liu, Z., Zhong, A., Li, Y., Yang, L., Ju, C., Wu, Z., Ma, C., Shu, P., Chen, C., Kim, S., Dai, H., Zhao, L., Zhu, D., Liu, J., Liu, W., Shen, D., Li, Q., Liu, T., Li, X., 2024. Tailoring Large Language Models to Radiology: A Preliminary Approach to LLM Adaptation for a Highly Specialized Domain, in: Cao, X., Xu, X., Reikik, I., Cui, Z., Ouyang, X. (Eds.), *Machine Learning in Medical Imaging*. Springer Nature Switzerland, Cham, pp. 464–473. https://doi.org/10.1007/978-3-031-45673-2_46
- Lou, J., Lu, W., Xue, F., 2021. A Review of BIM Data Exchange Method in BIM Collaboration, in: Lu, X., Zhang, Z., Lu, W., Peng, Y. (Eds.), *Proceedings of the 25th International Symposium on Advancement of Construction Management and Real Estate*. Springer, Singapore, pp. 1329–1338. https://doi.org/10.1007/978-981-16-3587-8_90
- Mishra, S., 2023. ESG Impact Type Classification: Leveraging Strategic Prompt Engineering and LLM Fine-Tuning, in: Chen, C.-C., Huang, H.-H., Takamura, H., Chen, H.-H., Sakaji, H., Izumi, K. (Eds.), *Proceedings of the Sixth Workshop on Financial Technology and Natural Language Processing*. Association for Computational Linguistics, Bali, Indonesia, pp. 72–78. <https://doi.org/10.18653/v1/2023.finnlp-2.11>
- Pradhan P., Costa L., Rybski D., Lucht W., Kropp J.P., n.d. A Systematic Study of Sustainable Development Goal (SDG) Interactions. <https://doi.org/10.1002/2017EF000632>
- Pu, H., Yang, X., Li, J., Guo, R., 2024. AutoRepo: A general framework for multimodal LLM-based automated construction reporting. *Expert Systems with Applications* 255, 124601. <https://doi.org/10.1016/j.eswa.2024.124601>
- Xu, Y., Tao, X., Das, M., Kwok, H.H.L., Liu, H., Kuan, K.K.L., Lau, A.K.H., Cheng, J.C.P., 2024. A blockchain-

based framework for carbon management towards construction material and product certification. Advanced Engineering Informatics 61, 102242. <https://doi.org/10.1016/j.aei.2023.102242>

Yan, H., Yang, N., Peng, Y., Ren, Y., 2020. Data mining in the construction industry: Present status, opportunities, and future trends. Automation in Construction 119, 103331. <https://doi.org/10.1016/j.autcon.2020.103331>

Automatic Compliance Checking of BIM Models for Architecture and Fire Protection Based on Knowledge Graphs and Machine Learning

Sihao Li¹, Guangyao Chen², Yangze Liang³ and Zhao Xu⁴

1) Ph.D. Candidate, Department of Construction and Real Estate, Southeast University, China. Email: lish@seu.edu.cn

2) Ph.D. Candidate, Department of Construction and Real Estate, Southeast University, China. Email: chengy@seu.edu.cn

3) Ph.D. Candidate, Department of Construction and Real Estate, Southeast University, China. Email: lyangz@seu.edu.cn

3) D. Eng., Prof., Department of Construction and Real Estate, Southeast University, China. Email: xuzhao@seu.edu.cn

Abstract: Compliance checking of Building Information Modeling (BIM) models is a critical process throughout the construction lifecycle, particularly during the design phase. Building design often involves the integration of multiple disciplines and complex spatial relationships, leading to errors. The growing volume and complexity of information embedded in BIM models have further complicated compliance checking. Traditional manual methods are not only time-consuming but also prone to mistakes. To address these challenges, this study proposes an integrated conceptual framework for automated BIM compliance checking, leveraging knowledge graph (KG) and machine learning. The framework aims to convert unstructured clauses in Chinese building standards into structured, interpretable, and extractable data, enabling the automatic detection of design errors in BIM models. The framework incorporates several key components. First, it constructs a knowledge graph by developing ontologies for Chinese building standards and training semantic role annotation models. A data extraction pipeline is designed using the Dynamo module in Revit to retrieve relevant information from BIM models. Finally, compliance checking logic is defined using Java to establish rules for matching the extracted building standard knowledge with BIM model information. The feasibility of this automated compliance-checking framework was validated using BIM models from two real-world projects, demonstrating its potential to streamline the compliance process and reduce errors in building design.

Keywords: BIM; machine learning; knowledge graph; compliance checking; natural language processing

1. INTRODUCTION

The design review process in building construction has evolved from hand-drawn blueprints to CAD (Balachandran et al., 1991) and BIM (Liu et al., 2022; Solihin et al., 2020), which is now integral to project approvals in many cities. In China, the rapid growth of BIM data supports digital advancements and smart city initiatives (Dimyadi and Amor, 2013). BIM compliance checks ensure safety, regulatory adherence, and public interest compliance but face challenges due to the complexity of BIM data. Manual reviews are inefficient, require high expertise, and can introduce biases (Li et al., 2024).

Automated Compliance Checking (ACC) offers a solution to these challenges (Noardo et al., 2022). Researchers have focused on standardizing regulations (Ismail et al., 2017) and integrating model data (Zhang and El-Gohary, 2017), leading to tools like Solibri Model Checker (Solihin et al., 2020) and Autodesk Navisworks (Hjelseth, 2015). These systems handle tasks like clash detection and stairway inspections but struggle with complex logic and interrelated data. Enriching model information remains a challenge, and current systems rely on hard-coded rules, requiring extensive programming and frequent updates to match evolving standards (Zhang and El-Gohary, 2017). While semantic inspection methods offer flexibility (Beach et al., 2015; Liebich et al., 2004), they fall short in automating rule construction and updates.

Knowledge graphs (KG) (Singhal, 2012), first introduced by Google in 2012, excel in semantic reasoning and flexible knowledge representation (Li et al., 2021). KGs can structure and formalize building standards, transforming unstructured clauses into computable data for machine interpretation. Soft-coding techniques and Information Extraction (IE) enable automated rule library creation, empowering domain experts to define rules without programming, improving BIM compliance efficiency.

However, existing KGs have limited scope, poor transferability, and focus primarily on representation rather than reasoning. Addressing these gaps, this study proposes an ACC framework combining ontology, NLP, and deep learning to automate KG construction for Chinese building standards. Logical relationships within KGs are used to interpret standards, while processes in Dynamo and Java establish links between BIM models and standards, automating compliance checks. An intelligent platform further enhances BIM review efficiency and accuracy.

2. METHOD

This study proposes an integrated conceptual framework for building engineering that combines NLP, machine learning, and knowledge graph techniques to automate the compliance checking of BIM models. The aim is to improve the efficiency and accuracy of this process through automation. The methods used in this research are designed to be easily updated and transferable, providing a pathway for the future development of knowledge

graphs across various building engineering disciplines. An overview of the proposed methodology is shown in Fig. 1, along with explanatory images for key concepts and steps. The process consists of three interrelated parts: knowledge graph construction, BIM model information extraction, and matching BIM models to the knowledge graph for compliance checking.

The first step is constructing the building standard knowledge graph, which begins with analyzing standards and studying relevant texts in the field. Based on this analysis, an ontology of architectural and fire protection elements is created. The building standard documents are preprocessed, with semantic role labels applied to them. A corpus of building standards is then built and annotated with these labels. A deep learning model, BERT-BiLSTM-CRF, is trained to automatically label the building standard texts. Syntactic analysis is performed using CFG, and knowledge graph triples are extracted from the annotated texts. Finally, Neo4j is used to construct and store the knowledge graph, where nodes and attributes are designed to preserve the logical relationships from the building standards, enhancing the graph's reasoning capabilities.

Next, BIM model checking data is predefined, and a data extraction method based on Dynamo is created to obtain the necessary files for checking. A matching algorithm is then developed to represent the logical rules of the knowledge graph triples, leading to an automated compliance checking method for BIM models. An intelligent BIM model checking platform is built and validated through two engineering case studies. The platform successfully performs automated compliance checking for both architectural and fire protection aspects of BIM models, generating comprehensive checking reports. Since the building standards used in this study are based on Chinese regulations, the research focuses on Chinese texts, with comparative English translations provided for clarity.

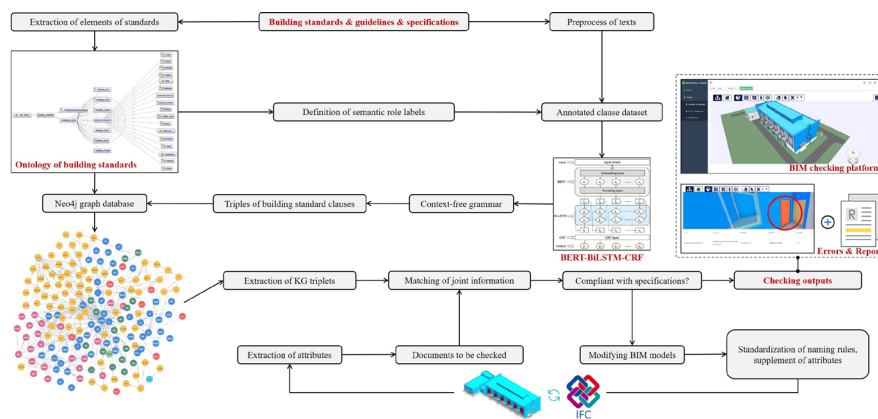


Figure 1. Overview of the proposed methodology framework

Building design and construction standards form the foundation for compliance checks. This study selects several widely used standards in architecture and fire protection (in Chinese) with broad coverage, comprehensive content, and extensive applicability. These standards include the "Unified Standard for Civil Building Design (GB 50352-2019)", "Residential Building Code (GB 50368-2005)", "Design Code for Residential Buildings (GB 50096-2011)", "Fire Protection Design Code for Buildings (GB 50016-2014)", "Technical Standard for Fire Emergency Lighting and Evacuation Indication Systems (GB 51309-2018)", "Fire Protection Design Code for Automobile Garages, Repair Garages, and Parking Lots (GB 50067-2014)", and "Fire Protection Design Code for Interior Decoration of Buildings (GB 50222-2017)". These standards serve as the data foundation for constructing the ontology of building standards, which is then used to build the knowledge graph (KG) through information extraction and syntactic analysis.

The process of constructing the KG in this study follows these steps: First, the ontology of building standards is developed. Then, the building standard texts undergo preprocessing. Based on this ontology, custom semantic role labels are designed, and the preprocessed texts are annotated with these labels to create a labeled dataset. A semantic role labeling model is trained using this dataset to automatically label the building standard texts. Afterward, syntactic analysis is performed on the annotated texts, which are parsed to extract data triples for the knowledge graph. Finally, the knowledge graph is constructed by combining the pattern layer and data layer.

2.1 Ontology Construction

The clauses in building standards define specific constraints for engineering design and construction and serve as the foundational knowledge for building the ontology and knowledge graph in this study. These clauses are categorized into mandatory and non-mandatory types. Mandatory clauses are those that must be followed, while non-mandatory clauses represent guidelines that should generally be adhered to under typical circumstances. Mandatory clauses take precedence over non-mandatory ones. In this study, a hierarchical analysis of the building standard knowledge and the logical relationships between various construction elements was conducted. A seven-

step method was employed to construct the building standard ontology, which forms the basis for semantic role labeling and knowledge graph development.

The focus of this research is on the compliance check of BIM models, with the ontology's scope defined as architecture and fire control. By integrating concepts, terminology, and attributes from existing ontologies such as Ontolingua and DAML, the construction of the building standard ontology in this study is enriched to improve its efficiency and completeness. The key terms included in the selected criteria are counted to provide a list of all terms.

The ontology development in this study follows a top-down approach. Initially, key concepts are extracted from the selected standard texts to form the classes within the ontology. Next, a hierarchical class structure is created through a stepwise subdivision from top to bottom. The building code is designated as the parent class, which is divided into two main categories: professional standards and building objects. These categories are then further subdivided into subclasses, each containing various underlying objects. Some of these underlying objects can be further classified into different types of instances, while others are directly treated as instances that do not require further subdivision. The professional standards category includes building standards and fire control standards, while the building objects category is divided into several subclasses such as project, building, space, site, storey, and component. A partial view of the constructed ontology for building standard texts is shown in Fig. 2.

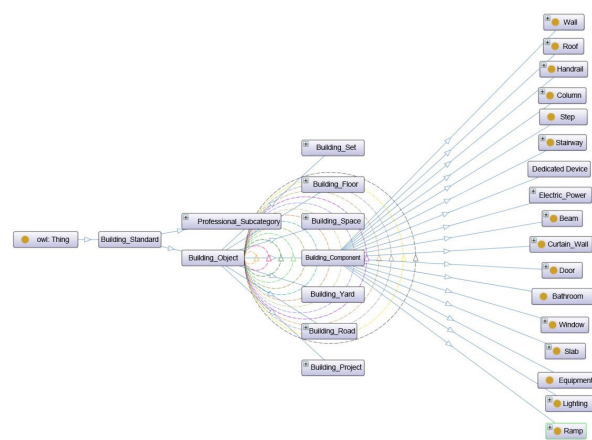


Figure 2. Ontology classes for building standards

2.2 Recognition of Semantic Role Labels

Based on the constructed ontology and Chinese word segmentation method, 6 semantic role labels have been customized, as shown in Table 1.

Table 1. Labels of semantic roles

Abbreviation	Definition	Interpretation
JD	Building object	Including building project, building site, building, building floor, building component, building space, building road, etc., referring to the semantic entities in the triples.
DS	Object property	Elements lower than JD, which are object properties of JD, including containment relationships and spatial relationships, referring to the semantic relationships between entities in the triples.
SS	Data property	Elements lower than JD, which are data properties of JD, including numerical properties, non-numerical properties, and measures, referring to the semantic properties in the triples.
SXZ	Property value	Requirements and conditions connected to property SS, with data formats such as Float, String, Boolean, referring to the semantic relationships between entities and properties in the triples.
BJ	Comparative term	Comparative relationship between property SS and requirement SXZ, including "greater than", "equal to", "less than", "should not be less than", and other comparative terms, referring to the semantic relationships between properties and property values in the triples.

LJ	Logical term	Expressing logical semantic relationships such as "parallel", "optional" including "and", "both", "except for", etc.
Abbreviation	Definition	Interpretation
JD	Building object	Including building project, building site, building, building floor, building component, building space, building road, etc., referring to the semantic entities in the triples.

Selected building code provisions are annotated using the BIO (Begin-inside-outside) sequence labeling method, where "B" indicates the beginning, "I" represents the middle position, and "O" denotes that the token does not belong to any label type. For example, the sentence "Residential buildings with a height greater than 33m should be equipped with fire elevators" is annotated, as shown in Figure 3. In this study, a total of 11,174 lines of data were annotated, and the deep learning model's training, development, and test datasets were constructed with an 8:1:1 ratio.

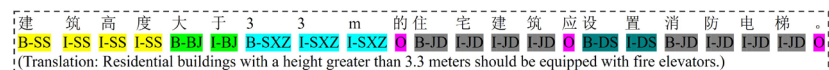


Figure 3. BIO sequence labeling data

After constructing the dataset, the BERT-BiLSTM-CRF model is introduced to automatically label the semantic roles of building standard texts. The model architecture consists of three levels. First, the pre-trained BERT layer is used to generate embedding vectors and output encoded sequence vectors $T=(t_1, t_2, \dots, t_n)$. Then, the Bi-LSTM layer extracts semantic features from the text data by integrating contextual information, and the resulting bidirectional output sequence vector h_i indicates the probability of each position being recognized with a BIO label. Finally, the CRF layer adjusts and optimizes the sequence vectors, outputting the most probable label sequence $Y=(y_1, y_2, \dots, y_n)$.

2.3 Extraction of the Triples

In the triples, entities and attributes correspond to JD, DS, and SS labels, which can be achieved by extracting entities under each label based on the output of the statistical model. Additionally, Context-Free Grammar (CFG) is used to perform syntactic analysis on the labeled results, helping to identify and refine the extraction of more complex relationships in the triples.

CFG is a formal language grammar that describes all possible string combinations generated by a probabilistic process, with sentence generation following predefined rules. By analyzing common semantic representations in building standard texts, four types of rules are classified after CFG parsing and adjustment:

<JD-SS-BJ-SXZ>: Constraints on certain numerical attributes of building objects. For example, "The fire resistance grade of a high-rise factory should not be lower than Grade II."

<JD-DS-SXZ>: Constraints on measures or non-numeric attributes of data. For example, "The floor of the toilet should adopt waterproof construction measures."

<JD-DS-JD>: Constraints on the existence of certain building objects, equipment, or systems, or constraints on category matching for building objects. For example, "Residential buildings should have a lighting power supply system."

<JD-JD-DS>: Constraints on the spatial relationship between certain building objects. For example, "Electrical wiring pipes should not be set next to the toilet."

Taking the first rule as an example, the automatic semantic labeling result is shown in Figure 4. By directly applying CFG, the syntactic tree is generated. According to the part-of-speech labeling rules, "high-rise factory" is a noun phrase (NP) terminal, where "high-rise factory" is the entity to be extracted. Both "high-rise factory" and "fire resistance grade" are labeled as nouns (N), making it difficult to distinguish between the entity and attribute. At the same time, "should not be lower than" is split into "should not" and "be lower than," corresponding to the auxiliary verb (Aux) and verb (V). For the knowledge graph triple, "should not be lower than" should be extracted as a whole, representing the relationship between the attribute and its value.

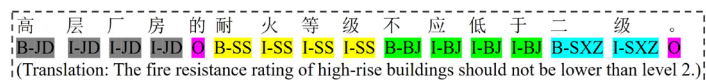


Figure 4. Automatic annotation results of semantic labels

Therefore, the triples cannot be directly obtained from the syntactic tree at this point. By combining the semantic role labels with the syntactic tree, the semantic labels were changed. The concept of JD corresponds to the entity in the triple, SS corresponds to the data attribute, and SXZ corresponds to the attribute value connected to the SS attribute. The knowledge graph triple for this rule and its minimal unit triple are shown in Figure 5. Using

a similar approach, adjustments are made to the labeled results, and the triples for the standard text are obtained.

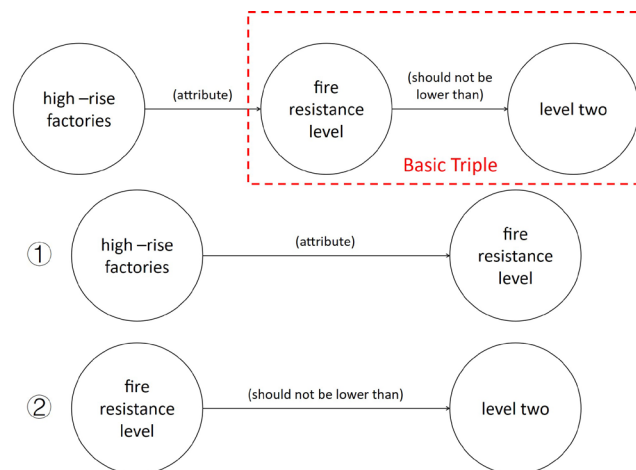


Figure 5. Knowledge graph triplet

2.4 Generation of Knowledge Graph

After obtaining the triples, the knowledge graph is generated and stored using the Neo4j database. The ontology triples constructed in Protégé software are parsed and converted into ".turtle files," which are then imported into Neo4j to create the knowledge graph network. To preserve logical relationships and meet the requirements for subsequent automatic review of BIM models, the knowledge triples of the clauses are organized. Additional relationships are created, such as (standard) -[clause]-> (clause number), (clause number) -[clause type]-> (mandatory/non-mandatory), and (clause number) -[clause constraint]-> (building object). Entity nodes corresponding to each clause are created individually, as shown in Figure 6.

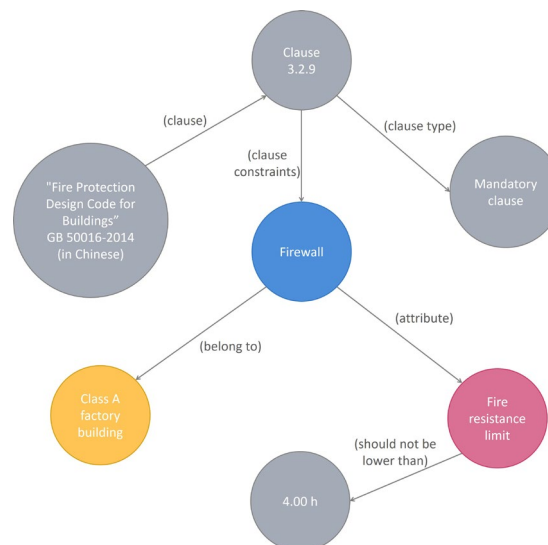


Figure 6. Example of a relational triplet

2.5 Generation of Knowledge Graph

In this study, the process of automatically checking BIM models involves data extraction and matching. Dynamo is used to extract specific data from BIM models, covering tasks such as adding, extracting, and completing model data, as shown in Fig. 7 and Fig. 8. Afterward, the extracted data is matched with the knowledge stored in the knowledge graph (KG) using triples, enabling the compliance check of the model. Although Dynamo's visual nodes can extract certain parameters from model elements, their capabilities are limited, as some parameters cannot be directly accessed using the built-in nodes. To overcome this, this study integrates Dynamo's visual nodes with PythonScript to call the Revit API for data extraction and compliance checking.

Some parameters necessary for compliance checks are not included in the default element parameters. Therefore, it is essential to define and add custom attribute parameters based on the specific checking requirements to ensure comprehensive model information extraction. In Revit, project parameters are a useful tool, acting as containers for element data. These parameters can be customized to include various attributes, such as fire resistance level, combustibility, and mitigation measures. Additionally, project parameters allow users to flexibly

define element attributes based on the specific requirements of the compliance checks. Revit elements are primarily categorized into Categories, Families, Family Types, and Family Instances. Family parameters are divided into type parameters and instance parameters. Type parameters are shared across all instances of a family type, meaning that modifying them will alter all instance parameters within that family. Conversely, modifying the instance parameters of a specific family instance will only affect that instance. In this study, custom parameters are added as instance parameters to facilitate the assignment and extraction of parameters for various components. The batch addition of these parameters follows a programming workflow in Dynamo. When creating a new family instance, the instance parameters will automatically include the custom parameters, allowing the required data to be input into the model.

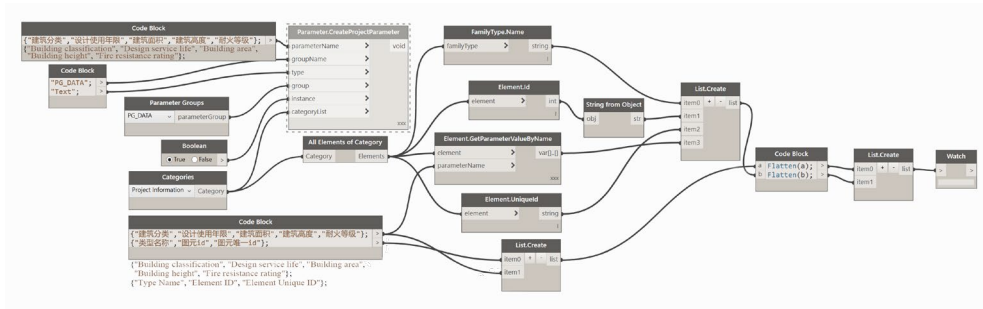


Figure 7. Extraction of project information parameters

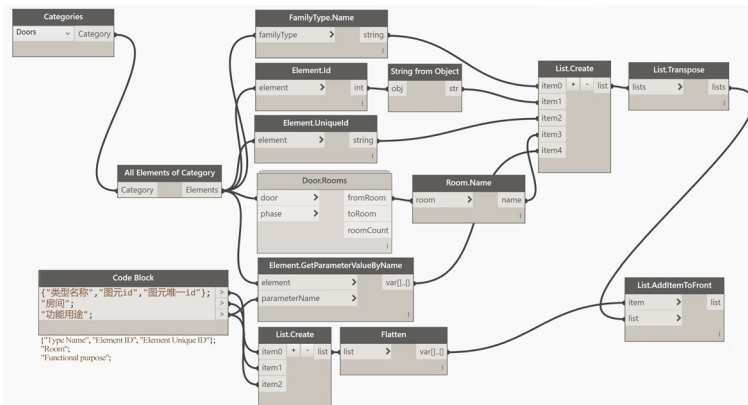


Figure 8. Information extraction of rooms associated with doors and windows

3. RESULTS

The experimental environment for the BERT-BiLSTM-CRF model in Section 2.2 for automatic semantic role labeling consists of a DELL R730 server, 64-bit Linux system, Visual Studio Code (Version 1.71), TensorFlow 1.13.2 framework, and an NVIDIA T4 GPU. The performance evaluation criteria of the model are accuracy, precision, recall and F1 according to the confusion matrix, and the calculation formulas are shown in Formula (1), (2) and (3). The performance of the semantic role annotation model for extracting various semantic labels is shown in Table 2.

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$F1 = 2 \times \frac{precision * recall}{precision + recall} \quad (3)$$

Table 2. FID score statistics for different models

Label	Precision (%)	Recall (%)	F1
BJ	100.00	100.00	1.000
DS	77.70	87.50	0.824
JD	96.55	95.45	0.960

LJ	100.00	95.24	0.976
SS	95.12	97.50	0.963
SXZ	95.65	95.65	0.957
Label	precision (%)	recall (%)	F1
BJ	100.00	100.00	1.000

A web platform integrating the knowledge graph (KG) and the described algorithms has been developed to facilitate BIM model compliance checking. The platform adopts a Browser/Server (B/S) architecture and consists of three layers, including Data Access Layer (DAL), Business Logic Layer (BLL), and User Interface (UI) Layer. The DAL handles storage for standard ontologies, the KG, and BIM model databases, enabling access through web service interfaces, drivers, and open standards. The BLL incorporates both business functions and logical entities. Business functions include view, system, model, and project management, while logical entities encompass web servers, database servers, engines, and scripts. The platform utilizes HTML5 and WebGL for its base structure, with functionality implemented through CSS and JavaScript (JS) scripts. Users interact with BIM data on the visualization platform via the UI, performing operations such as adding, deleting, editing, querying, and generating compliance reports.

A factory engineering project was used to verify the feasibility of this method, as shown in Fig. 9. For the building project, elements such as "steps," "stairs," "ramps," "railings," "doors," "firewalls," "roof access points," and "rooms" were selected for automated architectural and fire safety checks. The relevant data from the models was extracted and then analyzed and assessed using matching algorithms.



Figure 9. Compliance check platform for BIM models

Finally, an automated compliance report is generated for each BIM model. Non-compliant components are identified in the visualized model by their unique IDs and highlighted in different colors. For instance, as shown in Fig. 10, a door in the industrial building is flagged as non-compliant. According to Article 6.2.7 of the "Fire Protection Code for Building Design" (Chinese), doors leading into electrical rooms must be Class A fire-resistant doors. In this model, the door opens into the building's interior in the electrical room, but its "functional" value is listed as "Class B fire-resistant door," which does not meet the standard.

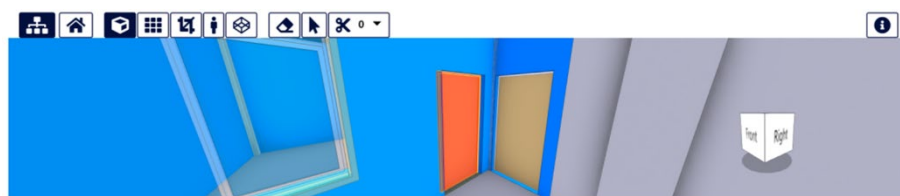


Figure 10. Highlighted door in the industry building

4. CONCLUSIONS

Based on the analysis of standards and BIM model data, an integrated framework for automated compliance checking of BIM models utilizing NLP and knowledge graphs has been developed. The feasibility and effectiveness of this approach were validated through two case studies. The theoretical value of the research lies in presenting a method for automating BIM model compliance checking and enhancing the automation and transferability of domain-specific knowledge graph construction. Its practical significance is reflected in the development and successful application of a BIM model checking system, offering both a theoretical foundation and practical reference for related research and applications. The main research contributions are as follows:

(1) Developed an ontology for the building standards domain using the "seven-step method" and created a semantic role labeling model for building standard texts based on BERT-BiLSTM-CRF.

(2) Integrated the semantic role labeling model with CFG-based analysis methods and constructed a knowledge graph stored in the Neo4j database, incorporating logical markers to strengthen the graph's reasoning capabilities.

(3) Developed script tools using Dynamo and the Revit API to enable parameter addition, extraction, and supplementation for BIM model data.

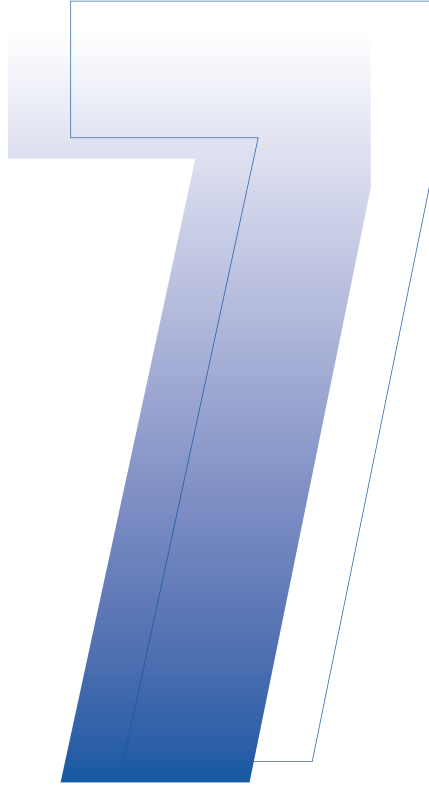
(4) Formulated logical rule expressions for information matching and proposed a method to match knowledge graph triples with BIM model data, enabling automated compliance checking.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support provided by the "National Natural Science Foundation of China (72071043)," the "National Key Research and Development Program of China (2022YFC3803600)," and the "SEU Innovation Capability Enhancement Plan for Doctoral Students (CXJH_SEU 25089)." Additionally, the authors thank the author of the template for the manuscripts of the proceedings of the 30th International Symposium on Automation and Robotics in Construction and Mining (ISARC 2013).

REFERENCES

- Balachandran, M., Rosenman, M. A., and Gero, J. S. (1991). A knowledge-based approach to the automatic verification of designs from CAD databases. In J. S. Gero (Ed.), *Artificial Intelligence in Design '91* (pp. 757-781). Butterworth-Heinemann.
- Beach, T. H., Rezgui, Y., Li, H., et al. (2015). A rule-based semantic approach for automated regulatory compliance in the construction sector. *Expert Systems with Applications*, 42(12), 5219-5231.
- Dimyadi, J., and Amor, R. (2013). Automated building code compliance checking: Where is it at? *Proceedings of CIB WBC 2013*, 172-185.
- Hjelseth, E. (2015). BIM-based model checking (BMC). *Building Information Modeling—Applications and Practices*, 33-61.
- Ismail, A. S., Ali, K. N., & Iahad, N. A. (2017). A review on BIM-based automated code compliance checking system. *2017 international conference on research and innovation in information systems (icriis)* (pp. 1-6). IEEE.
- Li, S., Wang, J., and Xu, Z. (2024). Automated compliance checking for BIM models based on Chinese-NLP and knowledge graph: An integrative conceptual framework. *Engineering, Construction and Architectural Management*.
- Liebich, T., Wix, J., and Qi, Z. (2004). Speeding-up the submission process: The Singapore e-plan checking project offers automatic plan checking based on IFC. *International Conference on Construction Information Technology (INCITE 2004)*, 245-252.
- Liu, H., Cheng, J. C. P., Gan, V. J. L., et al. (2022). A novel data-driven framework based on BIM and knowledge graph for automatic model auditing and quantity take-off. *Advanced Engineering Informatics*, 54, 101757.
- Li, X., Lyu, M., Wang, Z., et al. (2021). Exploiting knowledge graphs in industrial products and services: A survey of key aspects, challenges, and future perspectives. *Computers in Industry*, 129, 103449.
- Noardo, F., Wu, T., Arroyo Ohori, K., et al. (2022). IFC models for semi-automating common planning checks for building permits. *Automation in Construction*, 134, 104097.
- Singhal, A. (2012). Introducing the knowledge graph: things, not strings. *Official google blog*, 5(16), 3.
- Solihin, W., Dimyadi, J., Lee, Y.-C., et al. (2020). Simplified schema queries for supporting BIM-based rule-checking applications. *Automation in Construction*, 117, 103248.
- Zhang, J., and El-Gohary, N. M. (2017). Integrating semantic NLP and logic reasoning into a unified system for fully-automated code checking. *Automation in Construction*, 73, 45-57.



Part 7

Sustainable Construction and Development

DEVELOPING GNN-BASED SURROGATE MODELS FOR MULTI-OBJECTIVE SUSTAINABLE PERFORMANCE PREDICTIONS OF RESIDENTIAL BLOCKS

Zhaoji Wu ¹, Wenli Liu ², Jack C.P. Cheng ³, Zhe Wang ⁴, Helen H.L. Kwok ⁵, Cong Huang ⁶, and Fangli Hou ⁷

1) Ph.D. Candidate, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: zwubz@connect.ust.hk

2) Ph.D. Student, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: wliucd@connect.ust.hk

3) Ph.D., Prof., Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong SAR. Email: cejcheng@ust.hk

4) Ph.D., Asst. Prof., Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong SAR. Email: cezhewang@ust.hk

5) Ph.D., Postdoc, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong SAR. Email: hlkwokab@connect.ust.hk

6) Ph.D. Student, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: chuangax@connect.ust.hk

7) Ph.D., Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: fhouaa@connect.ust.hk

Abstract: Although building performance simulation using physical models is frequently utilized for performance prediction, its significant computational demands pose challenges to its implementation in the early design stage. Surrogate models have been proposed to replicate computationally expensive physics-based simulation models, but existing surrogate models for sustainable residential block design are limited in scope, focusing on specific cases. Graph neural network (GNN) could be a solution to enhance the generality of the surrogate models for residential block design. However, the optimal architectures of the surrogate model and the time costs compared with physics-based simulation models have not been discussed yet. To fill these gaps, this study explores the development of GNN-based surrogate models for multi-objective sustainable performance predictions of residential blocks. Firstly, we introduce a graph schema to represent the general geometric features and relations, and a regional dataset for training and testing of the surrogate models. Secondly, we propose two kinds of architectures (individual architectures for specific indicators and an integrative architecture) for the surrogate models. Thirdly, we train and optimize the models utilizing the graph schema, regional dataset and architectures. Finally, the optimized surrogate models are evaluated in two aspects: 1) the optimized models using the individual architectures for specific indicators and the ones using the integrative architecture are compared in terms of prediction accuracy and time costs; and 2) the time costs of the optimized model are analyzed by comparing with physics-based simulations. The results showed that surrogate models based on individual architectures outperform the model using the integrative architecture in terms of prediction accuracy and time costs for all sustainable performance indicators. Although the model preparation time of the surrogate models exceeds that of the physics-based simulations, the surrogate models reduce the calculation time from 6.346 min to 1.565 ms per case compared with the physics-based simulations.

Keywords: Surrogate model, Graph neural network, Building performance prediction, Sustainable building design, Residential block

1. INTRODUCTION

The subject of sustainable design for residential blocks is progressively gaining significance within the realm of global urbanization and the escalating demand for ecologically conscious habitats. In 2022, residential energy consumption accounted for 20.96% of the total energy utilized in the building sector worldwide (International Energy Agency 2023). Incorporating sustainable performance considerations during the initial design phase requires performance prediction, a fundamental process enabling the measurement of the level of sustainability achievable by designs. Building performance simulation using physical models is frequently utilized for performance prediction, such as structural and energy performance (Wong et al. 2023). However, its significant computational demands and time-intensive modeling pose challenges to its implementation in the early design phase (Attia et al. 2012). For example, in Natanian and Wortmann's study (Natanian and Wortmann 2021), the time taken for full energy simulation using EnergyPlus for a nine-block district was 40 min and 30 s per iteration (500 iterations in total). The extended computational duration would not be viable during the initial design phase, where it is preferable for the feedback time of the program to remain under the threshold of 10 seconds. (Miller 1968).

Surrogate models have been proposed to replicate computationally expensive physics-based simulation models (Westermann and Evins 2019). Two studies (Hu et al., 2023; Wang et al., 2021) have developed surrogate models to predict multiple indicators of sustainable performance in the early design stage of residential buildings

at block levels. These models offer substantial reductions in computational time (e.g., 500 times faster than physics-based simulation modeling (Hu et al. 2023)), while preserving an acceptable level of accuracy during the optimization process in the early design stage. However, these surrogate models are based on artificial neural networks (ANNs) in which the input structures are fixed. Hence, the models are case-specific and cannot represent general residential block design.

A graph is a data structure representing a set of nodes interconnected by edges (Zhou et al. 2020) that has been introduced to the architecture, engineering and construction (AEC) industry (Jia et al. 2023b) (e.g., building energy simulation (Wu et al. 2023a, b), generative design of sustainable buildings (Wu et al. 2024b, c), and compliance checking (Tao et al. 2024)). Graph neural networks (GNNs) are a class of neural networks specifically designed to operate on graph-structured data (Zhou et al. 2020) and have been adopted in surrogate models for building performance prediction (Hu et al. 2022; Lu et al. 2022; Jia et al. 2023a). However, there is a lack of research examining GNN-based surrogate models in the context of sustainable residential block design. Using GNNs in surrogate models has two advantages, which may address the issues found in the current surrogate models for sustainable residential block design. Firstly, GNNs can not only consider the impacts from the building itself but also incorporate the influences from the surrounding buildings. Secondly, GNNs can handle inputs with varying numbers of nodes or edges without requiring extensive modifications to the network architecture, and hence GNNs can enhance the generality of surrogate models.

In our previous study (Wu, et al., 2024), we developed a surrogate model for multi-objective sustainable performance prediction based on GNN. Firstly, a graph schema is proposed to represent the general topological relations among components in residential block layout design. Secondly, a dataset is established based on parametric design models of residential blocks and simulations of sustainable performance, including energy consumption, daylighting, and thermal comfort. Finally, an architecture using graph attention network (GAT) is proposed for multiple sustainable performance predictions. The results showed that the proposed model (GAT) outperforms the benchmark models (GCN and ANN) in terms of prediction accuracy, indicating that the inclusion of neural networks with message passing mechanisms that consider the impact of surrounding buildings leads to accuracy improvement. However, this is the preliminary study exploring GNN-based surrogate models for sustainable residential block design. Two issues remain to be solved. For one thing, the current surrogate models for predicting multiple indicators of residential blocks (Wang et al. 2021; Hu et al. 2023) were trained for each performance indicator individually. In contrast, in some studies on GNN-based prediction of multiple indicators (Lu et al. 2022; Li et al. 2023), one integrative model was trained for all the indicators. Consequently, when introducing GNN to the development of surrogate models for predicting multiple indicators of residential blocks, the adoption of architectures, the individual one or the integrative one, is an important issue to be addressed. For another, to facilitate smooth and efficient interaction between the designers and the simulation tool, the computational time needs to be short. However, the time costs of the GNN-based surrogate models for sustainable residential block design have not been discussed yet.

To solve these two issues, this study further explores the development of surrogate models for the early-stage design of residential blocks by leveraging GNN. Firstly, we introduce graph schema and regional dataset proposed the previous study (Wu et al., 2024). Secondly, we propose two kinds of architectures (individual architectures for specific indicators and an integrative architecture) for the GNN-based models. Thirdly, we train and optimize the GNN-based models utilizing the graph schema, regional dataset and architectures. Finally, the optimized surrogate models are evaluated in two aspects: 1) the optimized surrogate models using the individual architectures for specific indicators and the integrative architecture are compared in terms of prediction accuracy and time costs; and 2) the time costs of the optimized surrogate model are analyzed by comparing with physics-based simulations.

2. METHOD

2.1 Graph representation and dataset

A graph schema was developed to represent the general topological relations among components in residential block design (Wu et al., 2024). In the graph schema, nodes represent individual buildings within a residential zone and bi-directional edges represent positional relationships among the buildings. There is no edge connection between two buildings if these two buildings are completely obstructed by other buildings. Figure 1 shows the proposed graph model using one example of a residential zone with nine buildings. Nine nodes are used to represent the nine buildings. Bi-directional edges are established to represent the positional relationships among them, excluding buildings that are completely blocked by others (e.g., building 1 and building 3). The self-impact of a building on its sustainable performance is determined by building shape, building height, orientation, building volume, floor area, envelope area, window area, etc., while the impact of its surrounding buildings is determined by positional relationships (e.g., distances and relative angles) and obstructions between the two buildings. Accordingly, floor plan type, building height, window-to-wall ratio (WWR), north-south (N-S) projected length, west-east (W-E) projected length, shape factor and heat loss form factor are adopted as the node features, and relative angle, distance and visibility ratio are adopted as the edge features in the proposed graph schema.

The classification of floor plan types differs between regions. In this study, we take residential buildings in Hong Kong as an example and classify the floor plan type into six categories (Linear, L-shape, Y-shape, X-shape, Cruciform and Double-cruciform). To define the orientations of residential buildings, we introduce N-S projection length and W-E projection length to describe the building lengths along the north-south and west-east direction respectively (Figure 2 (a)). In addition, the shape factor (or shape coefficient) (determined by Eq. (1)) (Depecker et al. 2001) and heat loss form factor (determined by Eq. (2)) (Andrew 2021) are introduced as node features to incorporate the influence of building volume, floor area and envelope area.

$$SF = \frac{A_{envelope}}{V} \quad (1)$$

$$HLFF = \frac{A_{floor}}{A_{envelope}} \quad (2)$$

where SF is the shape factor, $HLFF$ is the heat loss form factor, V is the building volume, $A_{envelope}$ is the envelope area, and A_{floor} is the total floor area.

Figure 2 (b) defines the two edge features, distance and relative angle, which describe the positional relationships between two buildings. The centroids of the footprints of the two buildings are used to settle the distance and relative angle. To determine the obstruction conditions between two buildings, we propose a visibility ratio representing the portion of one building that can be “seen” by another. The visibility ratio is calculated by Eq. (3).

$$VP_{A \leftarrow B} = \frac{\sum_{i=1}^M \left(\frac{\sum_{j=1}^N \mathbb{1}(j \text{ is visible to } i)}{N} \right)}{M}, \quad (3)$$

where $VP_{A \leftarrow B}$ is the visibility ratio from building B to A (the portion of B that A can “see”), i is the sampling point on the exterior surface of building A, j is the sampling point on the exterior surface of building B, M is the number of sampling points on building A, N is the number of sampling points on building B, and $\mathbb{1}(j \text{ is visible to } i)$ is the indicator function that equates to 1 if there is no obstruction on the straight line from point j to point i , and 0 otherwise. Figure 2 (c) is a visualization of the visibility ratio in which building C is the obstruction between building A and B, the red point is one of the sampling points on building A, and the area in grey represents the area visible to the red point. The grid size of the sampling points adopted in this study is 10 m.

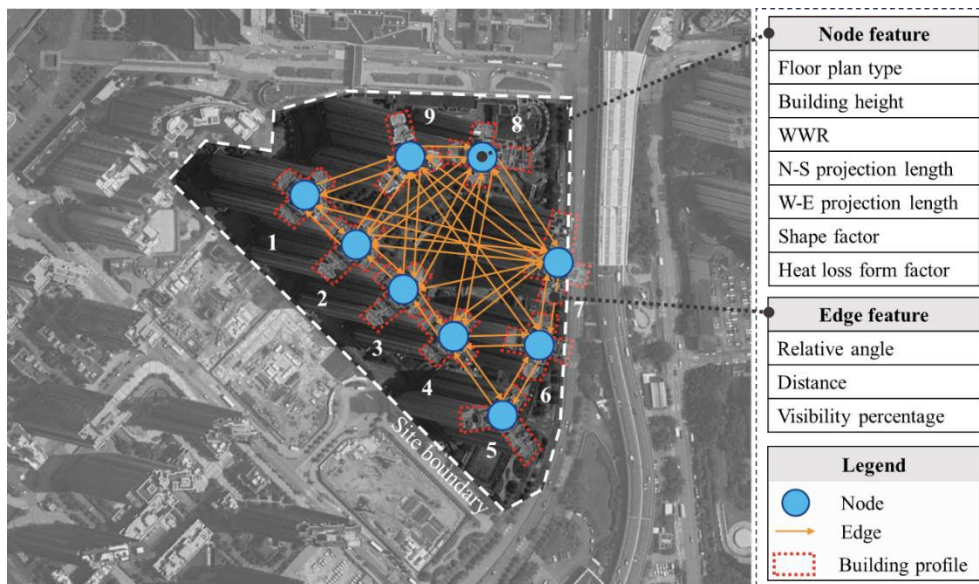


Figure 1. Graph representation of residential block layout design

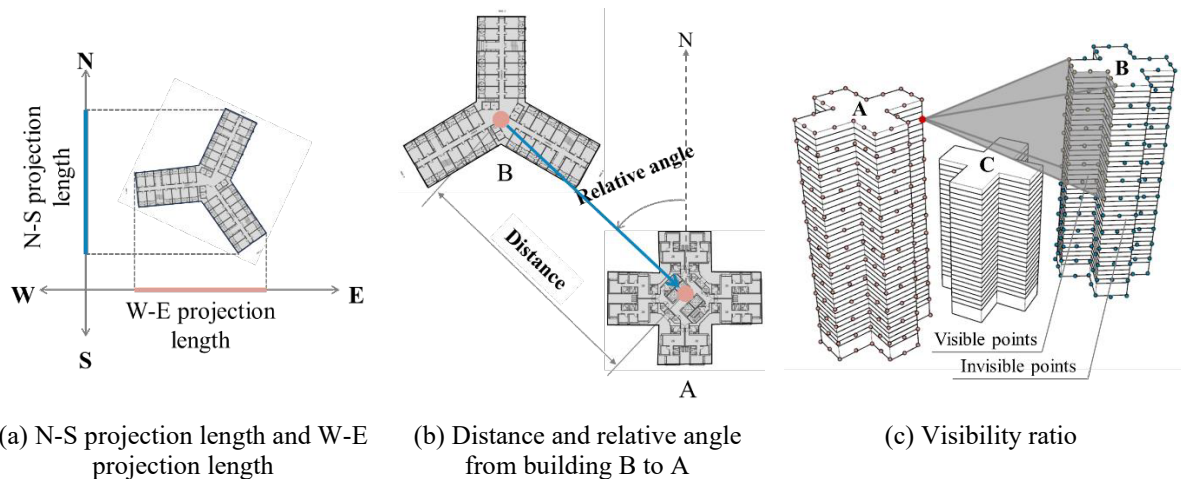


Figure 2. Definitions of N-S projection length, W-E projection length, distance, relative angle and visibility ratio

To enhance the generality of the surrogate model for residential block design, it is essential to collect comprehensive data that spans an entire region, where shared common patterns or frameworks in the design of residential blocks exist. In the previous study (Wu et al., 2024), we took residential blocks of public housing in Hong Kong as an example to generate a dataset for the training and testing of the GNN-based surrogate models. The dataset was generated based on parametric design models and performance simulation, considering the schedules and behaviors (such as window opening for natural ventilation (Wu et al. 2023c)) of the local residents. We used energy use intensity (EUI), annual comfort hours (ACH) and useful daylight illuminance (UDI) as the performance indicators of energy consumption, indoor thermal comfort and daylighting respectively. The dataset contains 9962 graphs, and each graph represents a residential zone, containing node features and performance indicators of each building and edge features. In this study, we introduce the dataset for the training and testing of the GNN-based surrogate models.

2.2 GNN architecture

We propose two kinds of architectures for our GNN-based surrogate models, individual architectures for specific indicators and an integrative architecture, as shown in Figure 3. The aim is to predict the sustainable performance indicators of each building in a residential zone, which is a node regression task (Zhou et al. 2020). The graphs containing node features and edge features of all the buildings serve as the inputs of the architecture. The main body of each architecture contains three parts: a first cluster of fully connected layers (FC layers) for pre-processing of the node features of the target building, a cluster of convolution layers (Conv layers) aggregating the edge features and node features from the target building itself and the surrounding buildings, and a second cluster of FC layers for post-processing after incorporating the information from the surroundings. In the individual architecture (Figure 3 (a)), the output only contains one performance indicator, and one model is trained for each of the performance indicators individually. In the integrative architecture (Figure 3 (b)), the outputs contain all the performance indicators, and one model is trained for all three performance indicators.

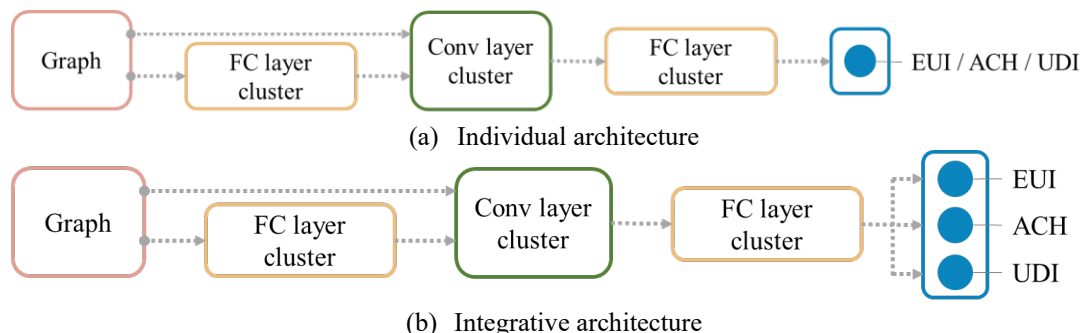


Figure 3. The two kinds of adopted architectures for the GNN-based surrogate models

Figure 4 shows the detailed architecture for predicting multiple indicators of sustainable performance of residential blocks, using the integrative architecture as the example. GAT (Brody et al. 2021) is applied to the Conv layers and the GATv2 operator (PyG Team 2024) is used to aggregate the node features and edge features from the target building and the surrounding buildings. In GAT, an attention mechanism is employed to assign different weights to the relations between nodes in a graph, which allows the network to focus on the most relevant

nodes during information propagation. Multiple heads are also employed in GAT, which enables the network to capture complementary aspects of the graph structure. Each Conv layer is connected to an FC layer before passing data to the subsequent Conv layer. At the beginning of the first cluster of FC layers and the end of the second cluster of FC layers, the number of channels gradually increases and decreases ($12 \rightarrow n/2 \rightarrow n$ and $n \rightarrow n/2 \rightarrow 3$) respectively to ensure a smooth change of the number of channels. The rectified linear unit (ReLU) is adopted as the activation function except the last FC layer, where the Sigmoid function is used. The architectures are parameterized with five parameters, the number of neurons (n), the number of heads in GAT (h), the number of layers in the pre-processing FC layer cluster ($N1$), the number of layers in the Conv layer cluster ($N2$), and the number of layers in the post-processing FC layer cluster ($N3$).

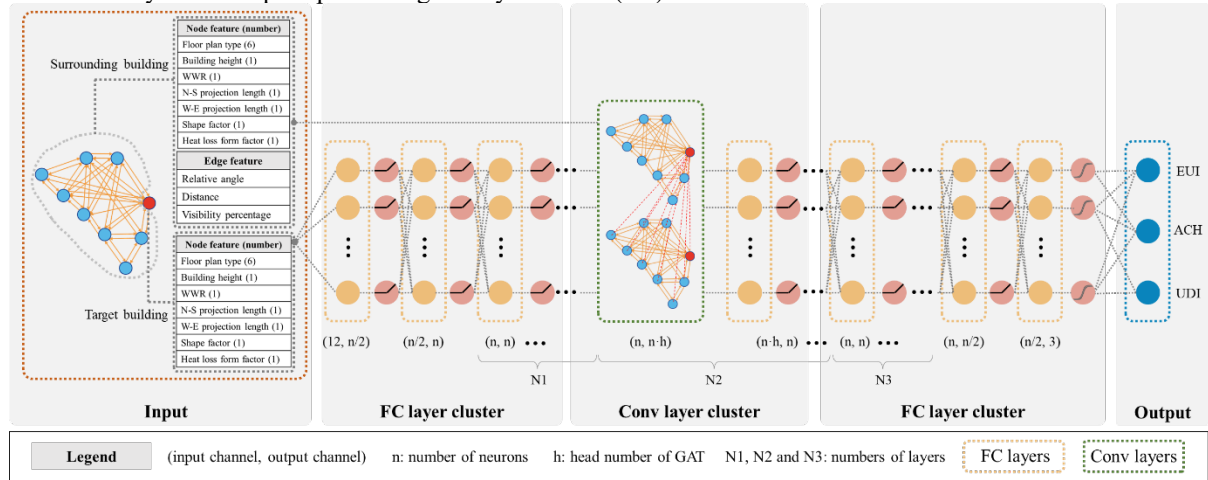


Figure 4. Details of the architecture (using the integrative architecture as the example).

2.3 Model training and evaluation

The dataset is randomly divided into a training set and a test set at the ratio of 75/25. The training data are then divided into smaller subsets and the model's parameters are updated based on the gradient computed from each mini-batch. Dropout is implemented to alleviate the overfitting issue. The training uses Adam as the optimizer and MSE as the loss function. The architecture parameters are optimized first, and the hyper-parameters of the models with optimal architectures are optimized subsequently. In the cases with the individual architecture, the model that exhibits the lowest coefficient of variation of the root mean square error (CV(RMSE)) value for the performance indicator in the test set is considered the optimal model. In the cases with the integrative architecture, the optimal model is the one with the lowest average CV(RMSE) of the three performance indicators. Grid search is used to optimize the architecture parameters and hyper-parameters. The search spaces of the parameters are listed in Table 1. To evaluate the performance of the models, the mean absolute error (MAE), root mean square error (RMSE), and CV(RMSE) are introduced, as given in Eqs. (4)–(6).

$$MAE = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i| \quad (4)$$

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (5)$$

$$CV(RMSE) = \sqrt{\frac{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2}{\frac{1}{m} \sum_{i=1}^m \hat{y}_i}} \quad (6)$$

where m is the number of all predictions, y_i is the i^{th} prediction and \hat{y}_i is the i^{th} ground truth.

The models are trained and tested using a desktop computer with the following specifications: Intel Core CPU i7-12700 at 2.10 GHz and 20 cores, NVIDIA GeForce RTX 3070 GPU, 32 GB RAM, and Windows 64 operating system. Python version 3.11.7, CUDA version 12.1 and PyTorch version 2.2.1 are used.

3. RESULTS AND DISCUSSION

3.1 Accuracy

Figure 5 shows the loss curves of the training of the GNN-based surrogate models using different architectures. All models reach convergence within 2000 epochs during the training. The optimal parameters and the optimal architectures are shown in Table 1 and Figure 6 respectively. The optimal values of each hyper-parameter (batch size, learning rate and dropout) are the same among the different architectures in the search space

used in this study. The optimal architectures for EUI and for ACH are similar, while the optimal architecture of UDI is different from the ones for EUI and for ACH. The different optimal architectures between EUI/ACH and UDI may result from the calculation mechanisms of the sustainable performance indicators. EUI and ACH are calculated based on thermodynamics while UDI is based on optics. Besides, although the optimal number of layers in the Conv layer cluster (N2), number of heads (h), number of layers in the pre-processing FC layer cluster (N1), and number of neurons (n) of the optimal individual architectures for EUI and for ACH are the same, the number of layers in the post-optimal processing FC layer cluster (N3) are different. The above result indicates that the calculations of EUI and ACH are similar in the pre-processing and the information aggregation, while the post-processing is different. Environmental parameters such as air temperature and relative humidity are calculated first based on thermodynamics for both EUI and ACH, while the thermal load is calculated for EUI and the thermal comfort index is calculated for ACH respectively later.

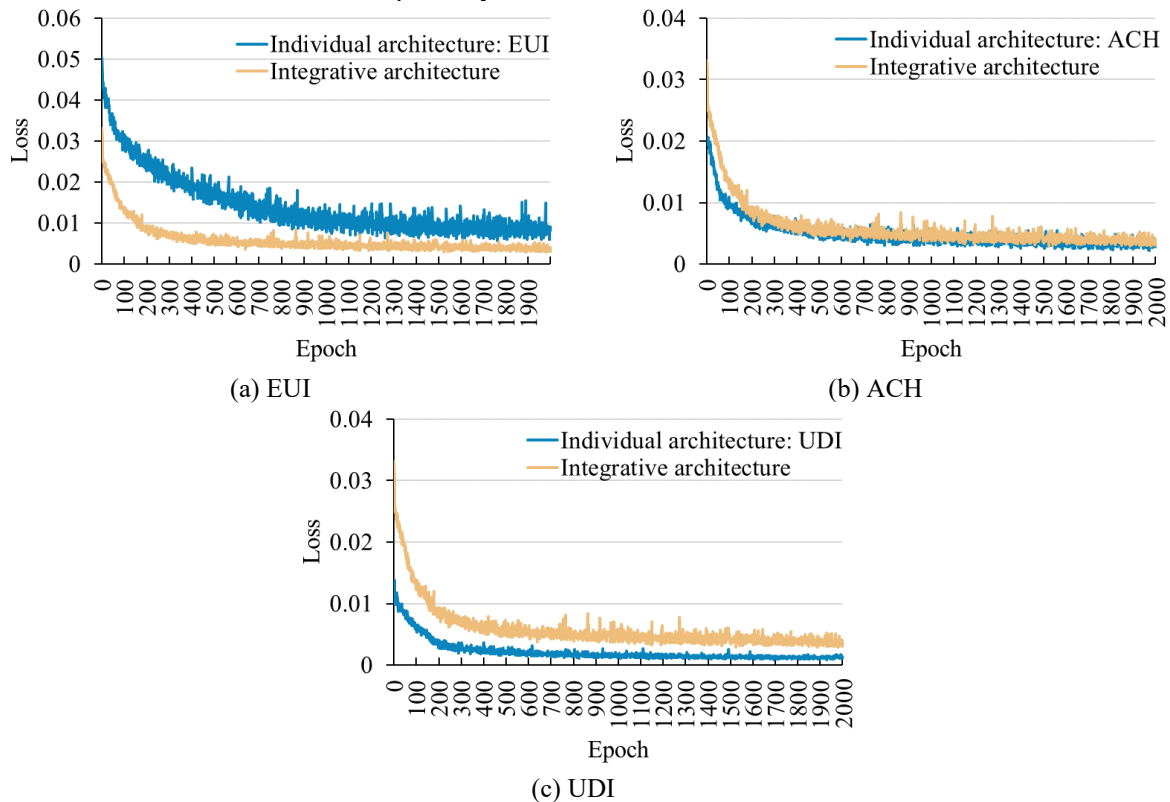


Figure 5. Loss curves of the training of the models using different architectures

Table 1. Optimal parameters of the individual architectures and the integrative architecture.

Parameter	Search space	Optimal value			
		Individual architecture for EUI	Individual architecture for ACH	Individual architecture for UDI	Integrative architecture
Architecture parameter					
Number of layers in the Conv layer cluster (N2)	[1, 2, 3, 4, 5, 6]	4	4	1	4
Number of heads (h)	[1, 2, 3, 4, 5, 6]	5	5	5	5
Number of layers in the pre-processing FC layer cluster (N1)	[0, 1, 2, 3]	1	1	0	1
Number of layers in the post-processing FC layer cluster (N3)	[0, 1, 2, 3]	0	2	0	1
Number of neurons (n)	[32, 64, 128, 256, 512]	64	64	128	128
Hyper-parameter					
Batch size	[32, 64, 128, 256, 512]	256			
Learning rate	[0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1]	0.001			
Dropout	[0, 0.1, 0.2, 0.3, 0.4]	0.2			

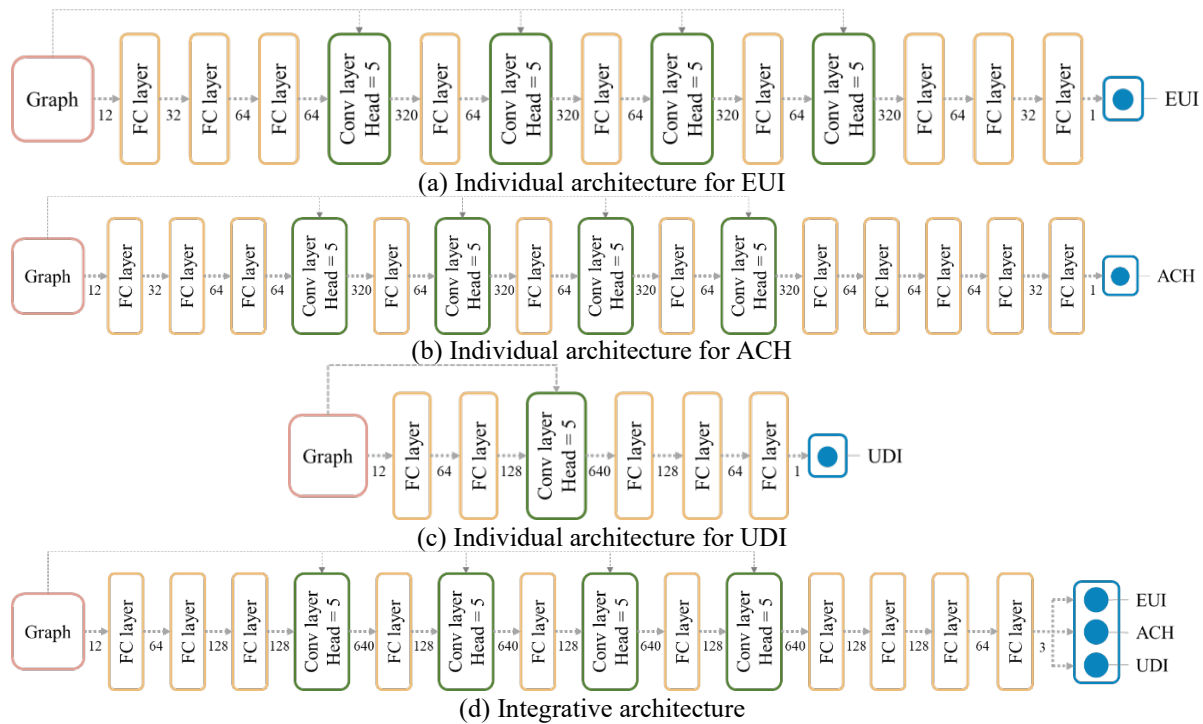


Figure 6. Optimal architectures (the numbers of channels are listed under the arrows between two layers).

In addition, the number of Conv layers determine the depth of information aggregation, and the different optimal numbers of layers in the Conv layer cluster (N2) indicate the different information aggregation mechanisms among different sustainable performance. The optimal number of layers in the Conv layer cluster (N2) of the optimal individual architecture for UDI is one, showing that the daylighting performance of the target buildings is only impacted significantly by their 1-hop neighbors. In contrast, the optimal numbers of layers in the Conv layer cluster (N2) of the optimal individual architectures for EUI and for ACH are both four, showing that the thermal environment-related performance of the target buildings is impacted significantly by their 4-hop neighbors. According to the definition of edges in this study (Section 3.1), edge connections are established between two buildings only if those two buildings are not completely obstructed by other buildings. The daylighting performance of the target buildings is impacted by shading and daylight reflectance from the adjacent buildings that the target buildings can “see”. Thus, it is reasonable that the optimal individual architecture for UDI uses one Conv layer (i.e., only considering influence from 1-hop neighbors). In contrast, the thermal conditions of the target buildings may not only be affected by the buildings adjacent to the target buildings, but depend on the heat balance among a certain hop pf neighbors. Therefore, the optimal individual architectures for EUI and for ACH use multiple Conv layers, and 4-hop neighbors are considered to affect significantly on the target buildings.

Table 2 shows the prediction accuracy of the GNN-based surrogate models using different architectures. For each sustainable performance indicator, the surrogate model using the individual architecture outperforms the model using the integrative architecture, with smaller RMSEs and CV(RMSE)s on the test set. The difference in the prediction accuracy between the individual and the integrative architecture is minor for EUI (CV(RMSE) difference of 0.5% ($|11.79\%-11.85\%| / 11.85\%$)), slightly higher for ACH (CV(RMSE) difference of 3.2% ($|7.63\%-7.88\%| / 7.88\%$)), and largest for UDI (CV(RMSE) difference of 21.9% ($|8.00\%-10.14\%| / 10.14\%$)). The optimal integrative architecture is close to the optimal individual architectures for EUI and for ACH, and is distinct from the optimal individual architecture for UDI, which accounts for the different prediction accuracy among the three performance indicators. During the model training, we use the minimal average CV(RMSE) of the three performance indicators as the optimization objective of the integrative architecture (Section 3.4). Therefore, the optimal result of the integrative architecture is a balance among the three performance indicators. The result shows that the balance tends to shift to the side of EUI and ACH. In the integrative architecture, the parameters of the model are shared by all the sustainable performance indicators, while the parameters in the models using the individual architectures are trained specifically for each sustainable performance indicator. Consequently, the prediction accuracy of the models using the individual architectures for specific indicators are higher than that of the model using the integrative architecture.

Table 2. Prediction accuracy using different architectures.

	EUI			ACH			UDI		
	MAE	RMSE	CV(RMSE)	MAE	RMSE	CV(RMSE)	MAE	RMSE	CV(RMSE)
Training set									
Individual architecture*	11.88	19.27	10.11%	1.41	2.17	5.48%	1.63	2.21	3.75%
Integrative architecture	9.65	16.42	8.62%	1.45	2.30	5.80%	3.59	5.71	9.69%
Test set									
Individual architecture*	13.77	22.48	11.79%	1.86	3.01	7.63%	2.99	4.73	8.00%
Integrative architecture	12.98	22.60	11.85%	1.89	3.11	7.88%	3.83	5.99	10.14%

3.2 Time cost

Table 3 shows the training time of the GNN-based surrogate models using different architectures. For the individual architectures, the training time is proportional to the complexity of the models. The individual architecture for ACH has the largest complexity and hence takes the longest time to train. The total training time of the models using the individual architectures is 5,391 s while that of the model using the integrative architecture is 8,040 s. If the models using the individual architectures were trained in parallel, the training time could be reduced to 2,115 s (determined by the longest training time among the individual architectures). The total calculation time of the models using the individual architectures is 1.565 ms per case on average, while that of the model using the integrative architecture is 2.338 ms. Therefore, the individual architectures outperform the integrative architecture in terms of the time costs. If both the prediction accuracy and the time costs are considered, we can conclude that the individual architectures should be adopted over the integrative architecture when developing GNN-based surrogate models for predicting multiple indicators of sustainable performance in the early design stage of residential blocks.

Table 3 shows the time costs of the sustainable performance predictions using the surrogate models and the physics-based simulations. For the surrogate model-based prediction, the most time-consuming part is the model preparation. It takes approximately one day to manually develop the simulation models in the simulation software, while it takes more than 90 days to prepare the surrogate models (i.e., more than 90 times longer than the physics-based simulation modeling). However, the surrogate model-based method achieves a dramatically faster calculation speed than the physics-based simulation modeling by 243,297 times (from 6.346 min to 1.565 ms) and 104,062 times (from 6.346 min to 2.338 ms) when using the individual architectures and integrative architecture respectively. The calculation time of the surrogate models are within one second and hence the surrogate models are suitable for the early design stage, in which the acceptable feedback time should be less than 10 seconds (Miller 1968).

Table 3. Time costs of indicator predictions using the surrogate models and the physics-based simulations.

	Surrogate model	Physics-based simulation
Model preparation	Dataset generation: ~90 d. Model training: <u>Individual architectures</u> EUI: 2,095 s, ACH: 2,115 s, UDI: 1,181 s, In total: 5,391 s. <u>Integrative architecture</u> 8,040 s.	Simulation modeling: ~ 1d.
Calculation (per case)*	<u>Individual architecture</u> EUI: 0.410 ms, ACH: 0.751 ms, UDI: 0.404 ms, In total: 1.565 ms. <u>Integrative architecture</u> 2.338 ms.	EUI and ACH (using EnergyPlus): 4.479 min, UDI (using Radiance): 1.867 min, In total: 6.346 min.

4. CONCLUSIONS

In this study, we developed GNN-based surrogate models to predict multi-objective sustainable performance indicators for residential blocks. The research consisted of several key steps. We introduced graph schema and regional dataset proposed the previous study (Wu et al., 2024). We proposed two kinds of architectures

(individual architectures for specific indicators and an integrative architecture) for the GNN-based models. Thirdly, we trained and optimized the GNN-based models utilizing the graph schema, regional dataset and architectures. Finally, the optimized surrogate models are evaluated in terms of accuracy and time costs.

The major findings of this study are summarized as follows.

- Surrogate models based on individual architectures outperform the model using the integrative architecture in terms of prediction accuracy and time costs for all sustainable performance indicators. Therefore, individual architectures are recommended for developing GNN models for predicting multiple performance indicators.

- Although the model preparation time of the surrogate models exceed that of the physics-based simulations, the surrogate models reduce the calculation time from 6.346 min to 1.565 ms per case compared with the physics-based simulations. This demonstrates that the GNN-based surrogate models can significantly accelerate the performance evaluation in the early design stage of residential blocks, and therefore facilitate a smoother performance-based design.

Some limitations have been identified and should be addressed in future studies.

- The dataset only includes buildings within a single residential zone, neglecting the potential impacts of surrounding buildings outside the site boundary. Future studies should incorporate these external factors that may affect buildings within the residential zone.

- This study focuses solely on three indoor sustainable performance metrics. Future studies should encompass a wider range of sustainable performance indicators, representing both indoor and outdoor environments.

REFERENCES

- Andrew S (2021) How to calculate form heat loss factor. - BuildPass. <https://www.buildpass.co.uk/blog/how-to-calculate-form-heat-loss-factor/>. Accessed 13 Jun 2024
- Attia S, Gratia E, De Herde A, Hensen JL (2012) Simulation-based decision support tool for early stages of zero-energy building design. *Energy and buildings* 49:2–15
- Brody S, Alon U, Yahav E (2021) How attentive are graph attention networks? arXiv preprint arXiv:210514491
- Depecker P, Menezo C, Virgone J, Lepers S (2001) Design of buildings shape and energetic consumption. *Building and Environment* 36:627–635. [https://doi.org/10.1016/S0360-1323\(00\)00044-5](https://doi.org/10.1016/S0360-1323(00)00044-5)
- Hu Y, Cheng X, Wang S, et al (2022) Times series forecasting for urban building energy consumption based on graph convolutional network. *Applied Energy* 307:118231. <https://doi.org/10.1016/j.apenergy.2021.118231>
- Hu Z, Zhang L, Shen Q, et al (2023) An integrated framework for residential layout designs: Combining parametric modeling, neural networks, and multi-objective optimization for outdoor activity space optimization. *Alexandria Engineering Journal* 80:202–216. <https://doi.org/10.1016/j.aej.2023.08.049>
- International Energy Agency (2023) Buildings - Energy System - IEA. <https://www.iea.org/energy-system/buildings>. Accessed 15 Nov 2024
- Jia Y, Wang J, Hosseini MR, et al (2023a) Temporal Graph Attention Network for Building Thermal Load Prediction. *Energy and Buildings* 113507
- Jia Y, Wang J, Shou W, et al (2023b) Graph neural networks for construction applications. *Automation in Construction* 154:104984. <https://doi.org/10.1016/j.autcon.2023.104984>
- Li M, Liu Y, Wong BCL, et al (2023) Automated structural design optimization of steel reinforcement using graph neural network and exploratory genetic algorithms. *Automation in Construction* 146:104677. <https://doi.org/10.1016/j.autcon.2022.104677>
- Lu J, Zhang C, Li J, et al (2022) Graph convolutional networks-based method for estimating design loads of complex buildings in the preliminary design stage. *Applied Energy* 322:119478. <https://doi.org/10.1016/j.apenergy.2022.119478>
- Miller RB (1968) Response time in man-computer conversational transactions. In: *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*. pp 267–277
- Natanian J, Wortmann T (2021) Simplified evaluation metrics for generative energy-driven urban design: A morphological study of residential blocks in Tel Aviv. *Energy and Buildings* 240:110916. <https://doi.org/10.1016/j.enbuild.2021.110916>
- PyG Team (2024) torch_geometric.nn.conv.GATv2Conv - pytorch_geometric documentation. https://pytorch-geometric.readthedocs.io/en/2.5.0/generated/torch_geometric.nn.conv.GATv2Conv.html#torch_geometric.nn.conv.GATv2Conv. Accessed 13 Jun 2024
- Tao X, Wu Z, Xu Y, et al (2024) Smarter smart contracts for automatic BIM metadata compliance checking in blockchain-enabled common data environment. *Advanced Engineering Informatics* 62:102627. <https://doi.org/10.1016/j.aei.2024.102627>
- Wang S, Yi YK, Liu N (2021) Multi-objective optimization (MOO) for high-rise residential buildings' layout

- centered on daylight, visual, and outdoor thermal metrics in China. *Building and Environment* 205:108263. <https://doi.org/10.1016/j.buildenv.2021.108263>
- Westermann P, Evins R (2019) Surrogate modelling for sustainable building design – A review. *Energy and Buildings* 198:170–186. <https://doi.org/10.1016/j.enbuild.2019.05.057>
- Wong BCL, Wu Z, Gan VJL, et al (2023) Parametric building information modelling and optimality criteria methods for automated multi-objective optimisation of structural and energy efficiency. *Journal of Building Engineering* 75:107068. <https://doi.org/10.1016/j.jobbe.2023.107068>
- Wu Z, Cheng JCP, Wang Z (2023a) An ontology-based framework for building energy simulation in the operation phase. In: *Advances in Information Technology in Civil and Building Engineering*. Springer International Publishing, Cham, pp 351–366
- Wu Z, Cheng JCP, Wang Z, Kwok HHL (2023b) An ontology-based framework for automatic building energy modeling with thermal zoning. *Energy and Buildings* 296:113267. <https://doi.org/10.1016/j.enbuild.2023.113267>
- Wu Z, Li M, Liu W, et al (2024a) A data-driven model for sustainable performance prediction of residential block layout design using graph neural network. In: *Proceedings of the 42nd Conference on Education and Research in Computer Aided Architectural Design in Europe (eCAADe 2024)*. Nicosia, Cyprus
- Wu Z, Wang Z, Cheng JCP, Kwok HHL (2024b) A knowledge-informed optimization framework for performance-based generative design of sustainable buildings. *Applied Energy* 367:123318. <https://doi.org/10.1016/j.apenergy.2024.123318>
- Wu Z, Wang Z, Cheng JCP, Kwok HHL (2024c) A knowledge graph model for performance-based generative design and its applications in accelerated design. In: *Proceedings of The 29th Annual Conference for Computer-Aided Architectural Design Research in Asia (CAADRIA 2024)*. Singapore
- Wu Z, Zhang Y, Mai J, et al (2023c) Adaptation-based indoor environment control with night natural ventilation in autumn in an office building in a hot-humid area. *Building and Environment* 243:110702. <https://doi.org/10.1016/j.buildenv.2023.110702>
- Zhou J, Cui G, Hu S, et al (2020) Graph neural networks: A review of methods and applications. *AI Open* 1:57–81. <https://doi.org/10.1016/j.aiopen.2021.01.001>

GEOSPATIAL GRAPH ATTENTION NETWORK FOR HIGH-RESOLUTION BUILDING FAÇADE PHOTOVOLTAIC POTENTIAL PREDICTION

Zheng Li¹, Jun Ma¹

1) Ph.D. Candidate, Department of Urban Planning and Design, The University of Hong Kong, Hong Kong, China. Email: zhengli@connect.hku.hk

2) Ph.D., Assistant. Prof., Department of Urban Planning and Design, The University of Hong Kong, Hong Kong, China. Email: junma@hku.hk

Abstract: Accurately predicting the photovoltaic (PV) potential of urban building façades plays a crucial role in the development of photovoltaics. This study proposes an innovative building façade PV potential prediction method based on the Geospatial Graph Attention Neural Network (GGAT). Compared to traditional methods, this approach considers the differences in solar radiation intensity at various heights of the building façade, enabling more precise identification of areas with higher PV potential on the façade. The study focuses on buildings in the Manhattan area of New York City and employs Rhino software and the Ladybug Tools plugin to conduct building solar radiation simulations, obtaining high-quality training data. During the modeling process, the concept of building height stratification is introduced, dividing the building façade vertically into 10 equal-height layers, with each prediction point representing the average solar radiation intensity within that height range. Experimental results indicate that GNN-based algorithms (especially GGAT) outperform traditional machine learning algorithms in predicting solar radiation on building façades. GGAT integrates geospatial features and graph attention mechanisms, enabling more accurate prediction of solar radiation on building façades. Solar radiation intensity exhibits significant differences both in the vertical direction of the building façade and in the horizontal direction (between census tracts). The stratified modeling method can reveal these differences, providing more comprehensive and detailed information for analyzing the PV potential of building façades.

Keywords: Façade photovoltaic potential, Graph attention network, Solar irradiation prediction, Machine learning.

1. INTRODUCTION

The world is facing increasingly severe energy crises and global warming issues, necessitating effective measures to address these challenges (Ahmed et al. 2022). As a major consumer of urban energy and a primary source of greenhouse gas emissions, buildings play a crucial role in achieving sustainable urban development and low-carbon transition. By installing solar photovoltaic (PV) systems on buildings, we can fully utilize clean and renewable solar energy resources for power generation, reducing buildings' reliance on fossil fuels and effectively lowering carbon emissions during building operations (Brown et al. 2024). This not only helps mitigate global warming but also makes significant contributions to achieving low-carbon transition and climate change goals in cities.

Building PV systems can be generally classified into two main categories based on their installation locations: rooftop PV systems and façade PV systems (Petter Jelle, Breivik, and Drolsum Røkenes 2012). The former involves installing solar panels on the rooftops of buildings, utilizing the roof space for power generation. This approach is relatively simple to install, does not occupy additional land resources, and does not affect the main structure and appearance of the building. Rooftop PV systems are suitable for various types of buildings, such as residential, commercial, and industrial buildings. On the other hand, façade PV systems integrate solar panels into the exterior walls of buildings, forming a building-integrated photovoltaic (BIPV) curtain wall. This approach not only generates electricity but also serves as a decorative material for the building's exterior, enhancing its aesthetic value.

Accurately predicting the installable area for PV systems on buildings is of great significance for the promotion and application of PV systems (Tian, Ooka, and Lee 2023). The size of the installable area directly determines the capacity of PV systems that can be installed on buildings, which in turn affects the power generation potential and economic benefits of the PV systems. By accurately assessing the installable area of buildings, we can determine the PV power generation potential of buildings and provide a basis for formulating PV system promotion policies and targets. Therefore, developing efficient and accurate methods for predicting the installable area of PV systems is of crucial practical significance for promoting the large-scale application of PV systems in urban buildings.

1.1 Related works

To gain a comprehensive understanding of the current research status and progress in building PV potential prediction, this study conducts a literature review and analysis based on two categories: rooftop PV potential and façade PV potential (Gassar and Cha 2021).

1.1.1 Rooftop PV Potential Prediction

Rooftops are ideal locations for installing PV systems, making rooftop PV potential prediction a research hotspot. Due to the relatively easy accessibility of building rooftop data and the simplicity of rooftop PV system installation, this area has received widespread attention, and research methods and techniques have become increasingly mature (Fakhraian et al. 2021).

For example, Singh et al. (Singh and Banerjee 2015) proposed a method for estimating the rooftop PV potential of a region, using Mumbai as a case study. Utilizing publicly available data, GIS analysis, and PVSyst simulation, they estimated the building footprint area ratio and available rooftop area ratio while considering factors such as irradiance, temperature, and tilt angle. The results showed that Mumbai's PV potential was 2190 MW with a capacity factor of 14.8%, capable of meeting 12.8-20% of the daily average demand and 31-60% of the early peak demand. With the development of big data and artificial intelligence technologies, some researchers have begun to explore the use of advanced algorithms for assessing rooftop PV potential. Zhong et al. (Zhong et al. 2021) proposed a general framework for estimating urban-scale rooftop solar PV potential using publicly available high-resolution satellite imagery. They developed a deep learning-based method for automatic extraction of rooftop areas. To address the labor-intensive issue of training rooftop extraction models, they developed a spatial optimization sampling strategy. In a case study in Nanjing, China, the manual cost of preparing the training dataset was reduced by approximately 80%, and the robustness of the extraction model was improved. The total rooftop area in Nanjing was 330.36 square kilometers, with an estimated potential installed capacity of 66 GW and an annual power generation of approximately 49,897 GWh in 2019.

In summary, rooftop PV potential prediction has been widely studied, with research methods and techniques continuously evolving and improving. From early GIS and remote sensing technologies to recent machine learning and deep learning methods, researchers have been exploring new approaches to enhance the accuracy and efficiency of rooftop PV potential prediction. These research findings provide important theoretical foundations and technical support for the planning, design, and promotion of rooftop PV systems.

1.1.2 Façade PV Potential Prediction

Compared to rooftop PV systems, the potential prediction of façade PV systems has received relatively less attention (Catita et al. 2014). However, with the development of BIPV technology, the application prospects of façade PV systems are becoming increasingly broad. In recent years, more and more studies have begun to focus on façade PV potential prediction and have proposed various novel methods and techniques.

Some studies utilize three-dimensional building models for simulation to assess the power generation potential of façade PV systems. For example, Brito et al. (Bruto et al. 2017) used a digital surface model obtained from LiDAR measurements and typical meteorological year data to calculate the PV potential of two typical case studies in Lisbon, Portugal, and compared it with the estimated local electricity demand. The results showed that the rooftop and façade PV potential exceeded the local non-base load demand and could meet 50-75% of the total electricity demand. Considering the solar potential of façades, PV generation could meet the electricity demand during winter noons. Economic analysis revealed that installing PV only on rooftops could result in a payback period of less than 10 years, while a 50-50 split between rooftops and façades would yield a payback period of 15 years. Additionally, Liu et al. (Liu et al. 2023) proposed an innovative method that, for the first time, utilized publicly available satellite imagery and vector maps to construct 3D building models for rural areas and precisely assessed the solar PV potential of rural rooftops and façades. The method was validated using two real 3D village models and on-site solar radiation measurements. The case study showed that south-facing and north-facing rural rooftops, as well as south-facing and west-facing façades, had the highest PV potential grades. North-facing rooftops with a slope of 30° accounted for 32.7% of the total rooftop solar PV potential and should not be neglected in future assessments. The method is cost-effective and can accurately assess rural solar PV potential at both micro and macro scales, contributing to the promotion of rural renewable energy penetration.

From the above review, it is evident that the existing assessment methods for façade PV potential still have some shortcomings. First, they often treat the entire façade as a whole, ignoring the differences in solar radiation intensity at different positions on the façade due to shading from surrounding buildings. This simplified approach may lead to biased assessment results and underestimate the actual PV potential of façades. When the façade is treated as a whole, the contribution of local high-radiation intensity areas may be averaged out, resulting in the entire façade exhibiting a lower average radiation level, masking the PV potential at specific positions on the façade. Second, most high-precision studies have a small spatial scale, and their assessment methods and conclusions may be difficult to generalize to the urban scale. These studies are often based on a limited sample of buildings and may not fully consider the complexity and diversity of the urban environment. Therefore, the existing research results may have limitations in guiding the formulation of façade PV policies at the urban level.

1.2 Objective

To address the aforementioned shortcomings, this study proposes a more refined urban-scale building

façade PV potential assessment method that can evaluate the solar radiation intensity at different heights of each building's façade. Compared to traditional methods, this novel approach takes into account the differences in radiation intensity at various heights on the façade, enabling more accurate identification of façade areas with higher PV potential. This method can optimize the economic viability of PV systems by avoiding unnecessary costs associated with installing PV systems in low radiation intensity areas.

2. METHODOLOGY

In urban environments, predicting the PV potential of building façades is a complex and challenging task. Building façades are often subject to shading from surrounding buildings, resulting in significant differences in solar radiation intensity at various positions on the façade. Therefore, considering the influence of surrounding buildings on the target building is crucial when predicting PV potential. Traditional methods often struggle to effectively capture the complex spatial relationships between buildings, but Graph Neural Networks (GNNs) (Scarselli et al. 2008) provide a new perspective for addressing this issue.

GNNs can abstract buildings as nodes in a graph structure and represent the spatial relationships between buildings through edges. By propagating and aggregating information on the graph, GNNs can aggregate the features of surrounding buildings onto the target building, thereby more comprehensively considering the impact of the surrounding environment on the target building's PV potential. However, in real urban scenarios, the influence of surrounding buildings on the target building is often imbalanced. For instance, buildings that are closer in proximity, larger in volume, or taller in height may have a more significant shading effect on the target building, while the influence of buildings that are farther away or smaller in volume is relatively weaker.

To address this issue, we introduce the Graph Attention Network (GAT) model (Veličković et al. 2018). Unlike conventional GNNs, the GAT model can adaptively adjust the influence weights between different buildings through an attention mechanism. Specifically, the GAT model calculates an attention coefficient based on the features of buildings (such as height, volume, etc.) to measure the relevance between different buildings. Surrounding buildings that have a greater impact on the target building will be assigned higher attention weights, while those with lesser influence will be correspondingly weakened. In this way, the GAT model can more accurately capture the complex spatial relationships between buildings, thereby improving the accuracy of PV potential prediction.

Furthermore, considering the uniqueness of the building façade PV potential prediction problem, we have innovated and extended upon the GAT model. We observed that, in addition to the features of the buildings themselves, many other geospatial features have a significant impact on PV potential, such as distance and azimuth angle. To fully utilize this geospatial information, we propose a novel model called the Geospatial Graph Attention Neural Network (GGAT). In the GGAT model, we incorporate more geospatial features into the calculation of attention coefficients, enabling the model to simultaneously consider building features and geospatial features, thereby more comprehensively characterizing the PV potential of building façades.

2.1 Building Solar Radiation Simulation

Acquiring high-quality training data is a crucial step in the study of building façade PV potential prediction. To train our proposed GGAT model, we require solar radiation intensity data for building façades. However, directly measuring the solar radiation intensity of every building façade in real-world scenarios is extremely difficult and expensive. To address this issue, we employ building solar radiation simulation methods to generate training data. First, we create three-dimensional models of the buildings in the study area using Rhino software. Next, we perform building solar radiation simulations in Rhino using the Ladybug Tools plugin. During the simulation process, we divide the building façades into multiple grids and calculate the solar radiation intensity received by each grid at different time points. After the simulation is complete, we export the solar radiation intensity data for each grid and associate it with the corresponding building. Through this approach, we obtain a dataset containing building IDs, grid IDs, timestamps, and solar radiation intensities.

2.2 Building Height Stratification

Accurately estimating the solar radiation intensity at different positions on building façades is crucial in predicting PV potential. Traditional methods often treat the building façade as a whole and only calculate the average solar radiation intensity. However, in reality, due to factors such as shading from surrounding buildings and differences in sky view factors, the solar radiation intensity at different heights of the building façade may vary significantly. To more precisely predict the PV potential on façades, we propose a stratified modeling approach.

Our method is based on the following assumption: positions at the same height on the building façade receive the same solar radiation intensity. This assumption simplifies the lighting conditions of the building façade, ignoring horizontal differences and focusing on vertical variations. Although this simplification may introduce some errors, it greatly reduces computational complexity, enabling us to perform building façade PV potential prediction more efficiently. Specifically, we divide the façade of each building vertically into 10 equal layers, with

each layer approximately 10% of the total building height. Through this stratification approach, we obtain 10 façade prediction points for each building, with each prediction point representing the average solar radiation intensity within that height range (Figure 1). Compared to traditional holistic modeling methods, our stratified approach can more finely characterize the vertical lighting differences on building façades, thereby improving the accuracy of PV potential prediction.

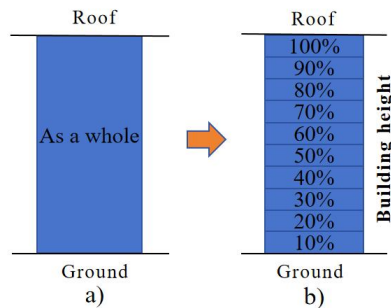


Figure 1. A brief diagram of the traditional method and our stratification method.

3. RESULTS AND DISCUSSION

To validate the effectiveness of the GGAT model, we selected buildings in the Manhattan district of New York City (NYC) as the research object. It includes approximately 45,000 buildings in total. To further enrich the dataset, we also introduced other common features of Manhattan buildings, such as building density, building shape coefficient, average height of surrounding buildings, etc. These features were obtained from the NYC open data website or can be derived by analyzing the three-dimensional building models. By combining these geospatial features with the solar radiation intensity data, we obtained a more comprehensive and rich training dataset.

3.1 Model performance comparison

Table 1 presents a performance comparison of six different algorithms in predicting solar radiation intensity and total radiation on building façades. These six algorithms are divided into two main categories: traditional machine learning algorithms and GNNs based algorithms. In the traditional machine learning algorithms, this study selected gradient boosting decision trees (GBDT), random forest, and deep neural networks (DNN). These three algorithms have been widely applied in many fields and can serve as benchmarks for comparison. In the GNNs algorithms, we selected graph convolutional networks (GCN), GAT, and GGAT. GCN and GAT are two models that have received extensive attention in the field of graph learning in recent years, while GGAT is a model specifically designed for the task of predicting solar radiation on building façades based on GAT.

From the experimental results, the three GNNs algorithms generally outperform the traditional machine learning algorithms. Whether in the task of predicting radiation intensity (RMSE and R-Squared metrics) or total radiation, the performance of GCN, GAT, and GGAT is better than GBDT, random forest, and DNN. This indicates that by introducing graph structures and attention mechanisms, GNNs algorithms can better model the spatial relationships between buildings, thereby improving prediction performance. Among the three GNNs algorithms, the GGAT model proposed in this study achieves the best performance. Compared with GCN and GAT, GGAT has significant advantages in both RMSE and R-Squared metrics for predicting radiation intensity and total radiation. This demonstrates that GGAT can more accurately predict the solar radiation on building façades by integrating geospatial features and graph attention mechanisms.

Table 1. The model performance on building façade solar irradiation predictions.

		Facade			
		Irradiation intensity		Total radiation	
		RMSE	R-Squared	RMSE	R-Squared
Machine learning	GBDT	0.7539	0.5181	1.0456	0.6912
	Random forest	0.6653	0.6063	0.9407	0.7500
	DNN	0.7347	0.5199	1.1035	0.6561
GNNs	Graph Convolutional Network (GCN)	0.6451	0.6340	0.8803	0.7932
	Graph attention network (GAT)	0.5945	0.6837	0.8213	0.8425
	GGAT (our work)	0.5774	0.7067	0.8005	0.8635

3.2 The distribution of solar irradiation intensity

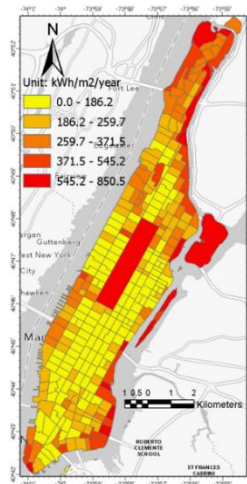


Figure 2. The distribution of façade irradiation intensity at the census tract-level (façade considered as a whole).

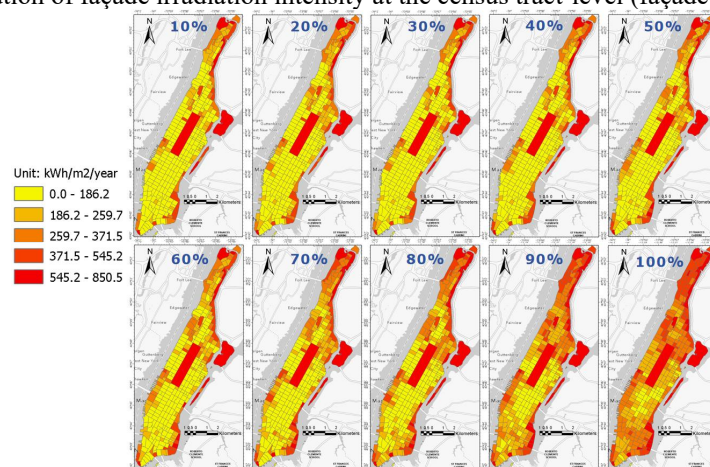


Figure 3. The census tract-level distribution of irradiation intensity when the façade is stratified.

Figure 2 and Figure 3 illustrate the spatial distribution of solar radiation intensity on building façades in the Manhattan area at the census tract level, comparing the differences between two modeling approaches: treating the building façade as a whole (traditional method) and stratifying by height (the method proposed in this study).

In Figure 2, the color of each census tract represents the average solar radiation intensity of all building façades within that tract. It can be observed that the solar radiation intensity exhibits significant spatial heterogeneity across the Manhattan area, with tracts in the southern and northern parts generally having higher radiation intensity compared to those in the central part. This may be related to factors such as building height, density, and street orientation. Figure 3 shows the distribution of solar radiation intensity under 10% building height stratification. Each row represents a height layer, from bottom to top: 0-10%, 10-20%, ..., 90-100%. It can be seen that as the height increases, the solar radiation intensity of each tract gradually increases, with colors transitioning from yellow-green to red. This indicates that the difference in radiation intensity along the vertical direction of the building façade is very significant.

By comparing Figure 2 and Figure 3, we can conclude that the traditional holistic modeling approach ignores the difference in radiation intensity along the vertical direction of the building façade, while the stratified modeling approach can reveal this difference. The solar radiation intensity of building façades not only varies significantly in the horizontal direction (between census tracts) but also in the vertical direction (building height). Therefore, it is necessary to comprehensively consider these two dimensions. The stratified modeling approach provides more comprehensive and fine-grained information for analyzing the PV potential of building façades, which helps to formulate more precise PV deployment strategies.

Overall, these two figures intuitively showcase the spatial distribution characteristics of solar radiation intensity on building façades in the Manhattan area and highlight the advantages of the stratified modeling approach compared to traditional methods. This provides important references for subsequent building PV potential assessment and planning.

5. CONCLUSIONS

The study proposes an innovative building façade PV potential prediction method based on the GGAT

model. By incorporating geospatial features into the graph attention network, the GGAT model can more comprehensively characterize the complex spatial relationships between buildings, thereby improving the accuracy of PV potential prediction. Experimental results demonstrate that the GGAT model outperforms traditional machine learning algorithms and other GNN algorithms in predicting solar radiation intensity and total radiation on building façades. This provides a new perspective for utilizing graph learning techniques to address problems related to urban building PV potential prediction.

Furthermore, this study introduces the concept of stratified modeling for building façades. By dividing the building façade vertically into multiple equal-height layers and predicting the average solar radiation intensity of each layer, we can more precisely depict the vertical lighting differences on building façades and reveal information overlooked by traditional holistic modeling methods. Experimental results show that the solar radiation intensity on building façades exhibits significant differences not only horizontally (between census tracts) but also vertically (building height). The stratified modeling approach can provide more comprehensive and detailed information from these two dimensions, offering important references for subsequent building PV potential assessment and planning.

In summary, the stratified modeling method based on the GGAT model proposed in this study provides a new solution for predicting building façade PV potential at the urban scale. This not only helps improve prediction accuracy and optimize the economic viability of PV systems but also provides an important basis for formulating more precise PV deployment strategies. Future research can further expand the GGAT model by incorporating more geospatial factors and exploring more efficient and intelligent building PV potential assessment methods. Meanwhile, applying this method to different cities and climatic conditions can verify its universality and robustness.

REFERENCES

- Ahmed, Asam, Tianshu Ge, Jinqing Peng, Wei-Cheng Yan, Boon Tuan Tee, and Siming You. 2022. "Assessment of the Renewable Energy Generation towards Net-Zero Energy Buildings: A Review." *Energy and Buildings* 256:111755.
- Brito, M. C., S. Freitas, S. Guimarães, C. Catita, and P. Redweik. 2017. "The Importance of Facades for the Solar PV Potential of a Mediterranean City Using LiDAR Data." *Renewable Energy* 111:85–94. doi: 10.1016/j.renene.2017.03.085.
- Brown, Darcy, Olubukola Tokede, Hong Xian Li, and David Edwards. 2024. "A Systematic Review of Barriers to Implementing Net Zero Energy Buildings in Australia." *Journal of Cleaner Production* 467:142910. doi: 10.1016/j.jclepro.2024.142910.
- Catita, C., P. Redweik, J. Pereira, and M. C. Brito. 2014. "Extending Solar Potential Analysis in Buildings to Vertical Facades." *Computers & Geosciences* 66:1–12. doi: 10.1016/j.cageo.2014.01.002.
- Fakhraian, Elham, Marc Alier Forment, Francesc Valls Dalmau, Alireza Nameni, and Maria Jose Casan Guerrero. 2021. "Determination of the Urban Rooftop Photovoltaic Potential: A State of the Art." *Energy Reports* 7:176–85.
- Gassar, Abdo Abdullah Ahmed, and Seung Hyun Cha. 2021. "Review of Geographic Information Systems-Based Rooftop Solar Photovoltaic Potential Estimation Approaches at Urban Scales." *Applied Energy* 291:116817.
- Liu, Jiang, Qifeng Wu, Zhipeng Lin, Huijie Shi, Shaoyang Wen, Qiaoyu Wu, Junxue Zhang, and Changhai Peng. 2023. "A Novel Approach for Assessing Rooftop-and-Facade Solar Photovoltaic Potential in Rural Areas Using Three-Dimensional (3D) Building Models Constructed with GIS." *Energy* 282:128920. doi: 10.1016/j.energy.2023.128920.
- Petter Jelle, Bjørn, Christer Breivik, and Hilde Drolsum Røkenes. 2012. "Building Integrated Photovoltaic Products: A State-of-the-Art Review and Future Research Opportunities." *Solar Energy Materials and Solar Cells* 100:69–96. doi: 10.1016/j.solmat.2011.12.016.
- Scarselli, Franco, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. 2008. "The Graph Neural Network Model." *IEEE Transactions on Neural Networks* 20(1):61–80.
- Singh, Rhythm, and Rangan Banerjee. 2015. "Estimation of Rooftop Solar Photovoltaic Potential of a City." *Solar Energy* 115:589–602. doi: 10.1016/j.solener.2015.03.016.
- Tian, Jia, Ryoza Ooka, and Doyun Lee. 2023. "Multi-Scale Solar Radiation and Photovoltaic Power Forecasting with Machine Learning Algorithms in Urban Environment: A State-of-the-Art Review." *Journal of Cleaner Production* 426:139040. doi: 10.1016/j.jclepro.2023.139040.
- Veličković, Petar, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. "Graph Attention Networks."
- Zhong, Teng, Zhixin Zhang, Min Chen, Kai Zhang, Zixuan Zhou, Rui Zhu, Yijie Wang, Guonian Lü, and Jinyue Yan. 2021. "A City-Scale Estimation of Rooftop Solar Photovoltaic Potential Based on Deep Learning." *Applied Energy* 298:117132. doi: 10.1016/j.apenergy.2021.117132.

REAL-TIME SENSOR-BASED CHARACTERIZATION OF RECYCLED COARSE AGGREGATES (RCA): ADVANCING SUSTAINABLE CONSTRUCTION THROUGH AUTOMATED QUALITY ASSESSMENT

Cheng Chang¹, Francesco Di Maio², and Peter Rem³

1) Ph.D. Candidate, Resource & Recycling, Department of Engineering Structures, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands. Email: c.chang-1@tudelft.nl

2) Ph.D., Asst. Prof., Resource & Recycling, Department of Engineering Structures, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands. Email: f.dimaio@tudelft.nl

3) Ph.D., Prof., Resource & Recycling, Department of Engineering Structures, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands. Email: p.c.rem@tudelft.nl

Abstract: The integration of recycled coarse aggregates (RCA) into construction projects has encountered industry resistance, primarily attributable to apprehensions about variable quality. This paper underscores the imperative need for reliable material quality assessments of RCA to ensure compliance with industry standards. Addressing this concern, we introduce a novel solution: a mobile, containerized sensor-based quality inspection system. This system is furnished with a 3D scanner Gocator, which, through optimized point cloud processing and streamlined segmentation algorithms, ensures rapid extrapolation of particle size distribution (PSD) from the RCA's surface point cloud data, producing outcomes closely aligned with conventional manual sieving techniques. Additionally, the application of laser-induced breakdown spectroscopy (LIBS) within this system has proven effective, consistently producing stable spectral data indicative of the material composition. The effectiveness of LIBS is further enhanced through the adoption of a cluster-based identification algorithm, which provides exceptional accuracy and precision in the spectral analysis. The system also includes conveyor belts capable of processing more than 100 tons of RCA per hour. This synergistic integration of technologies underpins a paradigm shift in RCA assessment, offering a scalable and adaptable model for enhancing the efficiency and reliability of End-of-Life material processing, aligning with global aspirations for sustainable infrastructural development.

Keywords: Recycled Coarse Aggregates (RCA), Quality inspection, Concrete recycling, Conveyor belt, 3D Scanner Gocator, Laser-induced breakdown spectroscopy (LIBS)

1. INTRODUCTION

The growing demand for construction materials, along with the need for sustainable waste management, has made the recycling of concrete from End-of-Life (EoL) infrastructure an increasingly practical and environmentally responsible solution (Kabirifar et al., 2020). This approach has gained significant attention in the construction industry, which is shifting towards more sustainable practices. This shift is particularly evident in the rising use of recycled coarse aggregates (RCA) derived from construction and demolition waste, a material that holds promise for reducing the construction industry's environmental footprint. However, despite the potential benefits, the widespread adoption of RCA presents significant challenges, primarily related to ensuring the quality and purity of RCA (Kim, 2022; H. Wu et al., 2023). These challenges become more complex when RCA is sourced from various types of dismantled infrastructures, which often contain contaminants such as mixed debris, organic materials, and other impurities (Alaejos Pilarand de Juan, 2013). If these contaminants are not effectively identified and managed, they can compromise the structural integrity, durability, and long-term reliability of new construction projects incorporating RCA (Poon & Chan, 2007; L. Wu et al., 2024). As such, addressing these concerns necessitates the development of standardized quality assessment protocols and stringent regulatory guidelines. Such measures are essential to ensure that RCA consistently meet or exceed the performance criteria of natural coarse aggregates. Without the establishment of these quality standards, the inconsistent quality of RCA may undermine its credibility as a reliable alternative to conventional materials, ultimately hindering its adoption on a larger scale.

Traditionally, RCA quality has been assessed using conventional methods that are labor-intensive, time-consuming, and heavily reliant on manual inspection (Marie & Mujalli, 2019; Tuan et al., 2022). While these approaches have served the industry in the past, their limitations—such as inefficiencies, inconsistencies, and susceptibility to human error—have become increasingly apparent as the sector seeks to enhance productivity and precision. With the growing complexity of demolition operations and the need for faster, more reliable results, there is an urgent demand for innovative quality inspection mechanisms that offer greater accuracy, efficiency, and reliability. Recent advancements in sensor technology have emerged as a transformative solution to these challenges (Cabral et al., 2023; Chang et al., 2025; Trotta et al., 2021). These technologies, through innovations in high-speed data acquisition and advanced processing algorithms, enable real-time, on-site characterization of RCA, offering precise assessments without compromising accuracy or reliability (Bonifazi et al., 2018; Nalon et al., 2022). By providing instant feedback during the production process, sensor-based systems facilitate more

informed decision-making and allow for immediate adjustments, reducing the likelihood of defects and material wastage. In addition, they optimize resource utilization, thus enhancing the overall efficiency of RCA production. This integration of sensor technology not only contributes to improved material quality but also aligns with broader environmental goals, reinforcing sustainable construction practices and enhancing the operational objectives of the construction industry.

This research investigates the potential of sensor-based quality inspection systems to revolutionize the analysis of RCA. These advanced systems facilitate granular analysis and effective detection of contaminants, offering a flexible and adaptive approach suitable for dynamic demolition environments. Through a series of rigorous experiments and field studies, this study evaluates the operational performance of these sensor-based systems, comparing their results with those obtained from conventional methods. The findings highlight the transformative potential of these technologies in improving the quality assurance process of RCA, emphasizing their role in promoting sustainable construction while maintaining stringent quality standards. Furthermore, this research demonstrates how these systems can be integrated into real-world scenarios, providing insights into their practical applicability, scalability, and long-term impact on the construction industry.

2. METHOD

The sensor-based quality inspection system is housed in a dedicated container, a design that serves several important purposes. This containerization simplifies the transport of the system to various demolition sites, improving mobility and ensuring more efficient use of resources. Additionally, it provides protection against adverse weather conditions, safeguarding the sensor equipment from environmental factors such as humidity, dust, or extreme temperatures that could compromise measurement accuracy. As a result, the integrity and precision of the data are preserved, ensuring consistent and reliable performance. Beyond protecting sensitive instruments, the enclosure also demonstrates the system's flexibility and adaptability in diverse field conditions. The modular nature of the container further facilitates easy deployment and potential upgrades, allowing for the integration of future technological advancements without major system overhauls. This design also supports rapid set-up and tear-down, reducing downtime between deployments and increasing operational efficiency across different demolition sites with varying environmental challenges.

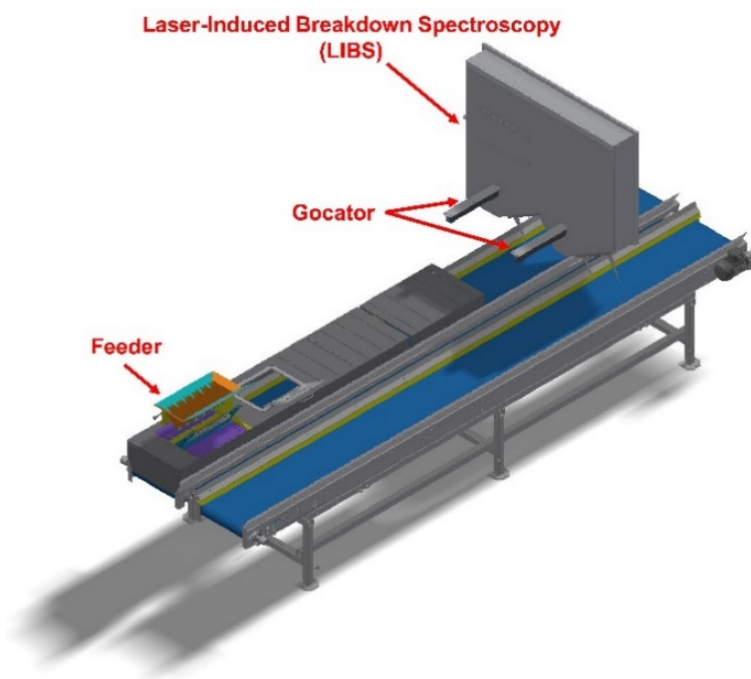


Figure 1 Sensor-based quality inspection system

The system (Figure 1) features two main sensors: the 3D scanner Gocator and Laser-Induced Breakdown Spectroscopy (LIBS). Both are positioned above the conveyor belt, ensuring optimal data collection without interference. The Gocator specializes in granulometric analysis of RCA, specifically measuring particle size distribution (PSD). This advanced device generates high-resolution 3D point cloud data, capturing details of the surface topology and granular distribution, which are critical for assessing the quality of RCA for reuse in construction applications. The Gocator's ability to produce accurate PSD measurements enhances the quality

assurance process, ensuring that only aggregates with appropriate size distributions are selected for further processing. LIBS, on the other hand, identifies contaminant compositions within the RCA, a crucial step in ensuring that the recycled material meets safety and regulatory standards. A notable feature of the system is the use of reflective mirrors, which allows simultaneous monitoring of RCA on two separate conveyor belts with just one Nd:TAG laser and a single spectrometer. This innovative design reduces operational costs and enhances the system's overall efficiency, offering both cost savings and increased throughput without compromising the quality of inspections.

The inspection process begins when RCA are introduced into the system via a feeder and deposited onto the conveyor belt, forming a triangular pile. This configuration ensures an even distribution of RCA from the center to the outer edges (Figure 2). It optimizes the Gocator's ability to inspect the surface and assess RCA properties accurately. As the RCA move along the conveyor, they are sequentially analyzed by both the Gocator and LIBS. The data from these inspections are instantly recorded in a computer system and uploaded to a secure cloud platform for long-term storage and retrieval, ensuring that inspection results are accessible for future reference and analysis. The conveyor belt runs at a constant speed of 0.529 m/s, enabling it to transport over 50 tons of RCA per hour, ensuring high throughput and efficiency.



Figure 2 Layered formation of RCA piles

3. RESULTS AND DISCUSSION

3.1 Particle Size Distribution (PSD)

(1) Resolution of Point Cloud Data

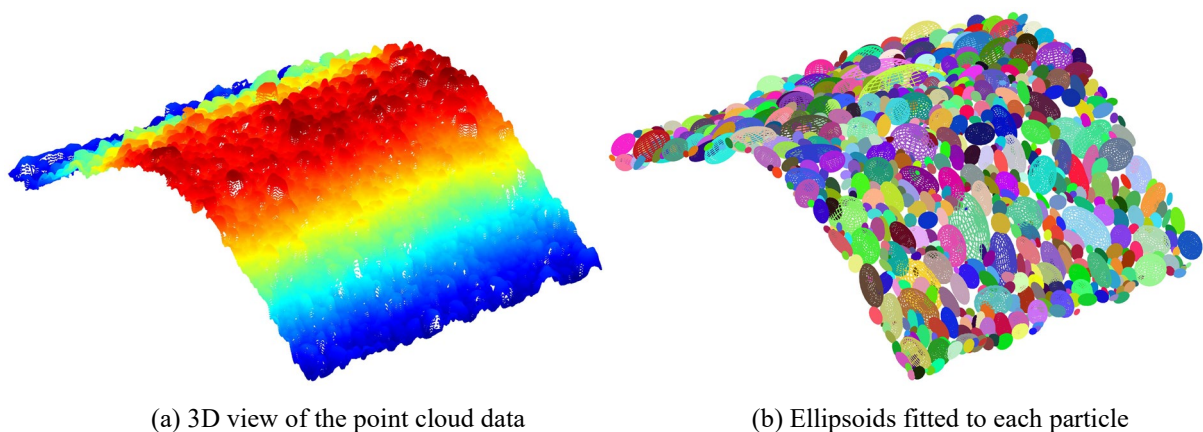


Figure 3 Point cloud processing

Given the uniform distribution pattern of constituents within the RCA piles, it is possible to extrapolate the PSD characteristics of the entire pile by analyzing only the outer surface layer. This approach provides an effective means to obtain a representative understanding of the overall PSD of the RCA piles. To facilitate this, a 3D scan of the RCA pile's external surface was conducted using the Gocator, which generated point cloud data (Figure 3(a)). The spatial resolution of the point cloud data plays a critical role in determining the accuracy and detail of the RCA particle analysis. Notably, the resolution varied depending on the spatial orientation, which has

important implications for the granulometric assessment.

Along the direction parallel to the conveyor belt's motion, two key factors influenced the point cloud resolution. First, the velocity of the conveyor belt, which was carefully controlled by adjusting the rotational speed of the motor drive. This speed was governed by two interrelated parameters: the output frequency and the number of poles. In this study, the motor drive was configured to operate at an output frequency of 50 Hz and a 4-pole design, resulting in a rotational speed of exactly 1500 rpm. Additionally, the gearbox ratio was set to 19, and the wheel perimeter was 402.116 mm, leading to a conveyor belt speed of 0.529 m/s. The second factor was the encoder resolution of the Gocator, which was calibrated to register 1024 ticks per full revolution. This calibration resulted in a point cloud resolution of 0.393 mm in the direction of the conveyor belt's motion.

However, when measurements were taken perpendicularly to the conveyor (across its width), the resolution varied between 0.375 mm and 1.100 mm. This variation is attributed to the scanner's field of view at different points. To maintain data consistency, the transverse resolution was standardized at 0.375 mm.

Vertical resolution, which refers to the height of the RCA piles, is another important consideration. The Gocator's internal mechanisms provide a resolution range from 0.092 mm to 0.488 mm. This variability demonstrates the scanner's flexibility, allowing it to adapt to different granulometric conditions, making it an effective tool for detailed analysis of RCA piles. For this study, the vertical resolution was set at 0.092 mm.

(2) Point Cloud Segmentation

The segmentation of the acquired point cloud data was performed using the FastScape algorithm (Braun & Willett, 2013). This algorithm, originally developed for terrain analysis, facilitates watershed segmentation (Steer et al., 2022). The point clouds were divided into distinct regions, each representing an individual RCA particle. To optimize processing speed, the algorithm was adapted to support parallel processing, enabling real-time operation.

To analyze the morphology of RCA particles, each segmented region was enclosed within an ellipsoidal envelope (Figure 3 (b)). This ellipsoidal shape approximated the form of the RCA particles and provided a means for detailed analysis of their structure, orientation, and morphology within the sample. A key feature of this representation was the selection of the second shortest axis as the primary gradation parameter. This axis was chosen because it effectively correlates with particle size and, by extension, the PSD.

To address the concern regarding the time-consuming nature of 3D scanning and point cloud segmentation, the proposed system incorporates several strategies to ensure real-time performance and practical processing speed, even in field conditions.

Efficient Data Acquisition and Preprocessing: The system employs a high-speed 3D scanning module capable of capturing point cloud data with minimal latency. Additionally, preprocessing steps, such as noise filtering and down-sampling, are implemented directly on the hardware to reduce the volume of data transferred for segmentation. These hardware-accelerated techniques significantly reduce processing overhead while preserving critical structural information.

Optimized Segmentation Algorithms: The segmentation algorithms focus on extracting essential features rather than performing exhaustive analyses, which helps to maintain a balance between accuracy and speed. Furthermore, the segmentation process is accelerated through the use of parallel processing and hardware acceleration, such as multi-threading and GPU-based computation, to handle large datasets efficiently.

By incorporating these strategies, the proposed system effectively addresses the typical bottlenecks in 3D scanning and point cloud segmentation. This ensures both practical usability and real-time operability in field applications, making it suitable for dynamic environments where speed and accuracy are crucial.

(3) PSD Calculation

The PSD derived from the analysis was clearly represented by the cumulative percentage retained graph (Figure 4). This graphical representation, based on calculations of ellipsoidal volume and the apparent density of the RCA, effectively illustrated the PSD characteristics of the RCA piles. Additionally, it provided valuable insights into the aggregate structure and its potential influence on concrete performance.

To validate the accuracy and reliability of this non-intrusive, surface-based method for determining PSD, a comparative analysis was conducted. The PSD results obtained from the 3D point cloud data were compared with those derived using the traditional, more invasive manual sieving method. This comparison aimed to evaluate the correlation and consistency between the surface PSD measurements and the overall PSD of the entire pile. Figure 4 shows the results of this comparison, where the cumulative percentage retained curves derived from the point cloud data are compared with those obtained from manual sieving, based on pilot scanning trials. To assess the precision of this method, the Root Mean Square Error (RMSE) was calculated between the predicted and manually obtained values. The RMSE values of 4.93%, 5.38%, and 4.27% across three experiments indicate minimal deviations, demonstrating the robustness and accuracy of the predictive methodology in estimating the PSD of RCA piles on a conveyor belt. Preliminary findings suggest a strong agreement between the two methods, indicating that the 3D scanning approach can reliably estimate the full PSD, significantly reducing manual effort.

and time.

In Figure 4, the predicted values of PSD are consistently higher than those obtained through manual sieving. This systematic discrepancy can be attributed to two primary factors. First, the ellipsoid approximation used to model particle shapes may overestimate the dimensions of irregularly shaped or angular particles. In the proposed method, particles are modeled as ellipsoids to approximate their shapes for computational efficiency. However, this simplification tends to overestimate particle sizes, especially for non-spherical particles. The ellipsoid fitting algorithm calculates the semi-principal axes of each particle by determining the minimum-volume bounding ellipsoid based on the point cloud data. The optimal fit is achieved by minimizing the sum of squared distances from the points to the ellipsoid's surface, subject to constraints that ensure the surface represents an ellipsoid. This process prioritizes enclosing the particle's volume, which may result in slightly inflated size predictions compared to physical sieving, where particles are forced through sieve meshes. Second, the resolution of the 3D point cloud data can affect accuracy. Lower resolutions may smooth out fine details, further contributing to an overestimation of particle sizes. In this study, the resolution of the 3D scanner varied in each direction, which balances computational load and measurement accuracy but may still introduce minor deviations.

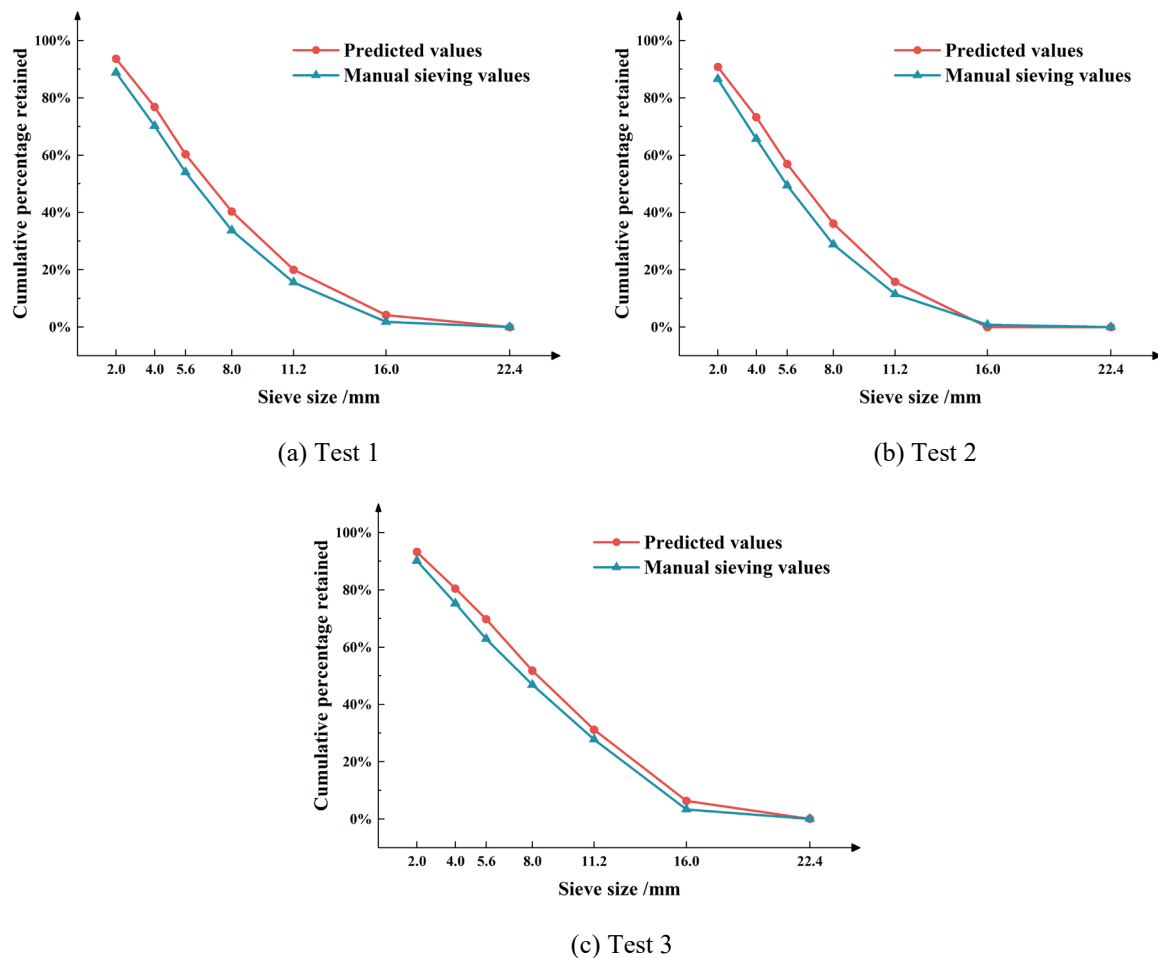


Figure 4 Cumulative percentage retained graphs

The accuracy of PSD estimation from the 3D point cloud data is influenced by several factors, including the speed of the conveyor belt. The spatial resolution of the point cloud, which directly impacts the accuracy of the PSD calculation, can be affected by the speed at which the RCA move along the conveyor. In this study, the conveyor belt was set to a constant speed of 0.529 m/s, which provided a balance between high-resolution data capture and real-time operation. However, variations in belt speed can influence the resolution of the 3D scan data. Specifically, a faster conveyor belt speed may reduce the time available for scanning, leading to a lower point cloud resolution, which could negatively affect the accuracy of the PSD calculation. Conversely, reducing the belt speed can enhance the resolution of the point cloud data, thereby improving the precision of the PSD estimation. Further studies could explore the potential for increasing conveyor belt speed while maintaining accuracy by optimizing the synchronization between the scanner's resolution and conveyor motion. Such optimization may

allow for higher throughput without compromising PSD estimation quality.

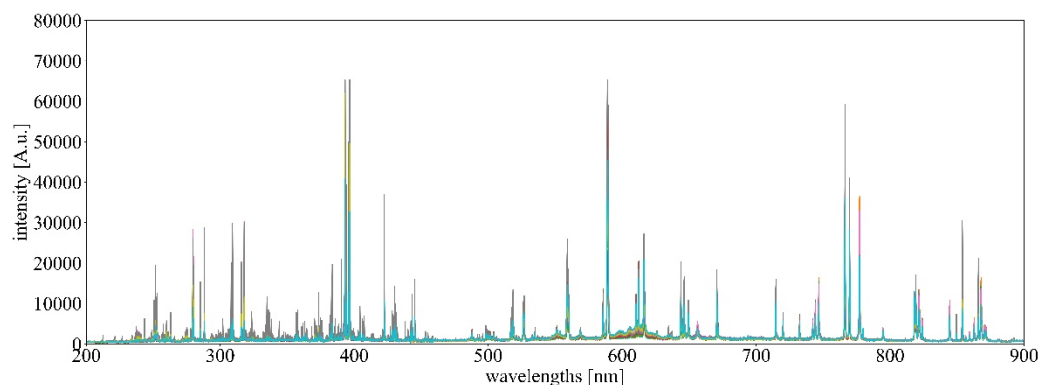
3.2 Contaminant Detection

To thoroughly evaluate the performance of LIBS in capturing spectra at a conveyor belt speed of 0.529 m/s, and to assess the potential influence of conveyor belt movement on the efficacy of the reflective mirrors within the system, a systematic study was conducted. In this study, spectral measurements were obtained for a variety of materials under two distinct conditions: when the conveyor belt was moving and when it was stationary.

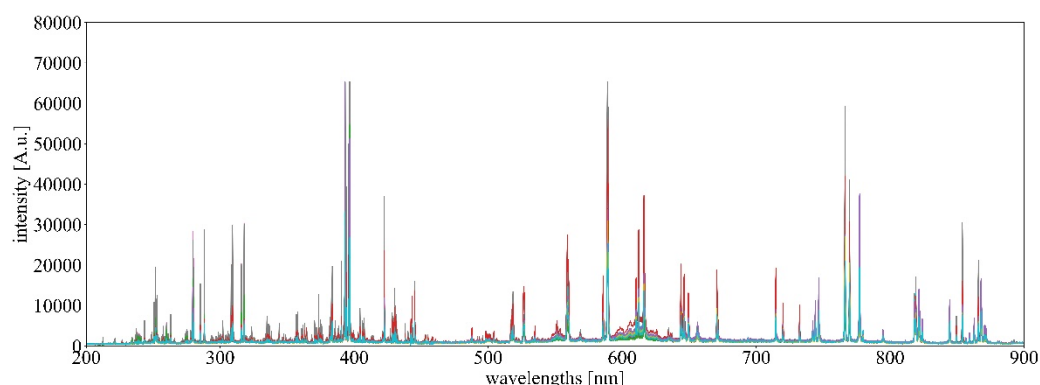
Figure 5 presents a comparison of LIBS spectra for brick materials under two conveyor belt conditions: (a) operational state (belt in motion) and (b) stationary state (belt at rest). The spectra cover a broad wavelength range of 200–900 nm, providing insights into the elemental composition of the brick samples. Several characteristic emission peaks can be identified in both conditions, such as those at approximately 280 nm, 393–397 nm, and 589 nm. These peaks correspond to the primary chemical elements found in brick and serve as markers for compositional analysis.

In the operational state (Figure 5(a)): The spectra exhibit slight variations in intensity, particularly in the 200–400 nm region, where weaker peaks experience minor shifts or noise. This phenomenon likely results from movement-induced disturbances, such as changes in the laser focus or micro-vibrations of the brick surface on the moving conveyor. However, the positions of the primary emission lines remain consistent, and their relative intensities are preserved. This suggests that the system is capable of compensating for dynamic conditions, ensuring that the material composition is reliably detected even during high-speed operations.

In the stationary state (Figure 5(b)): The spectra show smoother and more uniform peaks with reduced noise levels. This is due to the absence of mechanical disturbances, allowing the laser-material interaction to remain stable. The stationary condition serves as a baseline, providing a reference for assessing the robustness of the LIBS system under real-time operational conditions.



(a) Conveyor belt in the operational state



(b) Conveyor belt in the stationary state

Figure 5 Comparison of spectra of Brick in the operation and stationary states of the conveyor belt

A detailed comparison of the two spectra highlights that the primary spectral features, such as peak positions and relative intensities, remain nearly identical between the two states. This consistency demonstrates the robustness of the LIBS system in maintaining accuracy and reliability regardless of conveyor motion. Although slight variations in peak intensity are observed in the operational state, they are negligible in terms of analytical performance. Such findings validate the system's suitability for industrial applications, such as real-time quality

monitoring and contaminant detection in recycling processes, where the conveyor belt is in constant motion.

The influence of conveyor belt speed on the spectral data collected by the LIBS sensor was also assessed. While the conveyor belt was in motion (0.529 m/s), the stability and reliability of the spectral readings remained high, as shown in Figure 5. This suggests that even at relatively high speeds, the system can maintain accurate contaminant detection. Future experiments may examine the feasibility of increasing conveyor belt speeds while ensuring the spectral data's integrity, exploring whether faster belt speeds can be accommodated without sacrificing analysis quality.

Following the spectral data collection, a classification analysis was performed using the Cluster-based Identification Algorithm (Chang et al., 2022). This algorithm, recognized for its ability to group data based on inherent similarities, was effectively applied to identify patterns in the spectral data. This algorithm integrates chemometric methods, including Principal Component Analysis (PCA) and Chi-square distribution, to classify single-shot spectral data efficiently. The approach follows these key steps:

Dimensionality Reduction via PCA: Initially, the high-dimensional spectral data undergoes dimensionality reduction through PCA, which helps reduce data redundancy and noise, retaining only the most significant features of the spectra.

Transformation to a New Coordinate System: The PCA transforms the original spectral data into a new coordinate system where the components become independent normal distributions. This transformation simplifies the classification process by focusing on the principal components.

Chi-square Distribution for Classification: After standardizing the data using z-scores, the transformed spectral data is subjected to a chi-square test to assess whether the data fits the distribution of the material in question. A high chi-square value indicates a higher likelihood that the spectrum belongs to the identified material.

The model's performance was rigorously evaluated to ensure its suitability for real-world applications. Remarkably, the model achieved strong performance metrics: an accuracy rate of 0.92, a weighted average precision of 0.93, a weighted average recall of 0.93, and a weighted average F1-score of 0.94 on the validation dataset.

These performance metrics underscore the effectiveness of LIBS as a powerful analytical tool, particularly in the context of recycled concrete aggregates in industrial settings. The high accuracy and precision demonstrate that LIBS not only enables rapid analysis but also maintains high-quality results. These capabilities have the potential to transform quality assessment processes in the industry, improving both efficiency and reliability in recycled concrete aggregate processing. By leveraging such advanced analytical techniques, industries can optimize the use of recycled materials while ensuring stringent quality control.

3.3 Comparison with Other Methods

To assess the effectiveness of the proposed system, we compared its performance with existing techniques such as computer vision-based methods for particle detection and size estimation. Computer vision algorithms, such as traditional edge detection and more recent convolutional neural networks (CNNs), have been widely used for particle analysis in industrial settings. However, these methods often face challenges related to lighting conditions, occlusions, and texture variations on the surface of materials.

In comparison, the proposed method using the 3D scanning and LIBS system offers several advantages. Firstly, the 3D scanning technique provides a more accurate reconstruction of particle surfaces, which reduces errors caused by shadows and lighting variations. The LIBS system, on the other hand, allows for real-time elemental analysis, which enhances the detection of contaminants that may not be visible to traditional vision-based methods.

While both methods achieve similar accuracy in particle size detection, the proposed 3D and LIBS-based method stands out in terms of robustness and ability to detect hidden contaminants, making it more suitable for complex industrial environments. Further comparisons with more advanced machine learning-based image processing techniques, such as deep learning models, could provide additional insights into the potential advantages and limitations of our method.

4. CONCLUSIONS

The global focus on sustainable infrastructure development has placed increasing importance on the efficient use of EoL materials, especially in the construction and demolition sectors. As demand for recycled materials grows, ensuring that these materials meet strict quality standards for safe and effective reuse is becoming more urgent. This research evaluates the potential of advanced technologies to assess the quality of RCA sourced from demolition sites. It emphasizes the need for real-time, on-site inspection systems that provide immediate feedback during construction processes.

A sensor-based quality inspection system, housed in a durable container, exemplifies the integration of mobility, resilience, and accuracy—key features for field applications. This system, with its high-speed data acquisition and processing capabilities, not only ensures real-time performance but also offers scalability and flexibility. These qualities make it suitable for diverse demolition environments where quick decision-making is

essential. Furthermore, automating the quality assessment process reduces human error, enhancing the consistency and reliability of evaluations.

The detailed granulometric analysis enabled by the Gocator 3D sensor, along with contaminant detection through LIBS, provides a comprehensive understanding of RCA quality. This dual approach addresses key concerns, including PSD and contaminant content—factors critical to the suitability of RCA for reuse in construction. By evaluating these parameters, the system ensures compliance with industry standards while optimizing RCA for applications such as concrete mixes or road construction. Additionally, the system's real-time data capture, integrated with cloud-based storage platforms, represents a significant advancement in data management and accessibility. These features streamline decision-making, allowing quality control teams to remotely monitor and adjust parameters to consistently meet quality benchmarks.

The ability to monitor two conveyor belts simultaneously with a single laser and spectrometer demonstrates the system's efficiency. This feature enhances cost-effectiveness by reducing the need for additional hardware and operators, significantly lowering operational costs without sacrificing precision. This solution, with its reduced complexity, presents a compelling case for industry-wide adoption.

In conclusion, this research highlights the potential of integrating advanced technologies into the processing of RCA. By combining innovation with precision, the study promotes a future where the quality of recycled materials is established through objective, real-time data. The findings demonstrate that, with the right technological tools, industries can achieve a balance between operational efficiency and material quality. These advancements have practical implications for reducing waste, lowering carbon footprints, and supporting sustainability goals in the construction sector. As the world shifts toward sustainability, studies like this lay the foundation for a new approach to quality and efficiency in material reuse. This will be crucial in the transition to a circular economy, where materials retain value throughout their lifecycle.

ACKNOWLEDGMENTS

Fundings: This research was supported by the European Union's Horizon funded Projects "Innovative Circular Economy Based solutions demonstrating the Efficient Recovery of valuable material resources from the Generation of representative end-of-life building materials" (ICEBERG, grant agreement No. 869336) and "Digital, autonomous, Intelligent and Synchronous system for Continuous identification, Optimization and Value Extraction of Resources from the end-of-use built environment" (DISCOVER, grant agreement No. 723611).

REFERENCES

- Alaejos Pilar and de Juan, M. S. and R. J. and D. R. and V. I. (2013). Quality Assurance of Recycled Aggregates. In E. Vázquez (Ed.), *Progress of Recycling in the Built Environment: Final report of the RILEM Technical Committee 217-PRE* (pp. 229–273). Springer Netherlands.
- Bonifazi, G., Palmieri, R., & Serranti, S. (2018). Evaluation of attached mortar on recycled concrete aggregates by hyperspectral imaging. *Construction and Building Materials*, 169, 835–842.
- Braun, J., & Willett, S. D. (2013). A very efficient O(n), implicit and parallel method to solve the stream power equation governing fluvial incision and landscape evolution. *Geomorphology*, 180–181, 170–179.
- Cabral, J. S., Menegatti, C. R., & Nicolodelli, G. (2023). Laser-induced breakdown spectroscopy in cementitious materials: A chronological review of cement and concrete from the last 20 years. *TrAC Trends in Analytical Chemistry*, 160, 116948.
- Chang, C., Di Maio, F., Bheemireddy, R., Posthoorn, P., Gebremariam, A. T., & Rem, P. (2025). Rapid quality control for recycled coarse aggregates (RCA) streams: Multi-sensor integration for advanced contaminant detection. *Computers in Industry*, 164, 104196.
- Chang, C., Maio, F. Di, Rem, P., Gebremariam, A. T., Mehari, F., & Xia, H. (2022). Cluster-based identification algorithm for in-line recycled concrete aggregates characterization using Laser-Induced Breakdown Spectroscopy (LIBS). *Resources, Conservation and Recycling*, 185, 106507.
- Kabirifar, K., Mojtahedi, M., Wang, C., & Tam, V. W. Y. (2020). Construction and demolition waste management contributing factors coupled with reduce, reuse, and recycle strategies for effective waste management: A review. *Journal of Cleaner Production*, 263, 121265.
- Kim, J. (2022). Influence of quality of recycled aggregates on the mechanical properties of recycled aggregate concretes: An overview. *Construction and Building Materials*, 328, 127071.
- Marie, I., & Mujalli, R. (2019). Effect of design properties of parent concrete on the morphological properties of recycled concrete aggregates. *Engineering Science and Technology, an International Journal*, 22(1), 334–345.
- Nalon, G. H., Santos, R. F., Lima, G. E. S. de, Andrade, I. K. R., Pedroti, L. G., Ribeiro, J. C. L., & Franco de Carvalho, J. M. (2022). Recycling waste materials to produce self-sensing concretes for smart and sustainable structures: A review. *Construction and Building Materials*, 325, 126658.
- Poon, C. S., & Chan, D. (2007). Effects of contaminants on the properties of concrete paving blocks prepared with

- recycled concrete aggregates. *Construction and Building Materials*, 21(1), 164–175.
- Steer, P., Guerit, L., Lague, D., Crave, A., & Gourdon, A. (2022). Size, shape and orientation matter: fast and semi-automatic measurement of grain geometries from 3D point clouds. *Earth Surface Dynamics*, 10(6), 1211–1232.
- Trotta, O., Bonifazi, G., Capobianco, G., & Serranti, S. (2021). Recycling-oriented characterization of post-earthquake building waste by different sensing techniques. *Journal of Imaging*, 7(9).
- Tuan, N. M., Kim, Y., Lee, J.-Y., & Chin, S. (2022). Automatic Stereo Vision-Based Inspection System for Particle Shape Analysis of Coarse Aggregates. *Journal of Computing in Civil Engineering*, 36(2).
- Wu, H., Liang, C., Zhang, Z., Yao, P., Wang, C., & Ma, Z. (2023). Utilizing heat treatment for making low-quality recycled aggregate into enhanced recycled aggregate, recycled cement and their fully recycled concrete. *Construction and Building Materials*, 394, 132126.
- Wu, L., Sun, Z., & Cao, Y. (2024). Modification of recycled aggregate and conservation and application of recycled aggregate concrete: A review. *Construction and Building Materials*, 431, 136567.

EXPLORING RIVERINE LITTER DETECTION BY DEVELOPING COMPREHENSIVE DATASET AND DEEP LEARNING

Xiaohan Xu¹, Cheng Zhang^{2*}, and Yunfei Xia³, Xiaohui Zhu⁴, Peter Burgess⁵

1) Ph. D. student, Department of Civil Engineering, Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu Province, China. Email: Xiaohan.Xu22@student.xjtlu.edu.cn

2) Ph.D., Assoc. Prof., Department of Civil Engineering, Design School Intelligent Build and Environment Research Center, Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu Province, China. Email: Cheng.Zhang@xjtlu.edu.cn

3) Ph. D. student, Department of Civil Engineering, University of Liverpool, Suzhou, Jiangsu Province, China. Email: Yunfei.Xia22@alumni.xjtlu.edu.cn

4) Ph.D., Assoc. Prof., Department of Computing, Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu Province, China. Email: Xiaohui.Zhu@xjtlu.edu.cn

5) Ph.D., Prof., School of Environmental Sciences, University of Liverpool, Liverpool, United Kingdom. Email: Peter.Burgess@liverpool.ac.uk

Abstract: With the rapid growth of the economy, the problem of plastic litter in rivers is becoming increasingly severe, particularly in key river basins such as the Taihu Basin. Plastic litter not only disrupts aquatic ecosystems but also poses a threat to human health and regional economic development. Therefore, it is imperative to take effective measures to reduce plastic litter in rivers in order to protect the environment and promote sustainable development. This study proposes an efficient river plastic litter detection method by combining unmanned equipment and deep learning. A dataset comprising 1,347 RGB images of river trash, captured under diverse environmental conditions, was developed to offer a wealth of diversity for model training. YOLOv10-N is employed for object detection and an mAP@0.5 of 94.4% on the dataset is achieved. The research results highlight the potential of applying deep learning in environmental monitoring. In addition, the contribution of this study's dataset provides valuable resources for future model training, with diverse types of images enhancing the model's generalization capabilities and offering possibilities for more effective trash collection.

Keywords: Computer vision, UAV remote sensing, Plastics litter dataset, Deep learning, Object detection, YOLOv10-N.

1. INTRODUCTION

As the economy continues to develop and grow rapidly, environmental issues have become increasingly prominent and are now a widespread concern globally. According to a World Bank report (Kaza et al., 2018), it is estimated that by 2050, 3.4 billion tons of waste will be generated per year, significantly impacting the natural environment. Plastic pollution has emerged as a significant concern for both public health and environmental sustainability. In particular, the presence of plastics in aquatic ecosystems has garnered considerable attention due to the economic losses and environmental risks it poses, as well as its potential effects on human health (González-Fernández et al., 2021; Van Calcar & Van Emmerik, 2020). The United Nations Environment Programme (UNEP) reports that marine debris is distributed as follows: 15% is found floating on the ocean surface, 15% is suspended within the water column, and 70% settles on the seabed (Jambeck et al., 2015). Rivers are identified as the primary contributors to marine plastic litter, with annual estimates of plastic discharged from rivers to the ocean ranging from 0.8 to 2.7 million tons (Meijer et al., 2021). Despite their vital role in the transport and accumulation of marine debris, research on river ecosystems remains limited (Jia et al., 2023). As litter in lakes, rivers, and oceans continues to rise, removing debris from the water's surface has become a crucial effort to enhance the ecological environment (Compa et al., 2019). The water quality of rivers can be partially assessed through the types and quantities of floating debris in the waterways, serving as a useful indicator for measuring water quality (Lin et al., 2021). However, conventional manual cleaning methods are often costly and inefficient. With advancements in unmanned equipment technology, utilizing such equipment to assist or even replace manual water purification has emerged as a viable solution (Kamarudin et al., 2021).

In recent years, convolutional neural networks (CNNs) have established themselves as the leading techniques in computer vision, largely owing to their remarkable accuracy and highly automated feature extraction capabilities (Ham et al., 2018; Dhillon and Verma, 2020; Esteva et al., 2021). In terms of object detection, YOLO accounts for more than 34% of the models used for object detection, followed by Faster R-CNN with 25% (Wu et al., 2023). Maharjan et al. (2022) found that the pre-trained YOLOv5s model was most useful for unclassified plastic detection in rivers from drone images. Van Lieshout et al. (2020) employed a visual method to estimate the flux of large plastic debris in rivers, identifying plastics within the dataset without categorizing them, with an accuracy of 68.7%. However, the application of these advanced techniques is limited by the lack of targeted and high-quality datasets, making model training and performance optimization challenging. In view of this, the objective of the study is to improve the performance of different types of river litter detection by developing a comprehensive self-made dataset combined with the latest deep learning

techniques. This dataset not only covers multiple types of river litter, but also takes into account different environmental conditions and litter states, providing a rich and accurate data basis for model training. The main contributions of this study include:

- Using RGB images of floating plastic litter in rivers acquired by drones and other equipment, a comprehensive multi-category dataset containing 1,347 images was developed (Zhang et al., 2024), and an object detection model, YOLOv10-N, was applied to achieve 94.4% mAP@0.5 on this dataset.

2. METHOD

2.1 Object Detection Algorithm

This study introduces a method for detecting riverine debris through the analysis of Unmanned Aerial Vehicle (UAV) imagery. The approach employs the YOLOv10 algorithm to facilitate the automatic identification and classification of self-classified plastic waste within river environments. The research workflow is shown in Figure 1. Five types of plastic litter are collected in the campus litter station, including plastic bottles, plastic bags, cans, plastic boxes, and plastic cups. The images are taken in the river using drones and other equipment, and the obtained images are classified and annotated. Then use the YOLOv10 algorithm as the base model. YOLOv10 achieves training and inference without non-maximum suppression (NMS) through consistent dual allocation, which significantly reduces inference latency and improves prediction efficiency (Wang et al., 2024). In addition, YOLOv10 also adopts an overall efficiency-accuracy driven model design strategy to optimize the various components of the model architecture to minimize computational overhead while improving performance (Sapkota et al., 2024). In term of performance, YOLOv10 outperforms previous versions and other state-of-the-art models on the COCO dataset (Wang et al., 2024). After the model training is completed, the prediction accuracy of the model is measured by calculating Intersection over Union (IoU) and mean Average Precision (mAP). This enables the model to predict the litter in the river and output the category and bounding box of each detected object. In this way, the method effectively detect and classify litter in the river, providing an efficient technical support for river litter management and environmental protection.

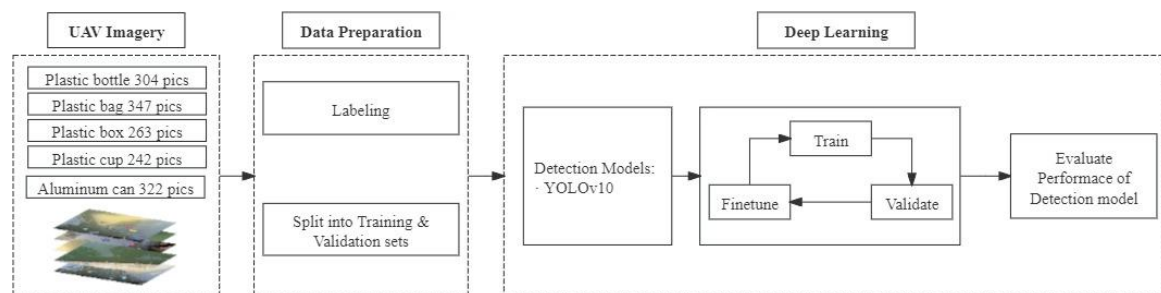


Figure 1. The research workflow for detecting plastic litters using the YOLO v10

2.2 Data Collection

As shown in Figure 2, data are collected from a representative water town in Suzhou, China. Aerial photography is performed using a DJI Phantom 4, which is equipped with a camera capable of 4K resolution. The drone is operated at a low altitude, ranging from 5 to 10 meters, to capture high-definition imagery. The camera is manually controlled and records images in the RGB spectrum during the daylight hours from 10:00 am to 5:00 pm. The native resolution of each photograph is an impressive 5472×3648 pixels. To evaluate the extent of plastic litter in the rivers, photographic surveys are also carried out at a closer altitude of 2 to 4 meters above the water's surface using an iPhone 13. Due to the insufficient number of images of floating plastic collected in field data. Diverse materials are utilized in creating a customized dataset for documenting riverine plastic waste. The materials include: (i) Fishing lines of two different diameters, 0.165 mm and 0.8 mm, which are employed to confine plastic debris within specific areas. (ii) Single-sided adhesive tape is used to secure the fishing line's other end to a hand or a vertical marker, thereby simplifying the manipulation of the waste. (iii) A variety of plastic items, including bottles, bags, cans, boxes, and cups, are included in the dataset. These items vary in size and color to represent the range of plastic litter found in the river. Specifically, the dataset encompasses various types of floating debris, ensuring a high level of diversity and representativeness among the samples. This diversity not only reflects the different types of waste present under various environmental conditions but also increases the variability during the model training process. As a result, the model's ability to generalize in complex scenarios is enhanced, thereby effectively improving detection precision and accuracy. Furthermore, the comprehensiveness of the dataset impacts the model's performance in real-world applications, as a rich and

diverse dataset can better capture the various features and situations encountered in reality. Consequently, this comprehensiveness enhances the model's performance and further validates the reliability and effectiveness of the research findings, providing a solid foundation for future applications.



Figure 2. Location of study sites (Background map: OpenStreetMap, 2024)

2.3 Labeling

Tzutalin is credited with the development of LabelImg, a software tool that was made publicly available in 2018 (Tabassum et al. 2020). When integrated with the YOLO (You Only Look Once) object detection algorithm, LabelImg forms a formidable duo for the creation and training of high-performance object detection models, as highlighted by Varnima & Ramachandran (2020). LabelImg is emerged as a prevalent choice for image annotation in the field of object detection, as noted by Pande et al. (2022). The repository for LabelImg, which includes its source code and additional documentation, is hosted on GitHub (HumanSignal n.d.). As depicted in Figure 3, within the customized dataset crafted for river plastic litter, each item is assigned a distinct descriptor. The precise spatial orientation of each object within the photographic frame is delineated through an annotation file that specifies the object's coordinates and associated labels. This methodical annotation is crucial for the accurate training and evaluation of object detection models.

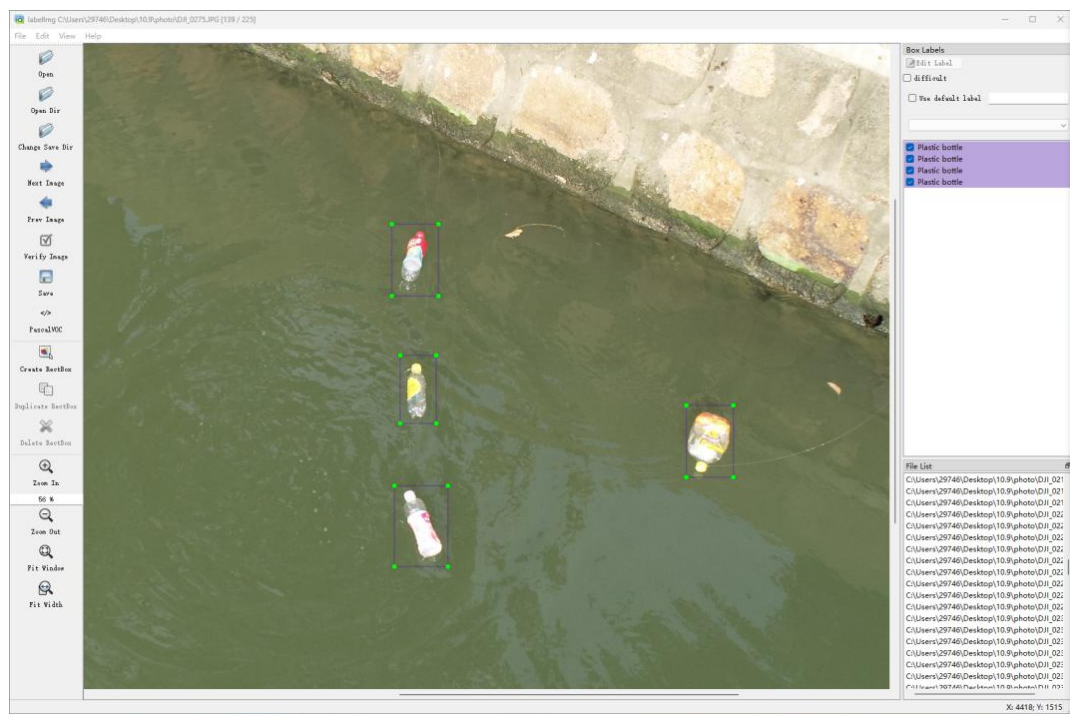


Figure 3. Image labeling steps for plastic bottles using LabelImg

3. RESULTS

3.1 Experimental Environment and Evaluation Metrics

The data processing and analysis for this project were completed on a HP 8A17 (LPC Controller - 5172) motherboard system equipped with an Intel 12th-generation Core i7-12700H 14-core processor, 16GB of memory (4800MHz), a 1TB Western Digital WD PC SN810 SDPCPNRY-1T00-1006 solid-state drive, and an NVIDIA GeForce RTX 3070 Ti Laptop GPU (8192 MB video memory). The operating environment is Windows 11 Home Edition (64-bit). The computational framework encompasses PyTorch version 2.1.1, Python version 3.9, and CUDA version 11.8. For the sake of maintaining uniformity and enabling the direct comparison of outcomes across different experiments, software milieu adhere to the specifications detailed in Table 1 for all subsequent trials.

Table 1. Training settings

Parameter	Value	Parameter	Value
Learning Rate	0.01	Weight Decay	0.0005
Batch Size	16	Momentum	0.937
Image Size	416x416	Epoch	200
Workers	8	Optimizer	Auto

In the domain of object detection, Intersection over Union (IoU) serves as a pivotal metric for assessing the accuracy of the detection model. IoU quantifies the extent to which the predicted bounding box aligns with the ground truth bounding box. The IoU is computed, as delineated in Equation (1), by taking the ratio of the area where the predicted box and the actual box overlap to the total area covered by both boxes combined (Maharjan et al., 2022). This calculation provides a measure of the model's precision in detecting objects within an image.

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}} \quad (1)$$

The mean Average Precision (mAP), as described by Equation (2), is a critical metric for assessing the overall effectiveness of an object detection model. This metric calculates the average precision across various confidence thresholds, offering a holistic view of the model's performance (Jackovljevic et al., 2020). Furthermore, the model's performance is also gauged by two fundamental metrics: Precision, as defined by Equation (3), and Recall, outlined in Equation (4). The evaluation index is based on four key statistics: True Negative (TN), True Positive (TP), False Positive (FP) and False Negative (FN). These metrics are essential for understanding the model's ability to accurately identify objects without producing excessive false positives or missing true positives.

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^N \text{AP}_i \quad (2)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (3)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (4)$$

3.2 Experimental Results

The experimental results presented in Table 2 demonstrate the remarkable detection performance of YOLOv10-N in identifying various types of riverine debris. For instance, plastic bottles and aluminum cans exhibit high precision and recall rates above 89%, suggesting that YOLOv10 is highly reliable in detecting these items. Notably, plastic boxes boast a near-perfect precision of 98%, with a commendable recall rate of 93.8%, indicating exceptional detection capabilities. As shown in figure 4a, the mAP at IoU thresholds of 0.5 and 0.5:0.95 further underscores the model's robustness, with mAP values exceeding 94.4% for all classes. This indicates that YOLOv10 not only detects debris with high accuracy but also does so consistently across different IoU thresholds.

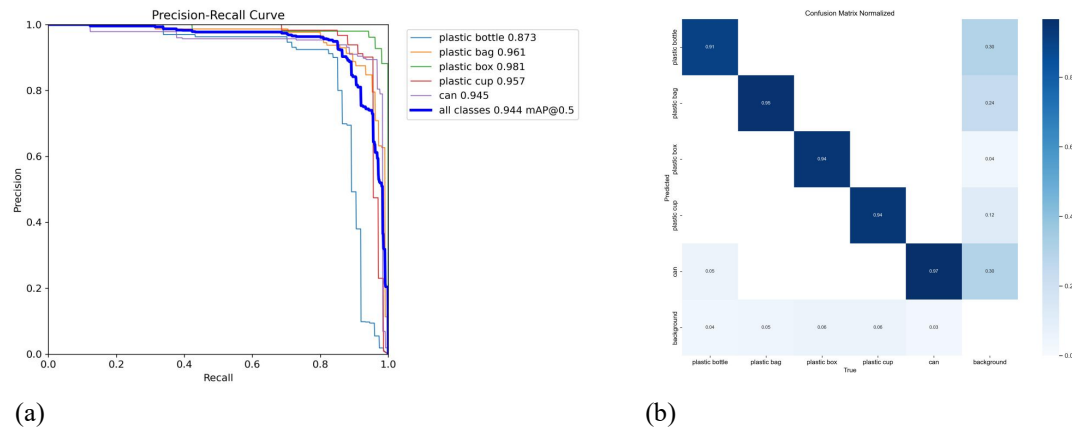


Figure 4. (a) Performance Evaluation of Object Detection Models
(b) Confusion Matrix Normalized for plastic litter Classification

Table 2. Experimental results

Class	P /%	R /%	CMN/%	mAP @0.5/%
Plastic bottle	89.9	85.1	91	87.3
Plastic bag	90.9	89.5	95	96.1
Plastic box	98	93.8	94	98.1
Plastic cup	93.7	89.2	94	95.7
Aluminum can	89.4	93.4	97	94.5
all	92.4	90.2	94.2	94.4

For Confusion Matrix Normalized (CMN), the diagonal elements are as close to 1 as possible, and the off-diagonal elements are as close to 0 as possible. Such a matrix indicates that the model's predictions on all categories are very accurate, that is, the model has high precision and high recall. As shown in Figure 5, the classification of Aluminum cans performed best with an accuracy of 97%, which shows that the model is very effective in identifying cans. The model's performance on most categories is satisfactory, but there is room for improvement in distinguishing plastic bottles. The accuracy and robustness of the model can be further improved by increasing the training data, optimizing feature extraction, or adjusting the model parameters.

The performance of the YOLOv10-N model was evaluated in the task of identifying river litter in complex scenarios. As shown in Figure 4b, when high intersection-over-union (IOU) conditions and target confidence thresholds (not less than 0.6) are set, YOLOv10-N demonstrates its excellent processing capabilities, especially in the case of target overlap and occlusion. Encouragingly, the YOLOv10-N model has demonstrated the capability to accurately detect plastic bottles in imagery captured within natural river settings, including areas with dense vegetation. The model exhibits a confidence threshold for object detection that averages approximately 66%. In addition, when the image contains multiple categories of litter and presents various morphological changes, YOLOv10-N can still effectively detect targets, showing its excellent adaptability.



Figure 5. The datasets image samples about successful prediction of YOLOv10-N

4. DISCUSSION

4.1 Analysis of Datasets

This research endeavors to create a specialized classification dataset specifically tailored for the identification of riverine plastic waste, a pressing environmental issue that necessitates effective monitoring and management strategies. Recognizing the urgency of addressing plastic litter in aquatic ecosystems, the dataset has been validated within the YOLOv10-N object detection algorithm, achieving both the high efficiency and precision of waste recognition processes. The dataset's quality and diversity are paramount, providing the algorithm with robust training and validation information essential for its performance. The dataset encompasses a wide variety of floating plastic waste typically found in riverine environments, including items such as plastic bottles, plastic bags, plastic boxes, plastic cups, and aluminum cans. Each category is represented with multiple examples, reflecting the real-world conditions where these items are often encountered. During the prediction phase, although the overall performance of the model was satisfactory, some aluminum cans were misidentified or missed due to their similarity to the background, indicating that the algorithm still needs to be further optimized in distinguishing objects similar to the background. Notably, the dataset includes various shapes of crushed cans and a range of forms for plastic boxes and cups, such as open cups and specialized containers like milk tea cups. This attention to detail in capturing different variations of plastic waste contributes to the dataset's representativeness and practicality, allowing for more accurate modeling of the complexities associated with plastic waste in aquatic settings. The diversity embedded within the dataset not only enhances its robustness but also ensures the model's wide applicability in real-world scenarios. By training the YOLOv10-N algorithm on a comprehensive set of plastic waste examples, the model is better equipped to handle the multifaceted challenges posed by varied environmental conditions. The scale and quality of the dataset play a crucial role in the effective training of deep learning models, as highlighted by Xiao et al. (2021), emphasizing the importance of high-quality inputs in achieving reliable outputs.

4.2 Limitation and future work

The research conducted in this study reveals several limitations that warrant further discussion and exploration. A primary challenge has been the difficulty in gathering comprehensive data from riverbanks that are densely populated with vegetation. This limitation has led to a focused study on a single river channel, which may restrict the broader applicability of the findings. To enhance the generalizability of this research, future studies should consider expanding data collection to include a variety of river environments. Such diversity in the study locations would allow researchers to draw more robust conclusions that can be applied to different ecological contexts. Additionally, the current focus has predominantly been on floating objects within the river channel. However, it is crucial to expand this focus to include aquatic plants, such as cyanobacteria and water hyacinth, which play significant roles in the river ecosystem. These organisms not only contribute to the ecological balance but also influence water quality and habitat dynamics. By incorporating the monitoring of these aquatic plants, future research can provide a more comprehensive understanding of the river's health and the factors affecting its biodiversity.

Moreover, the limitations associated with drone flight capabilities and monitoring in densely vegetated environments suggest the need for innovative technological solutions. Future studies should explore the integration of mast mounted cameras with drones to enhance data collection efficiency and coverage. Mast mounted cameras can operate effectively in shallow water areas along riverbanks, enabling researchers to access regions that drones alone might not effectively survey. Conversely, drones can cover a broader area from the air, providing valuable aerial perspectives and data. The collaborative operation of mast mounted cameras and drones presents an exciting opportunity for comprehensive monitoring of river environments. By leveraging the strengths of both technologies, researchers can achieve a more thorough and nuanced understanding of river ecosystems. This combined approach could facilitate the collection of a diverse array of data points, including water quality metrics, vegetation mapping, and assessments of aquatic species distribution.

In future research, an in-depth exploration of the performance of various deep learning models in river trash detection tasks is planned. Specifically, comparisons will be made between YOLOv10-N and other advanced object detection models, such as YOLO-world and the Segment Anything model, to evaluate their differences in detection accuracy, speed, and robustness. This comparison will aid in understanding the advantages and limitations of each model, facilitating the selection of the most appropriate model framework for future research and applications. Additionally, the quality and diversity of the dataset are recognized as essential for training efficient and accurate detection models. Therefore, incorporating data augmentation techniques to balance data distribution will be a part of the research, addressing the issue of class imbalance and enhancing the model's ability to recognize various types of trash. A variety of data augmentation methods will be explored, including image rotation, scaling, and color adjustment, to generate a more diverse set of training samples. Through these future research efforts, it is anticipated that the performance of river trash detection technology will be significantly enhanced, providing more reliable technical support for practical applications.

5. CONCLUSIONS

The focus of this study is on improving the performance of river litter detection by developing a comprehensive dataset of floating litter in rivers and applying it to deep learning models. The dataset covers 5 litter types and different morphologies, and is tested on the YOLOv10-N object detection algorithm. Experimental results show that the YOLOv10-N algorithm achieves 92.4% precision, 90.2% recall, and 94.4% at 0.5 IoU, respectively. By incorporating a diverse range of litter types, such as plastic bottles, bags, boxes, cups, and aluminum cans, the dataset provides a solid foundation for training deep learning models to recognize and classify plastic waste accurately. These remarkable results not only demonstrate the quality of the dataset itself, but also its key role in improving model performance. The dataset ensures high annotation accuracy and abundant sample size through a carefully designed collection and annotation process, providing the necessary training and validation information for deep learning models. This diversity and meticulous annotations enable the model to capture the subtle features of plastic litter, thereby achieving high accuracy in the detection process. Future work will further explore how to expand the diversity of the dataset to optimize other deep learning models and how to apply it to a wider range of environmental monitoring tasks.

ACKNOWLEDGMENTS

The authors would like to thank the support from the 2023 Suzhou Water Bureau Science and Technology Programme (RRSP10120240028); Suzhou Municipal Key Laboratory for Intelligent Virtual Engineering (SZS2022004).

REFERENCES

- Compa, M., March, D., & Deudero, S. (2019). Spatio-temporal monitoring of coastal floating marine debris in the Balearic Islands from sea-cleaning boats. *Marine Pollution Bulletin*, 141, 205–214. <https://doi.org/10.1016/j.marpolbul.2019.02.027>
- Dhillon, A., & Verma, G. K. (2020). Convolutional neural network: A review of models, methodologies and applications to object detection. *Progress in Artificial Intelligence*, 9(2), 85 – 112. <https://doi.org/10.1007/s13748-019-00203-0>
- Varnima, E.K., & Ramachandran, C. (2020). Real-time gender identification from face images using You Only Look Once (YOLO). In *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)* (pp. 48184). <https://doi.org/10.1109/icoei48184.2020.9142989>
- Esteva, A., Chou, K., Yeung, S., Naik, N., Madani, A., Mottaghi, A., Liu, Y., Topol, E., Dean, J., & Socher, R. (2021). Deep learning-enabled medical computer vision. *NPJ Digital Medicine*, 4(1), 5. <https://doi.org/10.1038/s41746-020-00376-2>
- Fan, C. (Nancy), Wall, G., & Mitchell, C. (2008). Heritage retail centres, creative destruction and the water town of Luzhi, Kunshan, China. In *Tourism in China* (1st ed., pp. 25). Routledge. <https://doi.org/10.4324/9780203886366>
- González-Fernández, D., Cózar, A., Hanke, G., Viejo, J., Morales-Caselles, C., Bakiu, R., Barceló, D., Bessa, F., Bruge, A., Cabrera, M., Castro-Jiménez, J., Constant, M., Crosti, R., Galletti, Y., Kideys, A. E., Machitadze, N., Pereira de Brito, J., Pogojeva, M., Ratola, N., ... Tourgeli, M. (2021). Floating macrolitter leaked from Europe into the Ocean. *Nature Sustainability*, 4(6), 474–483. <https://doi.org/10.1038/s41893-021-00722-6>
- Ham, S., Oh, Y., Choi, K., & Lee, I. (2018). Semantic segmentation and unregistered building detection from UAV images using a deconvolutional network. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences—ISPRS Archives*, 42, 419 – 424.
- HumanSignal. (n.d.). HumanSignal/labelImg: *LabelImg is now part of the Label Studio community. The popular image annotation tool created by Tzutalin is no longer actively being developed, but you can check out Label Studio, the open-source data labeling tool for images, text, hypertext, audio, video, and time-series data.* GitHub. <https://github.com/HumanSignal/labelImg> (Accessed: October 20, 2024).
- Jakovljevic, G., Govedarica, M., & Alvarez-Taboada, F. (2020). A deep learning model for automatic plastic mapping using unmanned aerial vehicle (UAV) data. *Remote Sensing*, 12(9), 1515. <https://doi.org/10.3390/rs12091515>

- Jambeck, J. R., Geyer, R., Wilcox, C., Siegler, T. R., Perryman, M., Andrady, A., ... Law, K. L. (2015). Plastic waste inputs from land into the Ocean. *Science*, 347(6223), 768–771.
<https://doi.org/10.1126/science.1260352>
- Jia, T., Kapelan, Z., de Vries, R., Vriend, P., Peereboom, E. C., Okkerman, I., & Taormina, R. (2023). Deep learning for detecting macroplastic litter in water bodies: A review. *Water Research*, 231, 119632.
<https://doi.org/10.1016/j.watres.2023.119632>
- Kamarudin, N. A., et al. (2021). Development of water surface mobile garbage collector robot. *Alinteri Journal of Agriculture Sciences*, 36(1), 534–540. <https://doi.org/10.47059/alinteri/v36i1/ajas21076>
- Kaza, S., Yao, L., Bhada-Tata, P., & Van Woerden, F. (2018). *What a waste 2.0: A global snapshot of solid waste management to 2050*. World Bank Publications.
<https://openknowledge.worldbank.org/handle/10.986/30317>
- Lin, F., Hou, T., Jin, Q., & You, A. (2021). Improved YOLO-based detection algorithm for floating debris in waterways. *Entropy*, 23(9), 1111. <https://doi.org/10.3390/e23091111>
- Maharjan, N., Miyazaki, H., Pati, B. M., Dailey, M. N., Shrestha, S., & Nakamura, T. (2022). Detection of river plastic using UAV sensor data and deep learning. *Remote Sensing*, 14, 3049.
<https://doi.org/10.3390/rs14133049>
- Meijer, L. J., Van Emmerik, T., Van der Ent, R., Schmidt, C., & Lebreton, L. (2021). More than 1000 rivers account for 80% of global riverine plastic emissions into the ocean. *Science Advances*, 7(18).
<https://doi.org/10.1126/sciadv.aaz5803>
- Pande, B., Padamwar, K., Bhattacharya, S., Roshan, S., & Bhamare, M. (2022). A review of image annotation tools for object detection. In *2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*. <https://doi.org/10.1109/icaaic53929.2022.9792665>
- Sapkota, R., Qureshi, R., Flores Calero, M., Badjugar, C., Nepal, U., Poullose, A., Zeno, P., Vaddevolu, U. B. P., Khan, S., Shoman, M., Yan, H., & Karkee, M. (2024). YOLOv10 to its genesis: A decadal and comprehensive review of the You Only Look Once (YOLO) series.
<https://doi.org/10.48550/arXiv.2406.19407>
- Tabassum, S., Ullah, S., Al-nur, N. H., & Shatabda, S. (2020). Poribohon-BD: Bangladeshi local vehicle image dataset with annotation for classification. *Data in Brief*, 33, 106465.
<https://doi.org/10.1016/j.dib.2020.106465>
- Van Calcar, C., & Van Emmerik, T. (2020). Abundance of plastic debris across European and Asian rivers.
<https://doi.org/10.5194/egusphere-egu2020-707>
- Van Lieshout, C., Van Oeveren, K., Van Emmerik, T., & Postma, E. (2020). Automated river plastic monitoring using deep learning and cameras. *Earth and Space Science*, 7(8). <https://doi.org/10.1029/2019EA000960>
- Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., & Ding, G. (2024). YOLOv10: Real-time end-to-end object detection. <https://doi.org/10.48550/arXiv.2405.14458>
- Wu, TW., Zhang, H., Peng, W., Lu, F., & He, P. J., (2023). Applications of convolutional neural networks for intelligent waste identification and recycling: A review. *Resources, Conservation and Recycling*, 190, 106813. <https://doi.org/10.1016/j.resconrec.2022.106813>
- Xiao, B., & Kang, S.-C. (2021). Development of an image data set of construction machines for deep learning object detection. *Journal of Computing in Civil Engineering*, 35(2).
[https://doi.org/10.1061/\(asce\)cp.1943-5487.0000945](https://doi.org/10.1061/(asce)cp.1943-5487.0000945)
- Zhang, C., Xu, X., & Xia, Y. (2024). Riverine Plastic Litter Dataset. Mendeley Data.
<https://doi.org/10.17632/7yvrhkxz22.1>

MAPPING AIR POLLUTION CONCENTRATIONS ON SIDEWALKS IN CENTRAL BUSINESS DISTRICTS BASED ON MOBILE MONITORING

Xujing Yu¹, Jun Ma²

1) Ph.D. Candidate, Department of Urban Planning and Design, The University of Hong Kong, Hong Kong, SAR.

Email: xujingyu@connect.hku.hk

2) Ph.D., Assis. Prof., Department of Urban Planning and Design, The University of Hong Kong, Hong Kong, SAR.

Email: junma@hku.hk

Abstract: Air pollution on urban sidewalks poses a huge threat to human health. We conducted an air pollution mobile monitoring campaign to measure PM_{2.5} levels on sidewalks in the Futian CBD area of Shenzhen, using portable air pollution sensors. The campaign involved collecting data during peak commuting times to capture variations in air quality. Calibration of the AirBeam3 devices against a standard instrument ensured data accuracy. Using GIS spatial analysis, we mapped air pollution patterns and identified notable differences between weekdays and weekends, with concentrations ranging from 3-9 µg/m³. Our findings indicate higher pollution levels on weekdays, particularly in the afternoon, correlating with increased traffic and economic activity. The study also highlights spatio-temporal heterogeneity, with morning pollution concentrations more pronounced in the northern financial district and afternoon levels higher in the southern area, characterized by commercial facilities. These patterns suggest that regional industrial distribution and traffic flow significantly influence air quality. By understanding the dynamics of pollution in urban environments, this research contributes to the development of effective strategies for improving air quality and public health in densely populated city centers.

Key words: PM_{2.5}, Mobile monitoring, GIS, Spatio-temporal heterogeneity

1. INTRODUCTION

Rapid urbanization and industrialization have made urban air pollution a critical environmental issue (Du et al., 2019; S. Wang et al., 2020). Road traffic emissions are a major contributor to this pollution and are closely linked to various health conditions (Jandacka et al., 2017; Tobias et al., 2020). To monitor air pollution and mitigate its negative impact on social development, governments worldwide have established networks of fixed-site monitors, including roadside stations, to measure air quality and set standards. However, due to high equipment and operational costs, these monitors are often sparsely distributed.

Since air pollutant concentrations can vary significantly over short distances and time periods, static networks alone cannot fully characterize the urban environment. Consequently, there is growing interest in using mobile monitoring to study the spatial distribution of air pollution. For example, Shi et al. (2018) used trams to collect long-term street-level PM_{2.5} data in Hong Kong, identifying key building design factors affecting pollution dispersion. DeSouza et al. (2020) identified PM_{2.5} hotspots using low-cost sensors mounted on trash trucks. Automobiles are the most commonly used vehicles for mobile air pollution sensing, as seen in studies from the United States (Messier et al., 2018; Miller et al., 2020), Europe (Shah et al., 2023), and China (C. Huang et al., 2022; G. Huang et al., 2023; Shi et al., 2016). Additionally, bicycles equipped with low-cost sensors are frequently used to monitor roadside air pollution (Hassani et al., 2023; von Schneidmesser et al., 2019).

Previous research using mobile monitoring has made significant progress, yet certain gaps remain. Firstly, most studies have utilized vehicles as the primary mobile monitoring platforms. While vehicle-mounted sensors can monitor urban road air pollution on a large scale, they are limited to specific routes

and cannot capture pollution levels on sidewalks. Walking is a crucial mode of travel, especially in central business districts (CBDs), where sidewalks are close to traffic (Rakowska et al., 2014) but often lack sufficient physical protection (Lv et al., 2021). Consequently, pedestrians are exposed to high pollution levels. Furthermore, some studies have monitored air pollution at specific sidewalk locations, such as bus stops (Xing et al., 2019) and intersections (Z. Wang et al., 2021) or in limited areas (Alas et al., 2022), they fail to fully capture regional spatiotemporal variations in pedestrian-level pollution. Understanding these patterns is crucial for urban planners to propose interventions that reduce pollution and protect pedestrians.

To address these gaps, this study focuses on the Shenzhen CBD, measuring PM_{2.5} levels on sidewalks to study spatial-temporal heterogeneity. The objectives are: (1) to use low-cost sensors for mobile monitoring of sidewalk air pollution, and (2) to employ GIS for data aggregation and spatial analysis to investigate spatiotemporal patterns.

2. METHOD

2.1. Study area

Shenzhen, a prominent mega-city in southern China's Guangdong Province, borders Hong Kong to the south and covers an area of 1,997.47 square kilometers. With a resident population of 17,790,100, Shenzhen is one of China's three major national financial centers. This study focuses on the Futian CBD area, chosen for its high pedestrian traffic and significant air pollution exposure risks. Air pollution data was collected from walkable roads in the CBD.

2.2. Data collection campaign

In this study, we focused on particulate matter with a particle size of less than 2.5 micrometers (PM_{2.5}) as a representative air pollutant, given its prominence in near-road environments. We measured PM_{2.5} using the low-cost AirBeam3 air quality monitor, which has been widely used in previous research (Johnston et al., 2019; Song & Kwan, 2023). The device features a Plantower PMS7003 sensor for PM_{2.5}, as well as a relative humidity and temperature sensor and a GPS locator, enabling meteorological data collection and spatial analysis. The temporal resolution of AirBeam3 data varies with the recording duration: data is recorded every second for sessions under 2 hours, averaged every 5 seconds for sessions between 2 and 9 hours, and averaged every minute for sessions exceeding 9 hours.

The data collection campaign took place from July 1 to July 8, 2024, during sunny weather. Eight research assistants were recruited to walk the sidewalks and collect air pollution data. Prior to the campaign, our team provided training on the monitoring routes and instrument usage, as shown in the Figure 1 (a). PM_{2.5} measurements were taken from 7:30 to 9:30 a.m. and 5:30 to 7:30 p.m., coinciding with peak commuting times and higher pollution levels. The instrument was worn on a lanyard around the chest, positioning the PM sensor intake and exhaust ports away from the direction of travel to minimize wind interference, as shown in the Figure 1 (b).



Figure 1. Photos from: (a) Training for research assistants; (b) The process of monitoring

2.3. Instruments calibration

We conducted a laboratory comparison between the low-cost PM_{2.5} instruments and a regulatory instrument to calibrate the AirBeam3. The eight AirBeam3 devices, labeled A through H, were placed side by side. The standard instrument used was the Profiler Model 212 (MetOne Instruments, Inc.), a robust optical particle counter widely used in air monitoring (Brattich et al., 2020). The calibration involved generating a PM_{2.5} linear calibration equation between the AirBeam3 and the standard instrument, following a method from previous studies (Zhou & Lin, 2019). Each AirBeam3 was tested and calibrated using 5 minutes average values over a 24-hour period. The fitting equation is shown in the Equation (1):

$$Y = \beta_1 X + \beta_0 \quad (1)$$

where Y is the PM concentrations recorded by the standard instrument, and X is PM concentrations recorded by the AirBeam3 sensors.

3. RESULTS

3.1. Results of instrument calibration

The coefficient of determination (R^2) was used to assess the relationship between the instruments. Figure 2 shows that R^2 values between the AirBeam3 devices and the standard instrument ranged from 90.8% to 96.5%, indicating a strong linear correlation. Thus, the eight linear equations derived were used for data calibration.

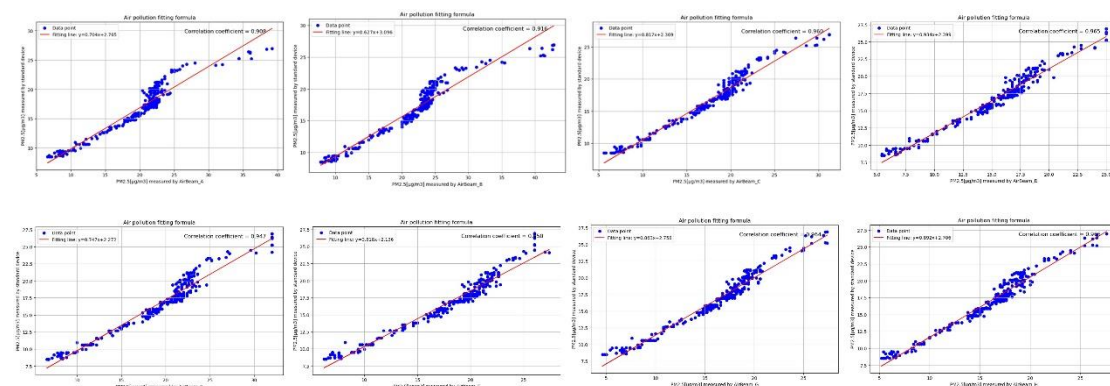


Figure 2 Measurement comparison between eight AirBeam3s and the standard instrument

3.2. Air pollution data aggregation and mapping

After data collection and calibration, air pollution data points were aggregated at the sidewalk segment level. To determine the spatial resolution (or segment length), we used semi-variogram modeling in GIS. The semi-variogram, a function of distance, increases over a certain range until reaching stability. We divided the time periods into four categories: weekday mornings, weekday afternoons, weekend mornings, and weekend afternoons.

The semi-variogram results, as shown in the Figure 3, indicating that the major range for morning data (17.4m on weekdays and 9.4m on weekends) is larger than for afternoon data (1.64m on weekdays and 1.50m on weekends). This indicates that morning data points are spatially correlated over a greater distance. To ensure consistent spatial resolution, we chose a 10m section of the road as the study unit.

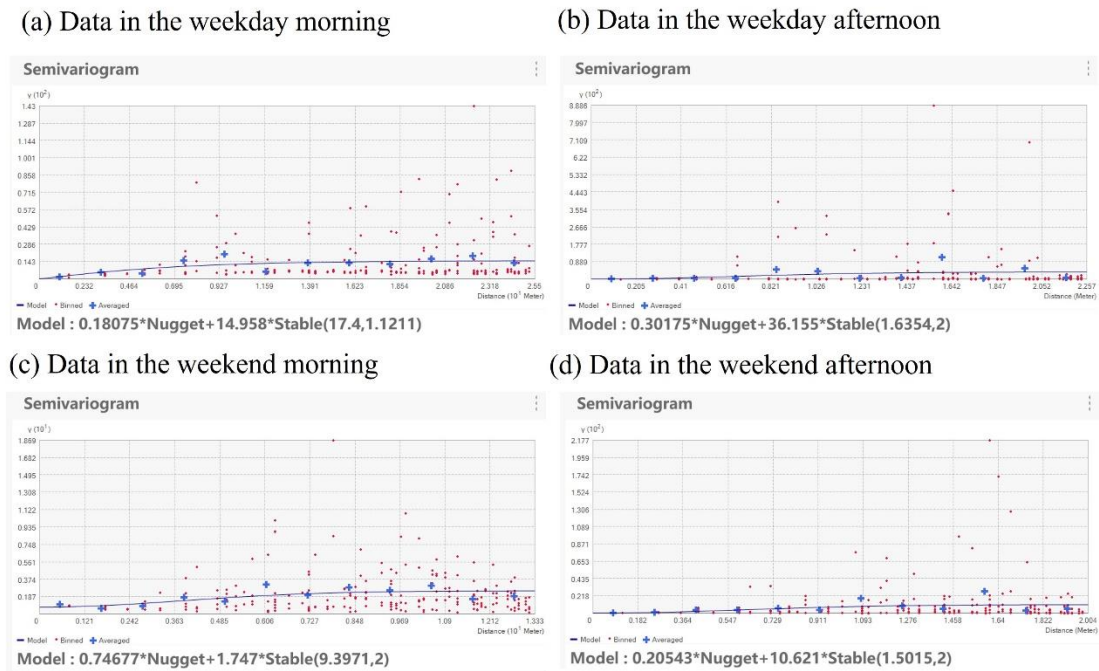


Figure 3 The semi-variogram models and the corresponding major ranges of the PM_{2.5} datasets

After data aggregation, PM_{2.5} distribution maps in Figure 4 illustrate the spatial and temporal variations in air pollution. Table 1 presents the descriptive statistics of air pollution on sidewalks in the Futian CBD area. The results indicate that average air pollution concentrations are higher on weekdays than on weekends. The highest average concentrations were recorded on weekday afternoons, while the lowest were observed on weekend mornings.



Figure 4 PM_{2.5} distribution maps in the: (a) weekday morning; (b) weekday evening; (c) weekend

morning; (d) weekend evening

Table 1 Descriptive statistics of air pollution on sidewalks

Time	Mean ($\mu\text{g}/\text{m}^3$)	Min ($\mu\text{g}/\text{m}^3$)	Max ($\mu\text{g}/\text{m}^3$)	Std ($\mu\text{g}/\text{m}^3$)
Weekday morning	7.5	2.14	25.1	2.06
Weekday afternoon	8.8	3.61	64.5	3.26
Weekend morning	3	2.37	67.5	2.14
Weekend evening	3.84	2.4	55.4	2.36

3.3. Hotspots analysis

To better demonstrate the spatial-temporal heterogeneity of air pollution, a hotspots analysis was conducted, with results shown in the figure. As shown in the Figure 5, the study area exhibits a higher number of hotspots compared to cold spots, indicating a pronounced clustering of high pollution on sidewalks in the CBD. Hot and cold spots are less evident on weekend mornings, likely because lower traffic and pedestrian flow result in fewer emissions, leading to a more balanced air pollution distribution. In contrast, on weekend evenings, increased economic activities in the CBD area leads to more noticeable concentrations of air pollutants.

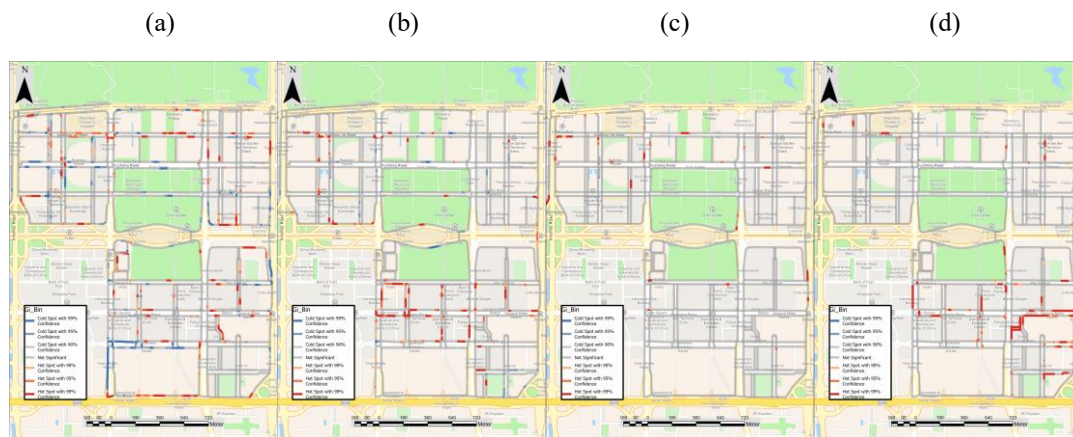


Figure 5 Hotspots analysis results in the: (a) weekday morning; (b) weekday evening; (c) weekend morning; (d) weekend evening

In addition to temporal variations, spatial heterogeneity was also observed. In the morning, pollutant concentrations are more pronounced in the northern part of the study area. This area, north of Shennan Boulevard, is dominated by financial facilities like the Shenzhen Stock Exchange and Ping An Financial Building, attracting significant morning rush hour traffic and emissions. In the afternoon, higher pollutant concentrations shift to the southern part of the study area, which features large restaurants and shopping centers such as Shenzhen One Avenue and the Leader Exhibition Center. These are key destinations for after-hours pedestrian and vehicular traffic. Thus, the spatial and temporal changes in economic activities influence regional air pollution concentrations.

4. CONCLUSIONS

In this study, we conducted an air pollution monitoring campaign in the Futian CBD area of Shenzhen, measuring PM_{2.5} levels on sidewalks using portable sensors. After calibrating the data, we mapped air pollution on weekdays and weekends. We found that mean PM_{2.5} concentrations ranged

from 3–9 $\mu\text{g}/\text{m}^3$ across different time periods. Using GIS spatial analysis, we examined the spatial and temporal heterogeneity of air pollution, suggesting links to regional industrial distribution and traffic flow variations. The study results can inform air pollution interventions on regional sidewalks.

REFERENCES

- Alas, H. D., Stöcker, A., Umlauf, N., Senaweera, O., Pfeifer, S., Greven, S., & Wiedensohler, A. (2022). Pedestrian exposure to black carbon and PM_{2.5} emissions in urban hot spots: New findings using mobile measurement techniques and flexible Bayesian regression models. *Journal of Exposure Science & Environmental Epidemiology*, 32(4), 604–614. <https://doi.org/10.1038/s41370-021-00379-5>
- Brattich, E., Bracci, A., Zappi, A., Morozzi, P., Di Sabatino, S., Porcù, F., Di Nicola, F., & Tositti, L. (2020). How to Get the Best from Low-Cost Particulate Matter Sensors: Guidelines and Practical Recommendations. *Sensors*, 20(11), 3073. <https://doi.org/10.3390/s20113073>
- deSouza, P., Anjomshoa, A., Duarte, F., Kahn, R., Kumar, P., & Ratti, C. (2020). Air quality monitoring using mobile low-cost sensors mounted on trash-trucks: Methods development and lessons learned. *Sustainable Cities and Society*, 60, 102239. <https://doi.org/10.1016/j.scs.2020.102239>
- Du, Y., Wan, Q., Liu, H., Liu, H., Kapsar, K., & Peng, J. (2019). How does urbanization influence PM_{2.5} concentrations? Perspective of spillover effect of multi-dimensional urbanization impact. *Journal of Cleaner Production*, 220, 974–983. <https://doi.org/10.1016/j.jclepro.2019.02.222>
- Hassani, A., Castell, N., Watne, Å. K., & Schneider, P. (2023). Citizen-operated mobile low-cost sensors for urban PM_{2.5} monitoring: Field calibration, uncertainty estimation, and application. *Sustainable Cities and Society*, 95, 104607. <https://doi.org/10.1016/j.scs.2023.104607>
- Huang, C., Hu, T., Duan, Y., Li, Q., Chen, N., Wang, Q., Zhou, M., & Rao, P. (2022). Effect of urban morphology on air pollution distribution in high-density urban blocks based on mobile monitoring and machine learning. *Building and Environment*, 219, 109173. <https://doi.org/10.1016/j.buildenv.2022.109173>
- Huang, G., Huang, X., Liu, C., Wu, L., Liu, G., Xing, Y., Li, J., & Yan, M. (2023). Characterizing spatiotemporal patterns of elevated PM_{2.5} exposures in a megacity of China using combined mobile and stationary measurements. *Atmospheric Environment*, 307, 119821. <https://doi.org/10.1016/j.atmosenv.2023.119821>
- Jandacka, D., Durcanska, D., & Bujdos, M. (2017). The contribution of road traffic to particulate matter and metals in air pollution in the vicinity of an urban road. *Transportation Research Part D: Transport and Environment*, 50, 397–408. <https://doi.org/10.1016/j.trd.2016.11.024>
- Johnston, J. E., Juarez, Z., Navarro, S., Hernandez, A., & Gutschow, W. (2019). Youth Engaged Participatory Air Monitoring: A ‘Day in the Life’ in Urban Environmental Justice Communities. *International Journal of Environmental Research and Public Health*, 17(1), 93. <https://doi.org/10.3390/ijerph17010093>
- Lv, H., Li, H., Qiu, Z., Zhang, F., & Song, J. (2021). Assessment of pedestrian exposure and deposition of PM₁₀, PM_{2.5} and ultrafine particles at an urban roadside: A case study of Xi’an, China. *Atmospheric Pollution Research*, 12(4), 112–121. <https://doi.org/10.1016/j.apr.2021.02.018>
- Messier, K. P., Chambliss, S. E., Gani, S., Alvarez, R., Brauer, M., Choi, J. J., Hamburg, S. P., Kerckhoffs, J., LaFranchi, B., Lunden, M. M., Marshall, J. D., Portier, C. J., Roy, A., Szpiro, A. A., Vermeulen, R. C. H., & Apte, J. S. (2018). Mapping Air Pollution with Google Street View Cars:

- Efficient Approaches with Mobile Monitoring and Land Use Regression. *Environmental Science & Technology*, 52(21), 12563–12572. <https://doi.org/10.1021/acs.est.8b03395>
- Miller, D. J., Actkinson, B., Padilla, L., Griffin, R. J., Moore, K., Lewis, P. G. T., Gardner-Frolick, R., Craft, E., Portier, C. J., Hamburg, S. P., & Alvarez, R. A. (2020). Characterizing Elevated Urban Air Pollutant Spatial Patterns with Mobile Monitoring in Houston, Texas. *Environmental Science & Technology*, 54(4), 2133–2142. <https://doi.org/10.1021/acs.est.9b05523>
- Rakowska, A., Wong, K. C., Townsend, T., Chan, K. L., Westerdahl, D., Ng, S., Močnik, G., Drinovec, L., & Ning, Z. (2014). Impact of traffic volume and composition on the air quality and pedestrian exposure in urban street canyon. *Atmospheric Environment*, 98, 260–270. <https://doi.org/10.1016/j.atmosenv.2014.08.073>
- Shah, R. U., Padilla, L. E., Peters, D. R., Dupuy-Todd, M., Fonseca, E. R., Ma, G. Q., Popoola, O. A. M., Jones, R. L., Mills, J., Martin, N. A., & Alvarez, R. A. (2023). Identifying Patterns and Sources of Fine and Ultrafine Particulate Matter in London Using Mobile Measurements of Lung-Deposited Surface Area. *Environmental Science & Technology*, 57(1), 96–108. <https://doi.org/10.1021/acs.est.2c08096>
- Shi, Y., Lau, K. K.-L., & Ng, E. (2016). Developing Street-Level PM_{2.5} and PM₁₀ Land Use Regression Models in High-Density Hong Kong with Urban Morphological Factors. *Environmental Science & Technology*, 50(15), 8178–8187. <https://doi.org/10.1021/acs.est.6b01807>
- Song, W., & Kwan, M.-P. (2023). Air pollution perception bias: Mismatch between air pollution exposure and perception of air quality in real-time contexts. *Health & Place*, 84, 103129. <https://doi.org/10.1016/j.healthplace.2023.103129>
- Tobías, A., Carnerero, C., Reche, C., Massagué, J., Via, M., Minguillón, M. C., Alastuey, A., & Querol, X. (2020). Changes in air quality during the lockdown in Barcelona (Spain) one month into the SARS-CoV-2 epidemic. *Science of The Total Environment*, 726, 138540. <https://doi.org/10.1016/j.scitotenv.2020.138540>
- von Schneidmesser, E., Steinmar, K., Weatherhead, E. C., Bonn, B., Gerwig, H., & Quedenau, J. (2019). Air pollution at human scales in an urban environment: Impact of local environment and vehicles on particle number concentrations. *Science of The Total Environment*, 688, 691–700. <https://doi.org/10.1016/j.scitotenv.2019.06.309>
- Wang, S., Gao, S., Li, S., & Feng, K. (2020). Strategizing the relation between urbanization and air pollution: Empirical evidence from global countries. *Journal of Cleaner Production*, 243, 118615. <https://doi.org/10.1016/j.jclepro.2019.118615>
- Wang, Z., He, H., Zhao, H., & Peng, Z. (2021). Spatiotemporal analysis of pedestrian exposure to submicron and coarse particulate matter on crosswalk at urban intersection. *Building and Environment*, 204, 108149. <https://doi.org/10.1016/j.buildenv.2021.108149>
- Xing, Y., Brimblecombe, P., & Ning, Z. (2019). Fine-scale spatial structure of air pollutant concentrations along bus routes. *Science of the Total Environment*, 658, 1–7.
- Zhou, S., & Lin, R. (2019). Spatial-temporal heterogeneity of air pollution: The relationship between built environment and on-road PM_{2.5} at micro scale. *Transportation Research Part D: Transport and Environment*, 76, 305–322. <https://doi.org/10.1016/j.trd.2019.09.004>

BLOCKCHAIN-DRIVEN TRANSPARENCY IN CONSTRUCTION TENDERS: SMART CONTRACTS AND SUSTAINABLE WASTE MANAGEMENT

Silvia Meschini¹, Daniele Accardo², Lavinia Chiara Tagliabue³, Stefano Rinaldi⁴, Rosa Meo⁵, Andrea Bracciali⁶,
Giuseppe Martino Di Giuda⁷

1) Assistant Professor, Department of Computer Science, University of Turin, Italy. Email: silvia.meschini@unito.it

2) Ph.D. candidate, Department of Management, University of Turin, Italy. Email: daniele.accardo@unito.it

3) Associate Prof., Department of Computer Science, University of Turin, Italy. Email: laviniachiara.tagliabue@unito.it

4) Associate Prof., Department of Computer Engineering, University of Brescia, Italy. Email: stefano.rinaldi@unibs.it

5) Associate Prof., Department of Computer Science, University of Turin, Italy. Email: andrea.bracciali@unito.it

6) Full Prof., Department of Computer Science, University of Turin, Italy. Email: meo@di.unito.it

7) Full Prof., Department of Management, University of Turin, Italy. Email: giuseppemartino.digiuda@unito.it

Abstract: The integration of blockchain and smart contracts in the construction industry has the potential to revolutionize the tender phase and enhance waste management practices. The prototype is designed to enhance transparency, efficiency, and trustworthiness in Italian public procurement. To analyze the national public procurement database with a view to identifying common issues, which are often related to the lack of trust among stakeholders, Large Language Models (LLMs) are exploited. Blockchain technology has the potential to facilitate this process by eliminating discrepancies and disputes, providing a decentralized, immutable ledger to notarize data related to digital models submitted during the tender phase, thereby ensuring transparency and tamper-proof data. The automation of bid evaluations based on predefined criteria such as those pertaining to waste management, is a key feature of smart contracts, which are of particular importance in the context of construction sustainability. Such assessments are enabled, allowing for unbiased and transparent evaluations based on quantifiable data, with the results recorded automatically on the blockchain. This results in a more efficient tender process and the promotion of sustainable practices, as projects with superior waste management are given priority. A tailor-made blockchain protocol is put forth to delineate requirements and facilitate data exchanges. It establishes standards and procedures for data submission, verification, and evaluation, ensuring secure and transparent interactions and enhancing stakeholder confidence in a fair and transparent evaluation process. In summary, the use of blockchain and smart contracts in the construction tender phase improves data integrity, transparency, and efficiency. The focus on waste management indicators allows for objective project evaluations and the promotion of sustainable practices. This innovative approach has the potential to transform public procurement, establishing a new global standard for the construction industry.

Keywords: Blockchain, Public procurement, Tender phase, LLM, Sustainability, Waste management, BIM, Smart contract, Performance-based evaluation.

1. INTRODUCTION

Some of the issues affecting the Construction Industry are particularly significant in Public Procurement. Despite the considerable impact of the European Union's (EU) single market on the region's gross domestic product (GDP) and expenditure, as well as the advent of digitalization and e-procurement platforms, there is still great room for improvement as can be seen from the latest single market scoreboard analysis based on 12 indicators of quality, transparency, speed and accountability indicators (European Commission, 2022) reveals significant room for improvement. Indeed, the processes in question remain predominantly "paper-based" rather than data-driven, resulting in Information Management and tender evaluation that are characterized by low transparency, time-consuming and error-prone tasks. The necessary information is frequently located within a plethora of documents, which are often incomplete and inconsistent with one another or with the initial requests, thereby negatively impacting the quality of the project. Consequently, this sector is characterized by inefficiency and the high prevalence of corruption and fraud. The Organisation for Economic Co-operation and Development (OECD) estimates that corruption and mismanagement results in losses amounting between 10 and 30% of global economic activity, or approximately 400 billion to 1.2 trillion euros per year (OECD, 2019). Italy exhibits considerable room for improvement in comparison to other EU countries. Its indicators are relatively low, with a decision-making speed of approximately 200 days, which falls below the efficiency threshold of 120 days. Additionally, there is a need to enhance transparency, facilitate the inclusion of small and medium-sized enterprises and improve data quality (European Commission, 2022; OECD, 2019). While current procurement procedures have been somewhat digitalized using platforms, there has not been an update to the implementing regulation that would enable the transition from a paper-based approach to a digital and data-driven one. This indicates that the process remains suboptimal in comparison to the potential currently offered by digital technology and information models. The sole automated task pertains to the mere control of the existence of the requested documents (Meschini et al., 2023), as opposed to the verification of bids' content and their consistency and completeness with respect to the requests of the Public Client. A significant advancement is necessary to effectively and comprehensively digitize the

process, enabling the utilization of information models and digitalized data, through algorithms that can automatically assess the content of the bidding models and related information (Pattini et al., 2022). This would enhance the process's efficiency, productivity, transparency and accuracy, reducing the likelihood of errors or the presence of incomplete or inconsistent bids.

2. BACKGROUND

2.1 Public Procurement and Construction industry

It was demonstrated that an improvement in PP productivity, even by a mere 1%, could result in savings of 43 billion per year in OECD countries (OECD, 2019). Indeed, it could have a high return on investment as Public Clients are significant purchasers and can lead through efficient processes with improved sustainability (Pellegrini et al., 2021). It is also crucial to achieve the objectives set out in the Agenda 2030, as promoted by the NRPP, which has identified digitalization, sustainability and ecological transition as its core principles, with the majority of its resources allocated to these areas. Consequently, EU directives and Italian regulations strongly encourage the utilization of digital and data-driven methodologies to augment the efficiency of the public procurement sector, alongside the implementation of green public procurement, which continues to encounter challenges in the context of prevailing paper-based procedures (Lavagna et al., 2019; OAV, 2022; Sapir et al., 2022). The application of the most economically advantageous tender criterion (MEAT) is necessary to allow for the incorporation of qualitative and quantitative criteria for the purchase of goods and services at an optimal life-cycle cost and quality.

This entails numerous variables that must be considered, particularly in the context of environmental verification. The current approach, in which a judging committee of experts is solely responsible for evaluation, lacks the use of digital tools or automation, which hinders an objective assessment and frequently results in outcomes that do not fully align with the expectations of customers. The implementation of data-driven and digital methods would facilitate the streamlining of procedures and processes for verifying compliance with client and regulatory IRs. This would allow for improved judgment on bids, which are currently entrusted entirely to a jury and fifteen subject to arbitrariness. The use of data-driven and blockchain methods can make the process more objective and transparent, while also facilitating the verification of sustainability protocols that are currently experiencing significant challenges in practice. The application of GPP is particularly limited in this regard (Lavagna et al., 2019; Meschini et al., 2023; Pellegrini et al., 2021) (Figure 1). Consequently, the complete digitalization of this phase would undoubtedly result in a notable reduction in decision-making times, largely due to the automation of numerous procedures, while simultaneously lowering costs. The digitalization of process management has recently been identified as a potential turning point for the sector. Furthermore, the European Union has proposed a common public procurement strategy, comprising six measures, with the aim of improving the sector through collaborative approaches between public administrations and stakeholders. These strategies encompass the advancement of digitalization, enabling data-driven decision-making, and the incorporation of digital methodologies throughout the entirety of the procurement process. This approach is designed to facilitate the implementation of fully digitalized processes, thereby realizing substantial time and cost saving. Additionally, it aims to enhance operational transparency and mitigate disputes (European Commission, 2017).

Perf. Ind.	AT	BE	BG	CY	CZ	DE	DK	EE	EL	ES	FI	FR	HR	HU	IE	IS	IT	LI	LT	LU	LV	MT	NL	NO	PL	PT	RO	SE	SI	SK
Indicator 1	29	24	34	16	42	25	23	32	48	28	14	22	23	33	21	14	37	18	34	23	37	15	19	16	52	24	42	13	51	30
Indicator 2	7	2	23	39	10	6	5	11	1	10	3	2	5	5	1	3	5	7	6	6	8	9	9	1	9	5	23	1	11	5
Indicator 3	4.8	3.7	5.8	3.8	5.7	2.1	6.5	9.2	3.3	4.3	7.0	8.4	7.9	9.4	3.3	n/a	6.0	n/a	6.2	4.5	10.9	4.4	3.8	n/a	7.9	2.4	8.7	6.8	4.7	6.1
Indicator 4	6	16	1	9	5	13	17	17	3	1	21	6	3	11	16	1	11	0	2	1	17	0	3	11	3	4	1	1	5	7
Indicator 5	30	37	87	94	81	66	50	80	85	22	58	8	1	49	15	90	40	73	93	70	72	92	21	25	58	65	86	83	68	96
Indicator 6	87	110	100	128	90	51	70	52	242	157	88	92	90	112	108	53	197	43	68	110	72	177	84	49	60	99	105	44	82	175
Indicator 7	57	33	74	88	68	61	52	82	51	48	56	41	50	75	67	92	48	94	91	63	92	92	67	46	55	45	53	58	85	82
Indicator 8	78	59	72	85	95	71	62	90	85	58	60	66	39	74	74	99	60	100	94	96	91	91	56	48	58	58	63	12	80	84
Indicator 9	11	31	35	22	19	18	25	30	37	30	28	42	38	38	14	6	27	9	42	8	36	16	16	10	46	35	47	6	40	33
Indicator 10	4	2	2	0	1	1	1	0	0	4	0	2	0	0	3	0	1	0	0	0	0	2	1	0	1	3	0	2	7	0
Indicator 11	51	90	0	0	6	99	43	5	88	29	6	88	0	12	33	25	94	80	5	98	4	3	43	47	66	58	0	4	2	1
Indicator 12	87	14	0	1	2	97	13	0	76	10	0	54	0	0	69	7	85	96	0	92	0	18	6	0	48	27	1	1	1	0

Figure 1. Scoreboard of EU Single Market – Public Procurement, 2022

2.2 Public procurement database scraping with LLM

The creation of new tenders frequently necessitates an examination of previously published tenders within the specified target area. The objective of this research is to propose the adoption of Retrieval-Augmented Generation (RAG), which is a Natural Language Processing (NLP) approach that combines information retrieval (IR) techniques with LLM text generation. This enables the production of more accurate and relevant responses to user queries (Gua et al., 2020). The paper demonstrates the information flow required to construct a recommendation system based on the national Anticorruption Authority (ANAC) database, which contains the

"metadata" of the Italian Public Administration (PA) tenders since 2010. Each text, comprising a few paragraphs describing each contract, is transformed into a numerical vector using Large Language Models (LLMs) provided by Cohere and OpenAI. The resulting embeddings are stored in an internal database (vector base) and are available for rapid retrieval when a content search is submitted by the user. In response to user queries, the system returns the k most similar procurement examples, determined by the strength of the embedding between the query and the vectors in the vector base. To assess the efficacy of the recommendation system, a metric known as NDCG (Normalized Discounted Cumulative Gain) is employed. This metric gauge the relevance and usefulness of the items presented in response to user queries. The measure is normalized against an ideal ordering of the items in the answers (Figure 2).

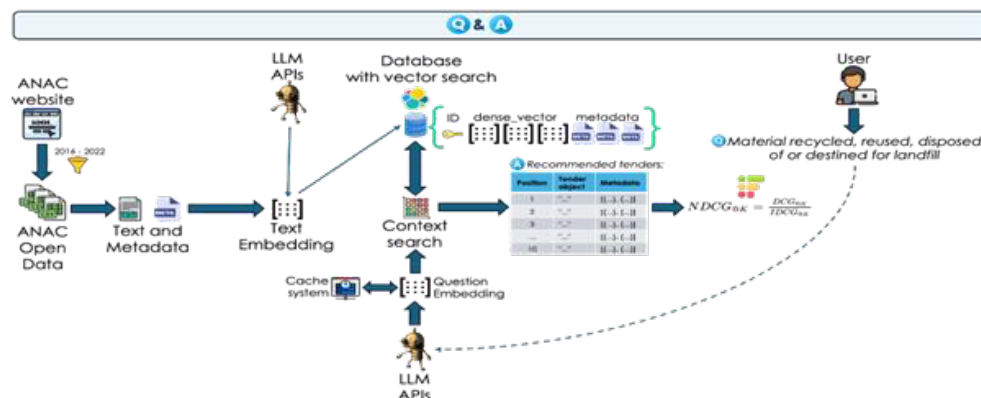


Figure 2. Database scraping workflow on the ANAC website for waste management

Although the BM25 model (Robertson et al., 2009) has a strong historical performance, it is evident that it has limitations when used as a baseline in this comparison. The results demonstrate that more sophisticated and expansive models tend to exhibit superior relevance ranking performance, although the extent of improvement may vary depending on the query and LLMs utilized. This indicates that the selection of a model for information retrieval tasks should consider the trade-offs between computational resources and the incremental gains in retrieval effectiveness (Nai et al., 2024).

Accordingly, the subsequent phases of the research will focus on:

- Implementation of a service using Flask and ReactJS that will facilitate web-based tool usage.
- Introduction of further evaluation metrics in addition to the NDCG and a deeper exploration of qualitative evaluation by domain experts, with the addition of more queries to be submitted to the system.
- Comparison of the proposed RAG framework with other existing frameworks, such as Fast-RAG.
- Utilization of open LLMs such as LLaMA in lieu of proprietary tools, including Cohere and OpenAI.

2.3 Automatic checking, evaluation and awarding in construction tenders

Among recent technologies, blockchain and Smart Contract have demonstrated considerable potential to enhance the traditional contract execution process. They facilitate a robust connection between stakeholders, materials and processes, and enable the use of computational language in contractual clauses, which prevents free interpretation or information asymmetry. This is evidenced by sector research (Kim et al., 2024; Pattini et al., 2022; Sreckovic et al., 2022). During the tendering phase, bidders are expected to develop their projects in accordance with the specifications set forth by the client and the relevant regulatory requirements. Smart Contracts can be drafted as a ledger that records all information exchanges among the parties, thereby ensuring high transparency and reliability. This is crucial for providing collaborative approaches and trustworthiness. A blockchain platform allows for the exploitation of a full digital process, enabling the faithful development of bidding models starting from either the preliminary or final BIM model, in accordance with the guidelines set forth in the notice documentation. The terms of the contract, the criteria for evaluating bids, and the notice documents can be translated into a Smart Contract, which will then be unchangeable and shared equally among all participants. Once submitted, the bids can be stored on the platform in an immutable state, thereby ensuring a transparent and auditable awarding process. The bidding models can be automatically evaluated by comparing them against the Information Requirements (IRs) and evaluation criteria defined in the notarized tender documentation. The findings of each bid verification can be compared and validated by executing the predefined Smart Contract. Once the procedure has been concluded, the scores assigned to each bid can be automatically compared, identifying the bid that has achieved the highest level of compliance with the specified requirements. The blockchain provides a distributed digital environment that can eliminate the issues of information asymmetry, ensuring information trust and transparency throughout the process (Kim et al., 2024; Pattini et al., 2022; Sreckovic et al., 2022). It is not necessary for the parties involved to establish trust in one another; they can instead rely on the system itself. As a

result, it is possible to provide superior project quality, transparency and objectivity, preventing clashes or obscurity in the process with potential disputes, additional expenses and time delays (Figure 3).

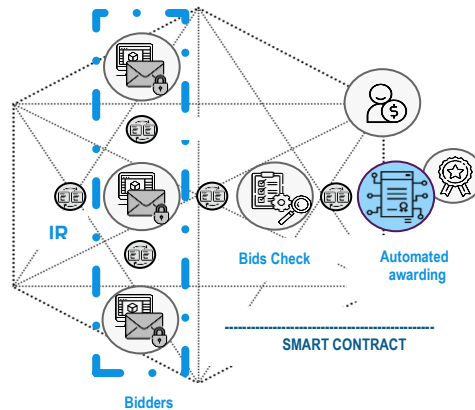


Figure 3. Automated tender process scheme

Aiming at achieving this objective, current approaches need to be revised from a model-based, open-source perspective, thereby enabling the full digitalization and data-driven approaches to tender evaluation. Such a shift could prove highly disruptive, necessitating a fundamental change in perspective (Meschini et al., 2023). The fundamental concept is that the system operates on a web platform that incorporates data, IFC (Industry Foundation Class) models and processes that have been formalized in BPMN 2.0 (Business Process Management Notation), thus making them machine readable (Object Management Group, 2011). The client disseminates the tender documentation via the platform, whereupon the bidder downloads it and the final IFC model. Subsequently, the bidder produces a bid and uploads the IFC bid model. A unique identifier, such as a hash, can be assigned to each bid and subsequently verified against the Client's and regulatory IRs. Models that fail to meet the requisite standards may be disqualified, with the rationale for this decision noted. Those that satisfy the requisite standards may then proceed to the evaluation phase. Subsequently, the judging commission initiates the evaluation process, which is conducted in accordance with the MEAT criterion. This process can be automated, thereby increasing the objectivity and transparency of the evaluation. The use of open formats, evaluation algorithms and score visualization and comparison through analytical dashboards allows for the evaluation of both the individual bid and the overall comparison, up to the award. Furthermore, the winning project model is stored on the platform. By leveraging blockchain and smart contracts, its consistency is guaranteed, as it becomes unreviewable and can be utilized as a benchmark for the subsequent construction phase. This provides a valid prototype of the building, developed with an information modelling and management approach and then structures the information according to the client's requirements and needs, ensuring consistency and integrity (Kim and Kim, 2024; Meschini et al., 2023) (Figure 4).

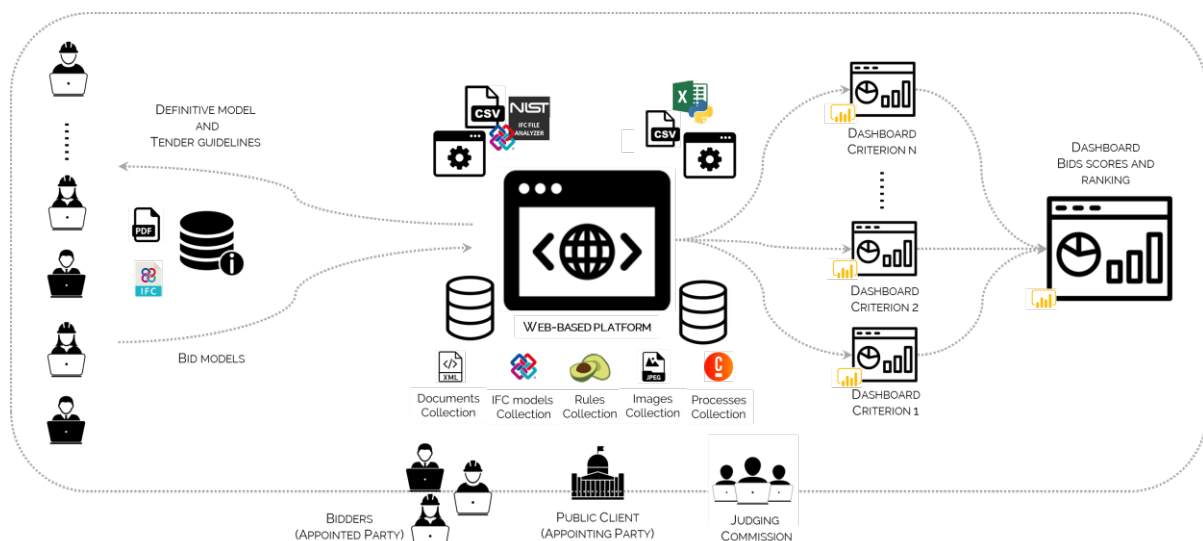


Figure 4. Model-based, open-source, semi-automated approach for bids evaluation and tender awarding

Table 1. Tendering process features

Subject	Features
Evaluator	Web platform with Judging Commission partial actions
Tools	Excel sheets, digital micro services, business intelligence, blockchain (Software, digital tools, simulation engines, codes, rule sets)
Formats	Open and machine readable (IFC, json, csv, xml, RDF, ...)
Evaluation references	Machine-readable rule sets, tender specification and regulations
Awarding criteria Weighting	Objective – notarized - transparent
Performance evaluation	Systematic, simultaneous and automated
Tender duration	Strongly reduced-few days

3. METHODOLOGY

3.1 Automating bids' evaluation

A replicable methodology was devised to automate bid evaluation by extracting essential data from BIM models. This process involves four key steps:

1. BPMN Normalisation: The procurement and criteria evaluation processes are structured using BPMN, allowing identification of information exchanges and automatable tasks.
2. Information Requirements (IRs) Definition: IRs are mapped in IFC, utilising standard or customised property sets.
3. Data Conversion and Evaluation: IFC models are converted into CSV sheets using the NIST IFC Analyzer tool, enabling the extraction of required quantities and scoring via an evaluation algorithm.
4. Analytic Dashboards Development: Dashboards are created to visualise and compare bids against individual criteria, sub-criteria, and overall scores.

Notably, the NIST tool enables viewing all entities and attributes simultaneously, generating spreadsheets from IFC files (Meschini et al., 2023). The procurement process is formalised in BPMN 2.0 to ensure a machine-readable, digitalised system linked to data-driven approaches and smart contracts.

The BPMN diagram includes four lanes: the web platform facilitating the exchange of information; the two upper lanes representing the Public Client and the Judging Committee; and the lower lane representing the Bidders. These actors exchange information, which can be notarised through the blockchain interface. The process begins with the Public Client uploading the tender documents to the platform, including the Information Requirements (IRs), accompanying documents, preliminary or definitive models, and the awarding criteria with their respective weights. Additionally, it is possible to notarise the bidding models, ensuring that all data becomes immutable and shared, thereby guaranteeing maximum transparency.

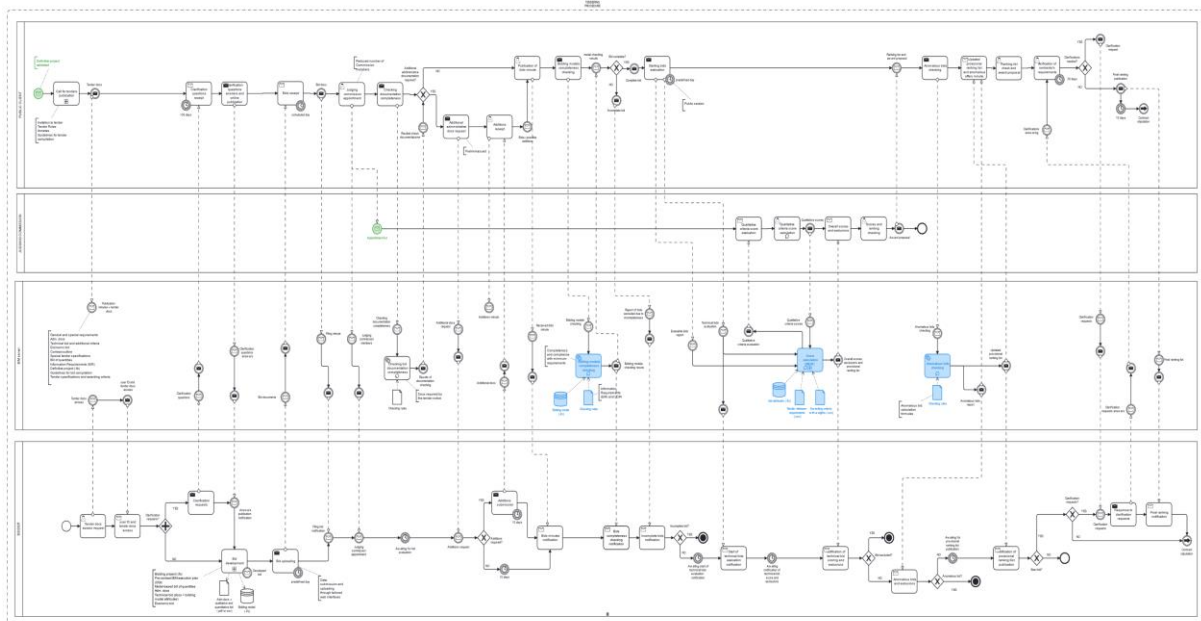


Figure 5. BPMN of the procurement process with automatable sub-processes highlighted

The verification of models can be automated to assess compliance with acceptance thresholds for each criterion and anomaly thresholds, enabling the automatic acceptance or exclusion of bidding models. A smart contract should record the verification results and link to the contract for bid evaluation against awarding criteria. This contract is activated if bidding parameters meet or exceed the minimum acceptance thresholds and do not fall within anomaly ranges. Once the best bid is determined, the smart contract for awarding the tender is triggered, with notarisation of the awarded model. This model can subsequently be used to monitor progress during the construction phase.

The impacts which can be envisioned include:

- Enhancing transparency and fostering collaboration among stakeholders.
- Reducing appointment delivery times through automated design validation.
- Ensuring automated and precise verification of compliance with information requirements.
- Supporting Green Public Procurement (GPP) through data-driven automation, aiding the Client in adopting circular economy principles, and improving waste management by optimising waste production.
- Guaranteeing rewards for stakeholders (via automated payment releases and token delivery), increasing objectivity and shortening decision times.

3.2 Potential of Smart Contract in construction industry

Smart contracts are self-executing agreements where the terms are directly encoded into the contract. They automatically enforce and execute the terms once predefined conditions are met, removing the need for intermediaries. This automation and transparency are particularly advantageous in the construction industry, where projects involve numerous parties and complex agreements (Pattini et al., 2022; Sreckovic et al., 2022; Kim et al., 2024). Implementing smart contracts in construction offers several benefits, including automating contract execution, which speeds up processes and reduces human error; recording every transaction on the blockchain, visible to all involved parties; significantly reducing administrative costs and delays by eliminating intermediaries; and ensuring the immutability of contract terms and project details, preventing tampering.

The main impacts are related to:

- **Efficiency:** Eliminates manual paperwork and reduces administrative overhead.
- **Security:** Immutable records prevent fraud.
- **Trust:** Transparent execution builds trust among stakeholders.

The process of defining a smart contract begins with identifying specific objectives and requirements, followed by gathering all necessary conditions and goals. The appropriate blockchain platform must then be selected, such as Ethereum or Hyperledger, based on scalability, security, and transaction costs. The development phase involves writing the smart contract code using languages like Solidity, ensuring it accurately reflects the agreed-upon conditions and actions. Rigorous testing is conducted in a controlled environment to ensure reliability and security. Once the contract passes all tests, it is deployed to the main blockchain network, making it accessible to all relevant parties. Continuous monitoring is crucial post-deployment to address any issues and implement updates.

Since the most critical part of designing and deploying a smart contract is the business logic, this can be automated using generative AI (Artificial Intelligence) (Petrovic et al., 2023). Generative AI, such as GPT-4, can create content based on input data, enabling the automation of smart contract drafting. This significantly reduces the time and effort required. By following predefined rules and conditions, AI ensures high accuracy and minimises errors, making the generated contracts ready for deployment. The workflow integration is straightforward: first, detailed requirements are gathered and fed into the AI to draft an initial contract. This draft is then reviewed by legal and technical experts to ensure compliance and accuracy. After thorough review, the contract undergoes rigorous testing in a controlled environment before deployment to the blockchain. Using generative AI for smart contracts offers several advantages, such as reducing legal and development costs, ensuring consistency across contracts, and speeding up the development cycle (Petrovic et al., 2023; Nguyen et al., 2024). However, it is essential to use high-quality input data, ensure regulatory compliance, and maintain human oversight to validate the AI-generated content.

4. RESULTS

4.1 PoC: sustainable waste management in a construction tender

As previously mentioned, several challenges typical of the construction industry are particularly evident in public procurement and the tendering phase. In response, the EU has been promoting digitisation and the adoption of data-driven processes to enhance productivity, delivering high returns on investment and positive spillovers across the supply chain while reducing errors, omissions, and variances. Since the introduction of

Directive 2014/24, significant emphasis has been placed on Green Public Procurement (GPP), recognising that the construction industry is one of the most impactful globally, accounting for approximately 36% of waste and necessitating a shift to circular economy models. Directive 2018/851 further underscored the importance of tracking and reducing construction and demolition waste through prevention, reuse, and recycling throughout the lifecycle.

In Italy, Legislative Decree 50/2016 marked a turning point by introducing GPP methodologies in public procurement and the MEAT (Most Economically Advantageous Tender) awarding criterion, recently extended to include BIM and IFC standards. However, challenges remain, particularly the low adoption of GPP and CAM (Minimum Environmental Criteria), largely due to the absence of simple, data-driven tools for tender evaluation with MEAT (Lavagna et al., 2019; Sapir et al., 2022; OAV, 2022). This reliance on paper-based procedures has prolonged procurement processes and affected the quality of awarded bids, wasting resources and limiting control over project performance and compliance with the client's Information Requirements (IRs).

A digital, automated method could address these issues by linking processes and data models through a web-based platform with blockchain notarisation, ensuring automation and trustworthiness. The proposed methodology was tested in a Design and Build procurement model, suitable for model-based approaches where building geometry is defined, and tenders are evaluated based on improving proposals. It focused on the semi-automated evaluation of waste management criteria—typically difficult to assess due to the vast data required. Incorporating these criteria from the tendering phase promotes GPP adoption and reduces environmental impact.

This approach could be extended to other CAM criteria and entire tender evaluations (Meschini et al., 2023). The case study involved a new school in Melzo, near Milan, using a Design and Build model with the MEAT approach. Although GPP and CAM were not mandatory, similar criteria were piloted to demonstrate how Information Modelling and Management can support and enhance the MEAT criterion's application (Figure 6).

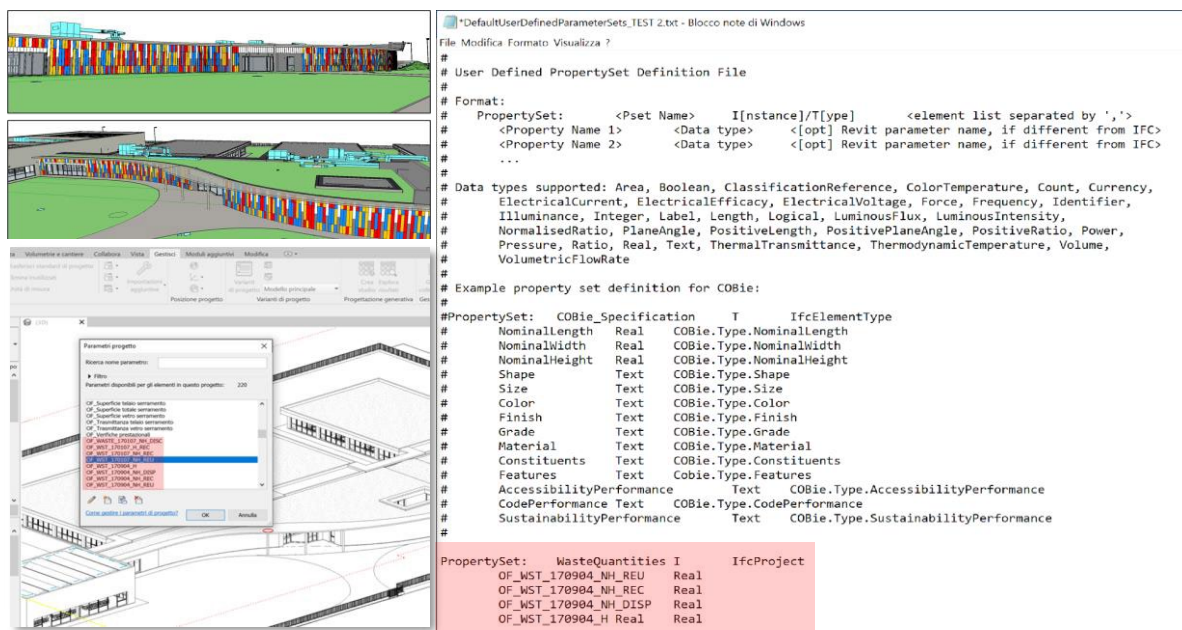


Figure 6. BIM model of the case study with parameters mapped in IFC

The criterion C 2.3, pertaining to the construction phase, necessitates the development and implementation of a waste management plan aimed at maximising the quantities of reused and recycled waste (Pellegrini et al., 2021). In line with the previously outlined methodology, related Information Requirements (IRs) were defined and exported using the open IFC format, incorporating a customised property set. This property set included, for each material identified by its corresponding European Waste Code (EWC), the four required percentages of waste quantities: reused, recycled, landfilled hazardous, and landfilled non-hazardous. To ensure maximum flexibility and customisation, the “DefaultUserDefinedParameterSets” feature was utilised. This process requires a .txt file containing a tailored script to define the property set named “Waste Quantities.”

4.2 Smart contract and blockchain for the PoC

A smart contract was established to execute and store data on the blockchain, involving the previously identified stakeholders: the Public Client (Authority/PA), the Bidders, and the Judging Committee of experts. All participants are registered and verified through the smart contract, incorporating Know Your Customer (KYC) and Anti-Money Laundering (AML) protocols, as instructed by the client, regulators, and other relevant authorities.

Stakeholders interact with the system by sending transactions to the blockchain, making payments, transmitting data, and invoking the smart contract's functionalities. The process is divided into three phases, each with specified deadlines:

1. Privacy-preserving committal bidding
 2. Validation
- Automated determination of the winner and awarding.

(1) Privacy preserving committal bidding

The code of the smart contract that rules the bidding is shown in Figure 7 with some constraints checked and imposed about the bidder/project/min deposit/bidding phase being currently open, in red.



Figure 7. Description of the protocol and interface associated with the committal deposit for bidders

(2) Validation

Figure 8 shows a bidder sending an offer consisting of documents, quotes, and a deposit, then the bids are encrypted (hashed) to preserve the privacy and only the bidder will be able to reveal them later.



Figure 8. Description of the protocol and interface associated with the validation of experts evaluation

The deposit is binding, serving as evidence of the bidder's genuine interest (skin in the game) and discouraging undesirable behaviours, such as refusing to disclose unfavourable bids or submitting multiple bids to maximise chances and advantages (various strategies could be employed). Data are stored off-chain, as on-chain storage is costly, with an immutable link to the smart contract typically provided via IPFS. Various levels of privacy and accessibility can be configured.

The bidding phase concludes either at a predefined deadline or through an action by the client. The prototype includes an interface for voting experts and the smart contract code governing the experts' votes (Figure 9). It ensures that only recognised experts can interact with the contract, verifying their credentials. Additionally, it confirms the correctness of the vote format, the voting phase, and that each expert casts only one vote. Once a vote is submitted, it is recorded, and an event is emitted, visible to all stakeholders outside the blockchain.

When a bidder reveals their previously private and immutable bid, the smart contract ensures its consistency with the encrypted commitment made earlier, maintaining trust and transparency.

(3) Awarding notarization

The smart contract code uses the "quantified" bids for a deterministic and verifiable determination of the best bid, and the winner is awarded. Their cryptographic address (i.e. the proxy for their validated identity in the blockchain world) has been identified. In this phase, no bidder can propose new tenders and all bidders are incentivized to reveal their so-far private bids, no one can know or alter them, nor bidders can retract them. As soon as they reveal the bid, the contract pays them back the deposit and the experts can evaluate them, assigning a numerical score.

```
//calculate the best bid
for (uint i = 0; i < projectsBids[_ID].length; i++) {
    uint totalSum = 0;
    Vote[] storage votes = project_bids_votes[_ID][i];
    for (uint j = 0; j < votes.length; j++) {
        totalSum += votes[j].expertsVote.beautiful;
        totalSum += votes[j].expertsVote.smart;
        totalSum += votes[j].expertsVote.materials;
    }
    if (totalSum > highestVoted) {
        highestVoted = totalSum;
        Winner = projectsBids[_ID][i].bidder;
    }
}
// winner found
projectAssignment[_ID] = Winner;
projectsCreated[_ID].project_address_winner = Winner;
projectsCreated[_ID].project_assigned = true;

emit projectAssignedWinner(_ID, Winner);

return("Address of the winner: ", Winner);
```

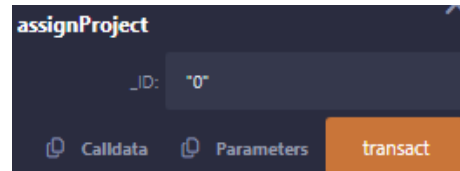


Figure 9. Description of the protocol and interface associated with the bidders' committal deposit

The justification for experts' scores can be recorded through a voting mechanism, or alternatively, an expert reputation framework could be established. A commit/reveal system, backed by a deposit, encourages proper behaviour from bidders, and appropriate strategies can be implemented to mitigate bias and arbitrariness in the evaluation of qualitative parameters (Figure 9).

5. CONCLUSIONS

This research proposes a methodology that integrates digital technologies to achieve comprehensive digitalisation and introduces automated processes to streamline traditionally manual and subjective tasks. By avoiding time-consuming and unreliable methods, the approach enhances transparency, reliability, and objectivity. Improved production, review, and validation of information through optimised information management ensure higher compliance with Client and regulatory Information Requirements (IRs). This reduces errors, omissions, reworking, and associated costs, while also mitigating risks of corruption. The methodology enables straightforward and rapid verification of CAM or other environmental protocol applications, facilitating the adoption of Green Public Procurement (GPP) and lowering the overall project impact.

Blockchain and smart contracts contribute significantly by enabling secure tracking of information transactions, ensuring reliability and confidentiality. They also allow for the automated verification, evaluation, and awarding of bids according to compliance standards defined by the client. This automation extends to payment processes, improving the efficiency of Public Procurement (PP). Furthermore, the need for a central trusted authority or intermediaries is eliminated, reducing opportunities for corruption.

Nevertheless, the methodology faces several challenges. These include defining IRs and IFC Property Sets for evaluation criteria and contract terms, which requires managing large datasets and normalising procedures in a machine-readable format. Transitioning to a distributed, process-driven approach with full transparency among stakeholders also poses challenges, particularly given the sector's close-knit nature and concerns over privacy. Additional obstacles involve automating smart contract generation with high reliability and addressing the lack of IT skills within Public Clients and the construction industry. The absence of updated regulations and standards further complicates adoption.

Another critical consideration is mitigating the environmental and energy impacts of blockchain technology. This may involve leveraging Sustainable High-Performance Computing (HPC) centres, such as the one under study at the University of Turin. These centres would utilise renewable energy sources, such as solar power, and implement waste heat reuse systems to minimise carbon footprints.

Future research will focus on addressing these challenges to advance the adoption of sustainable, efficient, and transparent digital solutions within Public Procurement and construction industry.

REFERENCES

- BPMN, Business Process Model and Notation, Version 2.0 Standard, Object Management Group, OMG (2011).
- European Commission, EU Single Market Scoreboard - Public Procurement, 2022
- European Commission, Making public procurement work in and for Europe, 2017
- Guu, K., Lee, K., Tung, Z., Pasupat, P., Chang, M., (2020), Retrieval augmented language model pre-training. In: International conference on machine learning. pp. 3929–3938. PMLR.
- Kim, M., & Kim, Y. W. (2024). Applications of blockchain for construction project procurement. *Automation in Construction*, 165, 105550.
- Lavagna, M., Bessi, A., Meneghelli, A., & Moschini, P. (2019). The environmental dimension of detailed design. Experiences and future perspectives. *TECHNE-Journal of Technology for Architecture and Environment*.
- Meschini, S., Di Giuda, G. M., Tagliabue, L. C., Locatelli, M., & Pellegrini, L. (2023). BPMN 2.0 to redefine Italian design-bid procurement in an innovative model-based, open-source approach. *EC3 conference*, 4, 1-8. doi 10.35490/EC3.2023.327
- Nai R., Sulis E., Fatima I., & Meo R. (2024), Large Language Models and Recommendation Systems: A Proof-of-Concept Study on Public Procurements. *NLDB Proceedings*.
- Nguyen, C. T., Liu, Y., Du, H., Hoang, D. T., Niyato, D., Nguyen, D. N., & Mao, S. (2024), Generative ai-enabled blockchain networks: Fundamentals, applications, and case study. *IEEE Network*.
- OECD G20-Compendium-of-Good-Practicesin-Infrastructure-Development., 2019
- OAV, Osservatorio Appalti Verdi, (2022). I numeri del Green Public Procurement in Italia; Legambiente: Rome, Italy.
- Pattini, G., Di Giuda, G.M., and Tagliabue, L.C. (2022), The Integration of Automatic BIM Validation and Smart Contracts for Design Compliance and Payment Reliability in the Design Process, in *Blockchain for Construction*, T. Dounas and D. Lombardi, Eds. Springer Nature, pp. 47–73.
- Pellegrini, L., Locatelli, M., Meschini, S., Pattini, G., Seghezzi, E., Tagliabue, L.C., Di Giuda, G.M. (2021) Information Modelling Management and Green Public Procurement for Waste Management and Environmental Renovation of Brownfields, *Sustainability* 13, no. 15: 8585. <https://doi.org/10.3390/su13158585>
- Petrović, N., & Al-Azzoni, I. (2023), Model-driven smart contract generation leveraging ChatGPT. In *International Conference On Systems Engineering* (pp. 387-396). Cham: Springer Nature Switzerland.
- Robertson, S., Zaragoza, H., et al. (2009), The probabilistic relevance framework: Bm25 and beyond. *Foundations and Trends® in Information Retrieval* 3(4), 333–389.
- Sapir, A., Schraepen, T., & Tagliapietra, S. (2022). Green public procurement: A neglected tool in the European green deal toolbox?. *Intereconomics*, 57(3), 175-178.
- Sreckovic, M., Sibenik, G., and Breitfuß, D., (2022), Capturing and Transforming Planning Processes for Smart Contracts, in *Blockchain for Construction*, T. Dounas and D. Lombardi, Eds., pp. 75–88.

ENHANCING RESILIENCE AND ADAPTABILITY OF INFRASTRUCTURE TRANSMISSION AND DISTRIBUTION NETWORKS: A REVIEW STUDY

Mohammad Saleh Nikoopayan Tak¹, Reyhaneh Moazami Goodarzi², and Yanxiao Feng³

1) Ph.D. Student, Hillier College of Architecture and Design, New Jersey Institute of Technology, Newark, NJ, USA. Email: mn552@njit.edu

2) Graduate Student, School of Architecture and Environmental Design, Iran University of Science and Technology, Tehran, Iran. Email: rayhaneh.moazami@gmail.com

3) Ph.D., Assistant. Prof., School of Applied Engineering and Technology, New Jersey Institute of Technology, Newark, NJ, USA. Email: yf43@njit.edu

Abstract: The growing frequency and intensity of climate-related events and natural disasters present substantial challenges to the resilience and adaptability of critical infrastructure, particularly electricity transmission and distribution networks. This study provides a review of existing literature and incorporates recent research findings to identify the primary factors influencing resilience and adaptability within these networks. The study emphasizes the importance of key areas including technical design strategies, infrastructure investments, facility design considerations, organizational capabilities, operational strategies, and supply chain factors. The findings offer essential insights for stakeholders in the energy sector aiming to enhance the resilience of transmission and distribution networks against climate change impacts and natural disasters. Additionally, the study underscores the importance of establishing standardized resilience metrics and advocates for future research focusing on cost-benefit analyses and data-driven approaches to predict and mitigate cascading failures and high-impact, low-probability (HILP) events.

Keywords: Electricity transmission and distribution resilience, Infrastructure adaptability, Climate-related natural disasters, Resilience assessment metrics

1. INTRODUCTION

Climate change has intensified the frequency and severity of natural disasters which significantly impact critical infrastructure systems. The challenge of high-impact low-probability (HILP) events, such as wildfires, earthquakes, windstorms, and floods, is highlighted by Surinkaew et al. (2022). These events pose significant threats to power systems, and the uncertainties they introduce necessitate comprehensive and accurate modeling techniques to improve resiliency. Electricity transmission and distribution (T&D) networks are particularly vulnerable, which can result in prolonged power outages that cause substantial economic losses and threaten public health and safety (Ekisheva et al., 2020; Fant et al., 2020). Enhancing the resilience and adaptability of these networks is essential to ensure a reliable electricity supply and to mitigate the adverse effects of climate-related events. Recent literature indicates a growing recognition of the challenges posed by climate change to T&D networks. The traditional reliability measures may not suffice for such extreme events, and there is a need for advanced modeling approaches that can capture the complexities associated with HILP events.

Addressing vulnerabilities and potential attacks is a significant concern. Valencia et al. (2021) reviewed methodologies for assessing the vulnerability of power systems using multilevel programming. They point out that while most research focuses on transmission systems using linear approximations, there is a need for more comprehensive models that can address vulnerabilities in distribution networks and incorporate defense strategies such as distributed generation and demand response. This aligns with the findings of Erenoglu et al. (2024), who stress the importance of distinguishing between reliability and resiliency. While reliability focuses on the system's ability to perform under normal conditions, resiliency pertains to the system's capacity to adapt and recover from extreme disruptions. Developing quantitative metrics for resiliency, as they suggest, is essential for better understanding and enhancing this emerging concept.

Studies have explored various aspects of resilience, including design strategies, the integration of advanced technologies, and organizational as well as operational aspects. This review study addresses key gaps by providing a comprehensive analysis of the challenges and critical factors affecting the resilience of T&D networks in response to natural disasters. By integrating insights from foundational studies and recent literature, this review aims to identify the challenges and critical factors that influence the resilience of transmission networks in the face of natural disasters. It also underscores the importance of resilience and adaptability in T&D networks, highlighting the need for standardized resilience metrics to improve assessment and response strategies under adverse conditions.

2. REVIEW METHOD

To ensure a thorough understanding of the subject matter, a selective literature review between 2017 to 2024 was conducted to include both foundational studies from earlier years and recent literature that reflects the latest advancements and challenges in the field. Databases such as IEEE Xplore, ScienceDirect, and Web of

Science were utilized. Keywords included "infrastructure resilience," "transmission networks," "distribution networks," "climate change," and "natural disasters."

Full-text analyses were then performed on the selected papers to extract detailed information on the factors influencing resilience, assessment metrics, and enhancement strategies. Emphasis was placed on peer-reviewed articles and authoritative reports that offered empirical data, case studies, or comprehensive reviews pertinent to the study's objectives. The selected studies cover a wide range of topics, including cyber-physical vulnerability, resilience metrics, renewable energy integration, outage management, and the impact of extreme weather events, which provide a robust foundation for this review. A summary of the selected papers and their key contributions is presented in Table 1.

Table 1. Summary of Key Studies on Infrastructure Resilience and Adaptability

Authors	Year	Focus of Study	Key Findings
Xu et al.	2021	Cyber-physical vulnerability and resilience of power systems	Proposed a framework to identify CPPS key features and emphasize the resilience techniques.
Dwivedi et al.	2024	Power system resilience and enhancement techniques	Presented power system resilience metrics, explored data-driven methods, and analyzed recent enhancements.
Gasser et al.	2019	Resilience definitions and assessment methods for energy systems	Classified literature by resilience approaches and stressed minimizing disruptive event impacts.
Serrano-Fontova et al.	2023	Fragility curves for resilience assessments in power systems	Classified fragility curves and compared results to show relevance in assessments.
Cicilio et al.	2021	Power system resilience and renewable energy integration	Emphasized new planning and operations to address uncertainties in resilience and renewable integration.
Daeli et al.	2022	Power grid resilience against extreme weather events	Outlined infrastructural strategies including grid hardening, redundancy, and adaptive operation.
Peng et al.	2023	Renewable energy integration and grid resilience	Emphasized advanced control algorithms and cybersecurity for grid stability and resilience.
Shittu et al.	2021	Electricity markets and resilience	Called for market redesign to address renewables, prosumers, and increasing disasters.
Xing	2020	Cascading failures in IoT systems	Discussed causes, models, and resilience strategies for cascading failures in IoT systems.
Surinkaew et al.	2022	Resilience of power systems to high-impact low-probability events	Proposed a resilience enhancement framework for power systems that incorporates comprehensive metrics and cost-effective strategies.
Venkateswaran et al.	2023	Quantitative analysis of power system resilience research	Identified research trends and emerging areas in power system resilience through bibliometric and correlation analysis.
Erenoğlu et al.	2024	Power system resiliency and implementation aspects	Distinguished resiliency from reliability and developed a comprehensive assessment framework.
Hawker et al.	2024	Strategies for managing extreme weather impacts on electricity grids	Recommended strategies for grid operators to enhance reliability against extreme weather events.
Voropai	2020	Transformations in electric power systems	Discussed prospects in power system transformations emphasizing flexibility, resilience, and survivability.
Valencia et al.	2021	Vulnerability assessment of power systems	Reviewed approaches and models for evaluating and mitigating power system vulnerability to attacks and failures.

Kumar et al.	2024	Challenges and opportunities in multi-microgrid operations	Analyzed methodologies for optimizing multi-microgrid operations, including energy dispatch, market strategies, and the role of blockchain technology.
Hossain et al.	2021	Grid resilience and reliability metrics and strategies	Proposed resilience risk factor and grid infrastructure density.
Almaleh	2023	Resilience models for Smart Interdependent Critical Infrastructures (Smart ICIs)	Provided a comprehensive evaluation of resilience models and measurements for Smart ICIs.
Vugrin et al.	2017	Resilience metrics for the electric power system	Developed Grid Resilience Analysis Process (RAP) for managing disruptions.
Malek et al.	2023	Outage management in distribution systems	Proposed power outage management strategy and resilience metrics.
Panteli et al.	2017	Metrics and quantification of operational and infrastructure resilience in power systems	Proposed framework enabling in-depth understanding of resilience level.
Amani et al.	2021	Vulnerability of power grids using complex network analysis and centrality metrics	Showed the applicability of centrality measures in identifying vulnerable points in power grids.
Mohanty et al.	2024	Power system resilience and strategies for a sustainable infrastructure	Developed a resilience metric grounded in the social welfare of power grid and water systems.
Rickerson et al.	2024	Value of resilience in distributed energy resources	Found no standardized approach for valuing resilience investments.
Anderson et al.	2020	Integrating the value of electricity resilience in energy planning and operations decisions	Demonstrated that incorporating duration-dependent resilience value can reduce energy lifecycle costs.
Lonergan et al.	2023	Ensuring resilient energy system infrastructure	Proposed a framework for insurers to promote resilient energy infrastructure through policies.

3. REVIEW RESULTS

The review process and results provide an overall understanding of infrastructure resilience in T&D networks, with key aspects summarized in Figure 1.

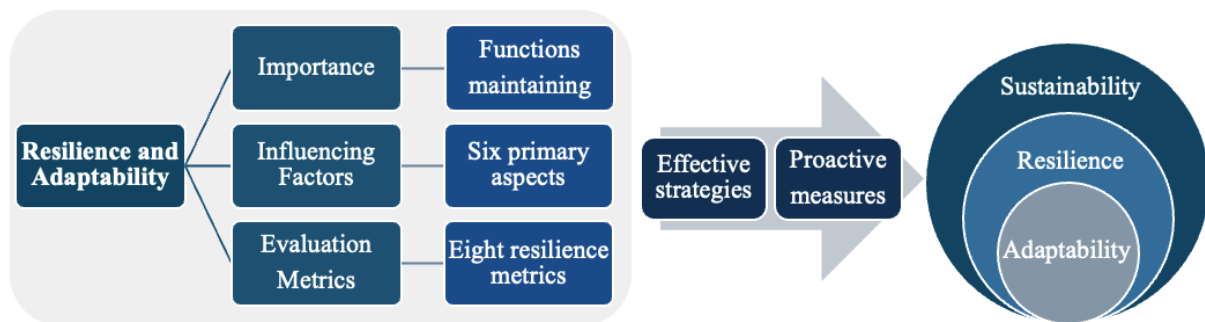


Figure 1. Summary of aspects of infrastructure resilience for the review study.

3.1 Importance of Infrastructure Resilience and Adaptability

The resilience and adaptability of electricity T&D networks are critical in the face of intensifying climate change impacts and natural disasters. These networks are fundamental to modern society and play a crucial role in delivering electricity continuously to homes, businesses, and essential services. Resilient T&D infrastructure can withstand disruptions and recover quickly while maintaining essential functions during extreme events. This is vital for minimizing economic losses and safeguarding public health and safety (Francis & Bekera, 2014).

Severe weather events such as hurricanes, wildfires, and ice storms pose significant risks to power systems and cause damage that leads to widespread outages (Li et al., 2016). The ability of T&D networks to resist and recover from such events reduces the duration and extent of power outages. Moreover, resilient electricity networks enable faster recovery of other dependent systems, such as communications, transportation, and

healthcare, which emphasizes the interconnected nature of modern infrastructure (Kandaperumal & Srivastava, 2020; Huang et al., 2021).

Adaptability complements resilience by allowing T&D networks to adjust to evolving conditions and incorporate new technologies. This is increasingly important as the energy sector integrates renewable energy sources, distributed generation, and smart grid technologies. Adaptable systems better manage the variability of renewable energy, optimize energy flows, and enhance system flexibility (Wang et al., 2015). Investing in the resilience and adaptability of T&D networks not only mitigates substantial economic losses from power outages (Sanstad et al., 2020) but also ensures that critical services remain operational during disasters. Proactively enhancing these qualities is essential to prepare for and mitigate the impacts of future climate-related events.

3.2 Factors Influencing Resilience and Adaptability

Resilience and adaptability in electricity T&D networks are influenced by a complex interplay of technical, organizational, operational, and supply chain factors. These factors can be categorized into six primary aspects: technical design strategies, infrastructure investments, facility design strategies, organizational capabilities, operational strategies, and supply chain factors.

(1) Technical Design Strategies

Technical design strategies play a vital role in enhancing the resilience of power systems. One of the key strategies involves the implementation of modular and decentralized architectures. By integrating distributed energy resources (DERs) into decentralized grids, systems can isolate failures more effectively and enable quicker restoration of services. This approach not only enhances resilience but also improves flexibility and scalability (Wang et al., 2015). Additionally, incorporating infrastructure redundancy, such as redundant lines and spare equipment, mitigates the risk of failures and facilitates faster recovery when disruptions occur. Redundancy serves as a safety net by providing alternative pathways and resources to maintain system functionality (Haimes, 2018). The advancement of monitoring and control technologies further strengthens technical resilience. Smart grid technologies, for instance, significantly improve situational awareness by providing real-time data and analytics. This heightened awareness allows for rapid response to disruptions, minimizing downtime and preventing the escalation of issues (Fan et al., 2021). Moreover, hardening control centers against physical and cyber threats is essential to ensure continuous operation during extreme events. By reinforcing these critical hubs, power systems can maintain command and control functions even under adverse conditions, which ensures the continuity of essential services (Gao et al., 2017).

(2) Infrastructure Investments

In terms of infrastructure investments, proactive measures are crucial for enhancing resilience. Grid hardening initiatives involve upgrading components to meet higher standards, which improves their resistance to extreme weather events. Such investments in robust materials and designs can significantly reduce the vulnerability of the infrastructure (Lin et al., 2017). Undergrounding power lines is another investment strategy that reduces exposure to weather-related damages. Burying overhead lines makes the system less vulnerable to wind, ice, and falling debris, which reduces the frequency of outages (Maliszewski & Perrings, 2012). Furthermore, upgrading aging infrastructure enhances overall system robustness by replacing outdated components with modern, more resilient ones (Ouyang & Dueñas-Osorio, 2012). Deploying distributed energy resources (DERs), such as microgrids, provides additional layers of resilience by offering backup power sources during outages, which ensures continuity of supply to critical loads (Hamidieh & Ghassemi, 2022; Hossain et al., 2016).

(3) Facility Design Strategies

Facility design strategies contribute significantly to resilience by addressing physical vulnerabilities. Elevation adjustments, such as raising equipment above known flood levels, protect critical assets from inundation during flood events (Francis & Bekera, 2014). Implementing flood protection measures, such as barriers and watertight seals, can limit the impact of flooding on facilities and help reduce damage and downtime (Wilkinson et al., 2019). Designing structures to withstand high winds and wildfires enhances their durability against such extreme events. Incorporating wind resistance and fireproofing measures into facility designs ensures that they remain operational or recover quickly after such incidents (Ouyang & Dueñas-Osorio, 2014). Securing equipment through proper anchoring prevents detachment and displacement during extreme events, which is essential for maintaining operational integrity (Francis & Bekera, 2014). Considering multi-hazard designs that address multiple threats simultaneously can improve overall resilience by providing comprehensive protection against various environmental challenges (Ouyang & Dueñas-Osorio, 2014). Moreover, resilience and adaptability strategies can vary depending on whether facilities are located in urban or rural areas. Urban facilities, characterized by higher population densities and complex infrastructure interdependencies, may require advanced monitoring systems and enhanced cybersecurity measures. In contrast, rural facilities might prioritize physical protection against environmental hazards and ensure connectivity despite geographical isolation.

(4) Organizational Capabilities

The development of strong organizational capabilities is essential for effective resilience planning and response. Conducting structured risk assessments and management processes allows organizations to prioritize

investments based on identified vulnerabilities and potential impacts (Chang et al., 2014). Effective emergency planning and response strategies ensure that restoration efforts are coordinated efficiently, which minimizes the time required to return to normal operations. Cross-sector collaboration enhances resource pooling and information sharing during emergencies, which can be critical for managing complex incidents (Aldrich & Meyer, 2015). Enhancing situational awareness through advanced analytics and real-time data collection improves an organization's ability to make informed decisions during crises.

(5) Operational Strategies

Effective operational strategies are essential for resilience. Controlled islanding and reconfiguration techniques involve isolating sections of the grid to prevent widespread failures, which helps contain issues and maintain service in unaffected areas (Wang et al., 2015). Demand response programs that manage consumer load can enhance grid flexibility by reducing demand during peak times or emergencies, thus alleviating stress on the system (Siano, 2014). Operating microgrids autonomously during disruptions ensures continuous power supply to critical facilities and can expedite the restoration process (Panteli et al., 2017). Moreover, investing in personnel training and ensuring the availability of skilled staff improve emergency response efficiency. Prepared and knowledgeable personnel are better equipped to handle crises effectively (Arab et al., 2015).

(6) Supply Chain Factors

Supply chain factors play a significant role in the resilience of T&D networks. Maintaining strategic reserves of spare parts enables quicker restoration by reducing downtime associated with sourcing and delivering critical components (Rose & Wei, 2013). Adopting multi-sourcing strategies, where organizations diversify their suppliers, reduces the risks associated with dependence on single sources and can mitigate supply chain disruptions (Golschmidt, 2021). Emphasizing flexibility and regional sourcing allows supply chains to respond more effectively to disruptions by leveraging local resources and reducing reliance on long-distance transportation (Whitney, 2014). Evaluating supplier resilience by assessing their capabilities and preparedness ensures the continuity of critical supplies during adverse events.

3.3 Resilience Metrics in Smart Grids and Transmission Systems

Assessing the resilience of smart grids and transmission systems requires robust and adaptable metrics that capture both the system's ability to withstand disruptions and its capacity for rapid recovery. Various metrics have been proposed in the literature to quantify different aspects of resilience. A comparison of different resilience metrics is presented in Table 2.

The time to recovery (TTR) metric evaluates how quickly a system can return to normal operation after a disruption (Hossain et al., 2021). This metric is crucial for assessing the effectiveness of restoration strategies and the system's rapidity in responding to adverse events. A shorter TTR indicates a more resilient system capable of minimizing downtime and associated losses. Another comprehensive measure is the resilience index (RI), which combines factors such as robustness, resourcefulness, redundancy, and rapidity to provide an overall assessment of system resilience (Vugrin et al., 2017; Almaleh, 2023). The RI offers a comprehensive evaluation by considering multiple dimensions of resilience, including preventive and restorative capabilities. Hossain et al. (2021) introduced novel metrics such as the resilience risk factor and grid infrastructure density. The resilience risk factor accounts for a region's susceptibility to disasters and its resilience score, which indicates the associated risk within a particular grid. Grid infrastructure density relates to the population and economic activity in an area. It serves as a parameter to determine grid resilience based on the concentration of critical infrastructure. In the context of distribution systems, Malek et al. (2023) developed improved resilience metrics aimed at enhancing outage management. These metrics consider priority loads and optimize the deployment of resources such as mobile emergency generators to maximize system resiliency. By integrating these metrics into their optimization model, they demonstrated effective strategies for restoring power and improving resilience during outages.

Table 2. Comparison of Resilience Metrics for Smart Grids and Transmission Systems

Metric	Definition	Components	Advantages	Limitations
Resilience Triangle	Tracks performance loss and recovery over time	Performance loss and recovery time	Simple visualization	Oversimplifies complex systems
Resilience Trapezoid	Extends the triangle with preparation phases	Preparation, impact, recovery, and adaptation	Detailed phase understanding	Complex; requires additional data
Time to Recovery (TTR)	Time taken to return to normal operations	Recovery duration	Easy to interpret and compare	Does not capture performance during recovery
Resilience Index (RI)	Combines robustness, redundancy,	Multiple resilience dimensions	Holistic assessment	Complex; needs extensive data

Resilience Risk Factor	resourcefulness, and rapidity Assesses regional disaster susceptibility and resilience	Disaster risk and regional indicators	Considers regional differences	Requires detailed regional data
Grid Infrastructure Density	Measures resilience based on population and economic activity	Population density and economic measures	Reflects critical infrastructure concentration	May miss other resilience aspects; less effective in rural areas
Improved Metrics for Outage Management	Considers priority loads and resource deployment	Priority loads and resource allocation	Optimizes restoration efforts	System-specific and not widely applicable
Complex Network Theory Metrics	Uses graph theory to assess vulnerabilities	Centrality measures and network topology	Identifies critical nodes and links	May oversimplify and requires detailed data

One widely recognized concept in the civil infrastructure is the resilience triangle, which illustrates the degradation of system performance during a disruption and the subsequent recovery over time (Bocchini et al. 2014; Hossain et al. 2021). This graphical representation helps in understanding the relationship between the initial impact of a disaster and the time required for recovery. Panteli et al. (2017) further extended this concept by introducing the resilience trapezoid, which provides a more detailed depiction of the different phases a power system undergoes during extreme events. Both concepts, as illustrated in Figure 2, help in understanding the relationship between system disruption and recovery.

Despite the availability of various metrics, challenges persist in standardizing them due to differing definitions of resilience and the inherent complexity of power systems (Almaleh, 2023; Amani & Jalili, 2021). The diversity of methodologies and the lack of universally accepted metrics hinder the ability to compare resilience levels across different systems and studies. Also, integrating these metrics into practical applications involves addressing issues related to data availability, modeling complexities, and the dynamic nature of smart grids.

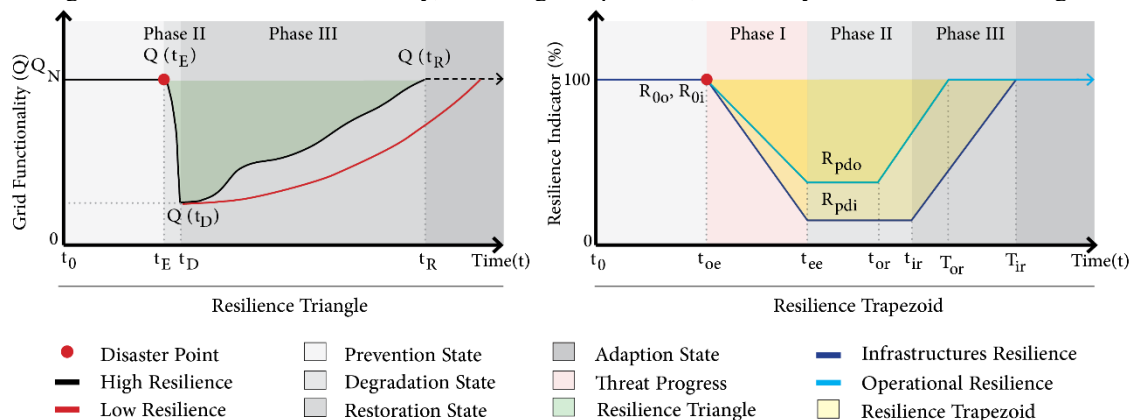


Figure 2. Illustration of the resilience triangle and resilience trapezoid concepts

3.4 Strategies for Enhancing Infrastructure Resilience

Technological challenges, regulatory hurdles, and financial constraints often impede the adoption of resilience measures. For example, undergrounding power lines is effective but cost-prohibitive in many regions (Sanstad et al., 2020). Additionally, the lack of standardized approaches for valuing resilience investments complicates decision-making for policymakers and utilities (Rickerson et al., 2024; Sanstad et al., 2020). Enhancing infrastructure resilience involves a multi-faceted approach.

Data-driven techniques and advanced analytics are identified as critical tools for enhancing resilience. The potential of data-driven methods and machine learning in improving power system resilience has been recommended. It is suggested that integrating these techniques can enhance predictive analytics and improve resilience strategies. The study by Xing (2020) noted that simulations and modeling are essential for understanding cascading failures and developing mitigation strategies.

Grid hardening has proven effective in reducing damage from extreme weather, particularly in hurricane-prone areas (Litalien, 2020). For instance, hardening transmission lines reduces service outages and promotes greater social equality by minimizing service disruptions for vulnerable populations. At the same time, implementing renewable energy sources and microgrids not only enhances resilience but also contributes to sustainability goals by reducing greenhouse gas emissions (Mohanty et al., 2024; Rickerson et al., 2024).

Microgrids and distributed energy resources provide supplementary electrical power during service disruptions. They support essential infrastructures, which promotes both resilience and sustainability (Mohanty et al., 2024).

The importance of organizational capabilities and regulatory frameworks in enhancing resilience cannot be overstated. Hawker et al. (2024) discuss the strategies adopted by system operators internationally to manage extreme weather impacts on electricity grids. This notion is echoed by Shittu and Santos (2021), who highlight the need for market redesigns that accommodate resilience, especially considering the increasing penetration of renewable energy sources and the emergence of prosumers.

Investments in resilience can lead to long-term savings by preventing costly outages. Cost-benefit analyses have shown that every dollar invested in resilience can save multiple dollars in disaster recovery (Anderson et al., 2020; Sanstad et al., 2020). Moreover, integrating the duration-dependent value of resilience into energy planning and operations can optimize investment decisions and reduce the lifecycle cost of energy (Anderson et al., 2020).

4. DISCUSSION

4.1 The Role of Emerging Technologies

Resilience of electricity networks, increasingly vital due to climate-related disasters, involves complex technical, organizational, and operational challenges (Ouyang & Dueñas-Osorio, 2014; Wang et al., 2015). The integration of advanced technologies and innovative concepts is a key aspect discussed in the literature. Voropai (2020) examines the transformations in electric power systems driven by innovative technologies and digitalization. The author emphasizes that the future of electric power systems will require greater flexibility, resiliency, and survivability, which calls for new approaches to system modeling, control, and operation. This perspective is supported by Kumar et al. (2024), who explored the role of multi-microgrid (MMG) operations in enhancing the resilience of smart distribution networks. They emphasize that MMGs, supported by robust information and communication technologies, can optimize energy dispatch and contribute to the resilience of power systems through energy sharing and trading.

A critical theme emerging from the literature is the need to address cascading failures within interconnected systems. Xing (2020) emphasizes that in the context of the Internet of Things (IoT) and smart grids, cascading failures can have profound impacts on system reliability and resilience. The integration of IoT devices and smart technologies in T&D networks introduces new vulnerabilities, which can trigger cascading failures and result in widespread outages. Understanding and modeling these failures are crucial for developing mitigation strategies and building sustainable, resilient power systems.

Moreover, the integration of renewable energy sources and the shift towards decentralized energy systems present both opportunities and challenges for resilience. The rise of microgrids and distributed energy resources offers potential for increased system flexibility and resilience (Hossain et al., 2016; Kumar et al., 2024). However, this also introduces new complexities in system operation and control, which necessitates advanced technologies and reliable control algorithms (Voropai, 2020; Peng et al., 2023).

4.2 The Challenges and Future Directions

Despite the advancements and proposed strategies, several gaps and challenges remain. In examining the broader research domain, Venkateswaran and Panteli (2023) provide a bibliometric and correlation analysis of power system resilience research from 2001 to 2022. Their study reveals that while there has been a significant increase in research output, certain natural hazards such as earthquakes and floods are underrepresented in resilience studies. This gap suggests that more attention is needed in these areas to develop comprehensive resilience strategies that address a wider range of potential threats.

There is a need for standardized resilience metrics and frameworks that can guide the assessment and enhancement of resilience across different systems and contexts (Erenoğlu et al., 2024). Although numerous strategies exist, their implementation is often hindered by financial and regulatory barriers. A comprehensive evaluation reveals the need for integrated approaches that consider technical feasibility, economic viability, and policy support. Additionally, human and institutional factors, such as organizational adaptability, cross-sector collaboration, and effective emergency planning, are essential components requiring further research and development (Aldrich & Meyer, 2015).

5. CONCLUSIONS

The increasing threat of climate change requires proactive measures to enhance the resilience and adaptability of electricity T&D networks. However, gaps remain in understanding the comprehensive factors influencing resilience and in developing holistic strategies that address both infrastructural and operational dimensions. Significant barriers such as financial constraints and regulatory hurdles also impede widespread implementation. This study highlights key factors influencing resilience, including technical design, infrastructure investments, facility design, organizational capabilities, operational strategies, and supply chain management. By integrating these factors into planning and operations, stakeholders can develop more resilient systems capable of withstanding and recovering from natural disasters. There is also a need for updated evaluations of the metrics

used to assess resilience levels in modern smart grids and transmission systems, considering the rapid evolution of technologies and the increasing complexity of power systems (Kumar et al., 2024). Future research is suggested to focus on cost-benefit analyses of resilience measures, the development of standardized metrics for resilience assessment, and the exploration of data-driven techniques and machine learning algorithms to predict and mitigate cascading failures and HILP events.

REFERENCES

- Aldrich, D. P., & Meyer, M. A. (2015). Social capital and community resilience. *American Behavioral Scientist*, 59(2), 254–269.
- Almaleh, A. (2023). Measuring Resilience in Smart Infrastructures: A Comprehensive Review of Metrics and Methods. *Applied Sciences*, 13(11), 6452. <https://doi.org/10.3390/app13116452>
- Amani, A.M., Jalili, M. (2021). Power Grids as Complex Networks: Resilience and Reliability Analysis. *IEEE Access*, 9. <https://doi.org/10.1109/access.2021.3107492>
- Amani, A. M., & Jalili, M. (2021). Power Grids as Complex Networks: Resilience and Reliability Analysis. *IEEE Access*, 9, 125852–125877. <https://doi.org/10.1109/ACCESS.2021.3107492>
- Anderson, K., Li, X., Dalvi, S., Ericson, S., Barrows, C., Murphy, C. C., & Hotchkiss, E. (2020). Integrating the Value of Electricity Resilience in Energy Planning and Operations Decisions. *IEEE Systems Journal*, 15(3), 3624–3635. <https://doi.org/10.1109/JSYST.2019.2961298>
- Arab, A., Khodaei, A., Han, Z., & Khator, S. K. (2015). Proactive recovery of electric power assets for resiliency enhancement. *IEEE Access*, 3, 99–109.
- Bocchini, P., Frangopol, D. M., Ummenhofer, T., & Zinke, T. (2014). Resilience and sustainability of civil infrastructure: Toward a unified approach. *Journal of Infrastructure Systems*, 20(2), 04014004.
- Chang, S. E., McDaniels, T., Fox, J., Dhariwal, R., & Longstaff, H. (2014). Toward disaster-resilient cities: Characterizing resilience of infrastructure systems with expert judgments. *Risk analysis*, 34(3), 416–434.
- Cicilio, P., Glennon, D., Máté, Á., Barnes, A. K., Chalishazar, V., Cotilla-Sanchez, E., ... Kapourchali, M. H. (2021). Resilience in an Evolving Electrical Grid. *Energies*, 14. <https://doi.org/10.3390/en14030694>
- Eastman, C., Teicholz, P., Sacks, R., & Liston, K. (2011). *BIM Handbook: A Guide to Building Information Modeling for Owners, Managers, Designers, Engineers, and Contractors* (2nd ed.). John Wiley & Sons.
- Daeli, A., & Mohagheghi, S. (2022). Power grid infrastructural resilience against extreme events. *Energies*, 16(1), 64.
- Dwivedi, D., Mitikiri, S. B., Babu, K. V. S. M., Yemula, P. K., Srinivas, V. L., Chakraborty, P., & Pal, M. (2024). Technological advancements and innovations in enhancing resilience of electrical distribution systems. *International Journal of Critical Infrastructure Protection*. <https://doi.org/10.1016/j.ijcip.2024.100696>
- Ekisheva, S., Papic, M., Pakeltis, M. J., Tillis, G. B., McClure, M., & King, D. J. (2020). Assessment of North American transmission outages by fault type. In 2020 IEEE Power & Energy Society General Meeting.
- Erenoglu, A. K., Şengör, İ., & Erdinç, O. (2024). Power system resiliency: A comprehensive overview from implementation aspects and innovative concepts. *Energy Nexus*, 15, 100311.
- Fan, D., Ren, Y., Feng, Q., Liu, Y., Wang, Z., & Lin, J. (2021). Restoration of smart grids: Current status, challenges, and opportunities. *Renewable and Sustainable Energy Reviews*, 143, 110909.
- Fant, C., Boehlert, B., Strzepek, K., Larsen, P., White, A., Gulati, S., ... & Martinich, J. (2020). Climate change impacts and costs to US electricity transmission and distribution infrastructure. *Energy*, 195, 116899.
- Francis, R., & Bekera, B. (2014). A metric and frameworks for resilience analysis of engineered and infrastructure systems. *Reliability Engineering & System Safety*, 121, 90–103.
- Gao, H., Chen, Y., Mei, S., Huang, S., & Xu, Y. (2017). Resilience-oriented pre-hurricane resource allocation in distribution systems considering electric buses. *Proceedings of the IEEE*, 105(7), 1214–1233.
- Gasser, P., Lustenberger, P., Cinelli, M., Kim, W., Spada, M., Burgherr, P., ... Sun, T. Y. (2019). A review on resilience assessment of energy systems. *Sustainable and Resilient Infrastructure*, 6. <https://doi.org/10.1080/23789689.2019.1610600>
- Goldschmidt, K., Kremer, M., Thomas, D. J., & Craighead, C. W. (2021). Strategic sourcing under severe disruption risk: Learning failures and under-diversification bias. *Manufacturing & Service Operations Management*, 23(4), 761–780.
- Haimes, Y. Y. (2018). *Modeling and managing interdependent complex systems of systems*. John Wiley & Sons.
- Hamidieh, M., & Ghassemi, M. (2022). Microgrids and resilience: A review. *IEEE Access*, 10, 106059–106080.
- Hawker, G., Bell, K., Bialek, J., & MacIver, C. (2024). Management of extreme weather impacts on electricity grids: An international review. *Progress in Energy*, 6.
- Hossain, E., Roy, S., Mohammad, N., Nawar, N., & Dipta, D. R. (2021). Metrics and Enhancement Strategies for Grid Resilience and Reliability During Natural Disasters. *Applied Energy*, 290, 116709. <https://doi.org/10.1016/j.apenergy.2021.116709>

- Hossain, M. A., Madloul, N. A., Rahim, N. A., Selvaraj, J., Pandey, A. K., & Khan, A. F. (2016). Role of smart grid in renewable energy: An overview. *Renewable and Sustainable Energy Reviews*, 60, 1168–1184.
- Huang, H., Mao, Z., Panyam, V., Layton, A., & Davis, K. (2021). An ecological robustness-oriented approach for power system network expansion. *arXiv preprint arXiv:2107.06178*.
- Kandaperumal, G., & Srivastava, A. K. (2020). Resilience of the electric distribution systems: Concepts, classification, assessment, challenges, and research needs. *IET Smart Grid*, 3(2), 133–143.
- Kumar, A., Singh, A. R., Raghav, L. P., Deng, Y., He, X., Bansal, R. C., ... & Naidoo, R. M. (2024). State-of-the-art review on energy sharing and trading of resilient multi microgrids. *iScience*, 27.
- Li, F. G., Pye, S., & Strachan, N. (2016). Regional winners and losers in future UK energy system transitions. *Energy Strategy Reviews*, 13, 11–31.
- Li, Y., He, L., Liu, F., Li, C., Cao, Y., & Shahidehpour, M. (2017). Flexible voltage control strategy considering distributed energy storages for DC distribution network. *IEEE Transactions on Smart Grid*, 10(1), 163–172.
- Lin, K. H. E., Lee, H. C., & Lin, T. H. (2017). How does resilience matter? An empirical verification of the relationships between resilience and vulnerability. *Natural Hazards*, 88, 1229–1250.
- Lin, M., & Dueñas-Osorio, L. (2012). Time-dependent resilience assessment and improvement of urban infrastructure systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 22(3).
- Litalien, Z. (2020). A review of methods to better predict and reduce the risk of hurricane damage to the energy sector.
- Lonergan, K. E., Greco, S. F., & Sansavini, G. (2023). Ensuring/insuring resilient energy system infrastructure. *Environment Systems & Decisions*, 43. <https://doi.org/10.1007/s10669-023-09928-9>
- Malek, A. F., Mokhlis, H., Mansor, N. N., Jamian, J. J., Wang, L., & Muhammad, M. A. (2023). Power Distribution System Outage Management Using Improved Resilience Metrics for Smart Grid Applications. *Energies*, 16(9), 3953. <https://doi.org/10.3390/en16093953>
- Maliszewski, P. J., & Perrings, C. (2012). Factors in the resilience of electrical power distribution infrastructures. *Applied Geography*, 32(2), 668–679.
- Mohanty, A., Ramasamy, A. K., Verayiah, R., Bastia, S., Dash, S. S., Cüce, E., ... & Soudagar, M. E. M. (2024). Power system resilience and strategies for a sustainable infrastructure: A review. *Alexandria Engineering Journal*, 105, 837–850. <https://doi.org/10.1016/j.aej.2024.06.092>
- Ouyang, M., & Dueñas-Osorio, L. (2014). Multi-dimensional hurricane resilience assessment of electric power systems. *Structural Safety*, 48, 15–24.
- Panteli, M., Mancarella, P., Trakas, D. N., Kyriakides, E., & Hatziaargyriou, N. D. (2017). Metrics and quantification of operational and infrastructure resilience in power systems. *IEEE Transactions on Power Systems*, 32(6), 4732–4742.
- Peng, F. Z., Liu, C., Li, Y., Jain, A. K., & Vinnikov, D. (2023). Envisioning the future renewable and resilient energy grids—A power grid revolution enabled by renewables, energy storage, and energy electronics. *IEEE Journal of Emerging and Selected Topics in Industrial Electronics*, 5.
- Policy, E. S., & Planning. (2023). Resilient Electric Grid.
- Rickerson, W., Gillis, J., & Bulkeley, M. (2024). The Value of Resilience for Distributed Energy Resources: An Overview of Current Analytical Practices. *National Renewable Energy Laboratory*. <https://doi.org/10.2172/2394652>
- Rose, A., & Wei, D. (2013). Estimating the economic consequences of a port shutdown: The special role of resilience. *Economic Systems Research*, 25(2), 212–232.
- Sanstad, A. H., Zhu, Q., Leibowicz, B., Larsen, P. H., & Eto, J. H. (2020). Case studies of the economic impacts of power interruptions and damage to electricity system infrastructure from extreme events. Lawrence Berkeley National Laboratory. <https://doi.org/10.2172/1607982>
- Serrano-Fontova, A., Li, H., Liao, Z., Jamieson, M. R., Serrano, R., Parisio, A., & Panteli, M. (2023). A Comprehensive Review and Comparison of the Fragility Curves Used for Resilience Assessments in Power Systems. *IEEE Access*, 11. <https://doi.org/10.1109/access.2023.3320579>
- Shittu, E., & Santos, J. R. (2021). Electricity markets and power supply resilience: An incisive review. *Current Sustainable/Renewable Energy Reports*, 8, 42–50.
- Siano, P. (2014). Demand response and smart grids—A survey. *Renewable and Sustainable Energy Reviews*, 30, 461–478.
- Surinkaew, T., Shah, R., & Islam, S. (2022). Risk and resiliency assessments of renewable dominated edge of grid under high-impact low-probability events—A review. In *2022 IEEE Global Conference on Computing, Power and Communication Technologies (GlobConPT)* (pp. 320–326).
- Valencia, J. P. H., López-Lezama, J. M., & Restrepo-Cuevas, B. J. (2021). Assessing the vulnerability of power systems using multilevel programming: A literature review. *Revista Ingenierías Universidad de Medellín*, 20(38).
- Venkateswaran, B., & Panteli, M. (2023). Power system resilience during 2001–2022: A bibliometric and

- correlation analysis. *Renewable and Sustainable Energy Reviews*, 188. <https://doi.org/10.1016/j.rser.2023.113862>
- Voropai, N. I. (2020). Electric power system transformations: A review of main prospects and challenges. *Energies*, 13(21), 5639.
- Vugrin, E. D., Castillo, A. R., & Silva-Monroy, C. A. (2017). Resilience Metrics for the Electric Power System: A Performance-Based Approach (No. SAND2017-1493). Sandia National Lab.(SNL-NM), Albuquerque, NM (United States).
- Wang, Y., Chen, C., Wang, J., & Baldick, R. (2015). Research on resilience of power systems under natural disasters—A review. *IEEE Transactions on power systems*, 31(2), 1604-1613.
- Whitney, D. E., Luo, J., & Heller, D. A. (2014). The benefits and constraints of temporary sourcing diversification in supply chain disruption and recovery. *Journal of Purchasing and Supply Management*, 20(4), 238–250.
- Wilkinson, M. E., Addy, S., Quinn, P. F., & Stutter, M. (2019). Natural flood management: small-scale progress and larger-scale challenges. *Scottish Geographical Journal*, 135(1-2), 23-32.
- Xing, L. (2020). Cascading failures in Internet of Things: Review and perspectives on reliability and resilience. *IEEE Internet of Things Journal*, 8(1), 36–50.
- Xu, L., Guo, Q., Sheng, Y., Muyeen, S. M., & Sun, H. (2021). On the resilience of modern power systems: A comprehensive review from the cyber-physical perspective. *Renewable and Sustainable Energy Reviews*, 152. <https://doi.org/10.1016/j.rser.2021.111642>



Part 8

OpenBIM and Generative Design

STREAMING SENSOR REGISTRATION AND UPDATING PROCESS IN BIM-BASED DIGITAL TWINS

Cheng-Wen Fu¹, Tzong-Hann Wu², Yun-Tsui Chang³, and Shang-Hsien Hsieh⁴

1) Research Assistant, Department of Civil Engineering, National Taiwan University, Taipei, Taiwan. Email: awenfu0424@gmail.com

2) Postdoctoral Research Fellow, Department of Civil Engineering, National Taiwan University, Taipei, Taiwan. Email: tzonghannwu@ntu.edu.tw

3) Doctoral Researcher, Department of Civil Engineering, National Taiwan University, Taipei, Taiwan. Email: ytchang@caece.net

4) Prof., Department of Civil Engineering, National Taiwan University, Taipei, Taiwan. Email: shhsieh@ntu.edu.tw

Abstract: In response to the global climate crisis, better building energy management to achieve higher energy efficiency and lower carbon emissions is becoming increasingly important. However, data collection and integration of essential information can be challenging for building managers, resulting in ineffective collaboration between sensor providers and building managers and further data maintenance issues. This is also related to the high costs of building management platforms. To address this situation, this research proposes a streamlined collaboration and management platform realized through the visual platform of Autodesk Platform Services (APS) and Building Information Modeling (BIM). This platform aims to streamline the complex processes of constructing digital twins, including sensor installation and spatial coordinate management, energy data collection, and updates. At the same time, a streamlined user interface is designed to enable a cohesive and integrated workflow, facilitating data updates and maintenance for building managers in the future and significantly reducing the difficulty of developing and maintaining digital twins.

Keywords: Digital Twin, Building Information Modeling (BIM), Internet of Things (IoT), Data Integration

1. INTRODUCTION

The application of digital twin technology in the construction sector is advancing rapidly, particularly in enhancing the energy efficiency of existing buildings. Compared to traditional methods, digital twin technology can optimize energy use through real-time data monitoring and simulation, which is crucial for transforming existing buildings into "Net Zero Energy Buildings (NZEB)" (Kaewunruen et al., 2018). However, the deployment of digital twin technology faces several challenges.

First, the high initial capital investment is a major barrier, especially for small and medium-sized building owners, due to the high cost of integrating Internet of Things (IoT) devices, sensors, and data processing systems (Cespedes-Cubides & Jradi, 2024). Additionally, the complexity of digital twin systems requires skilled personnel and management teams, which increases the need for training and infrastructure upgrades, making the deployment process more complicated and lengthy. Another challenge is the inefficient management of historical energy data and digital models. Differences in data storage methods often lead to difficulties in updating and managing data, reducing the effectiveness of digital twin technology. Addressing these issues is, therefore, crucial.

Building on the framework proposed by (Lee et al., 2019), which integrates government open data with sensor data for smart construction site management, this research further developed an integrated workflow for registering, managing, and viewing building information and sensor data. This research also referred to the model management interface proposed by (Nagy, 2017) to establish a visual platform to support this workflow. This approach emphasizes data integration and enhances the connection between BIM and IoT, resulting in broader data management capabilities and improved scalability. These improvements simplify the implementation and maintenance of digital twins, ultimately enhancing building projects' management efficiency.

In addition, this workflow simplifies the implementation and maintenance of digital twins, allowing users to utilize the platform for data visualization and other technologies without requiring prior experience in BIM modeling or software operation. It enhances the management efficiency and applicability of construction projects.

2. METHOD

2.1 System Design for Integrated BIM and IoT Data Management

The system architecture follows a typical three-tier structure, consisting of the Application Layer, Business Layer, and Data Layer, moving from the user interface to the core data management functions. Each layer contains specific functional modules to support the integrated management of Building Information Modeling (BIM) and Internet of Things (IoT) sensor data. The application can be divided into three types of data access, corresponding to different users and their interactions with Building Information Modeling (BIM) models and Internet of Things (IoT) data, as shown in Figure 1. The first type is Building Model Management, where building managers can manage all data on the platform, including creating model projects, updating model versions, editing sensor information, and issuing authorization codes. The second type involves Sensor Registration, where

sensor registrants can upload sensor data to the model and link real-time or historical data to the corresponding data tables. The third type is the Visualization of Models and Sensor Data, allowing general users to view building models and real-time sensor data through the platform. This workflow supports the realization of digital twins (DT) for buildings and reduces the complexity of data maintenance.

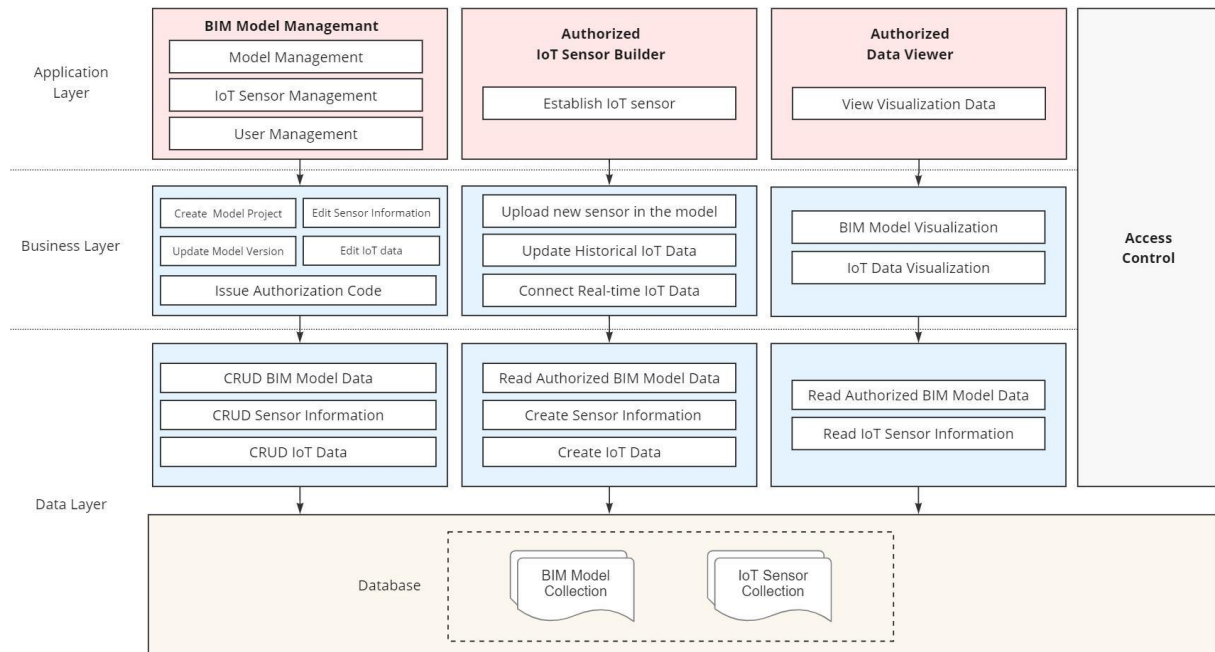


Figure 1. The three-tier diagram of the proposed platform

2.2 Database Design in the System

Regarding the storage method for BIM models and IoT data, instead of directly embedding sensor components within the BIM model and storing data there, the data is independently stored in a separate database, as shown in Figure 2. Different versions of BIM models are recorded in the project table, allowing efficient model version control, while sensor data is linked to individual projects. This method facilitates the independent management of models and sensor data, ensuring the accuracy of the most up-to-date model and sensor information. Additionally, this approach helps maintain system scalability and ease of maintenance.

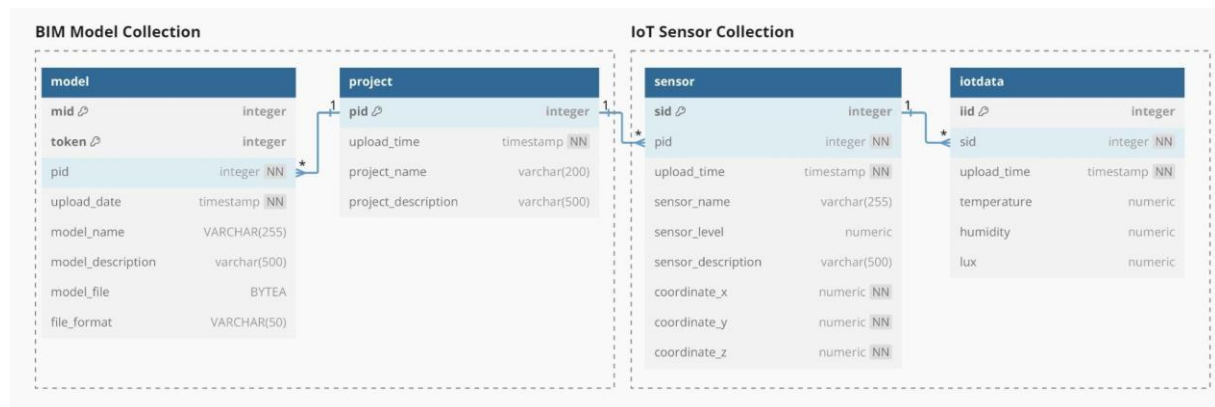


Figure 2. The database diagram for the proposed workflow

2.3 Collaboration Workflow in the Platform

The workflow for data management, sensor registration, and data visualization is illustrated in Figure 3. First, the data manager needs to log in and either create a new project or update the model version of an existing project. If a new project is selected, the system will create a new sensor database for the project. If an existing project requires an update, the latest BIM model will be uploaded and updated in the model database to allow sensor registrant later in the project. Next, the sensor registrant must enter an authorization code to access the

previously uploaded and authorized model by the data manager. The uploading of sensor data takes place in two steps: first, establishing the sensor's coordination and description within the model in a 3D environment and then uploading and linking the sensor data, thus allowing the BIM model and IoT data in the project to be interconnected. Finally, data viewers can use the platform to select models that are open for viewing by the data manager. This allows them to view the building model and its real-time sensor data, providing a better understanding of building operations and energy management. Overall, the platform offers an integrated workflow from model management to sensor setup and data visualization, reducing the complexity of digital twin implementation and data maintenance.

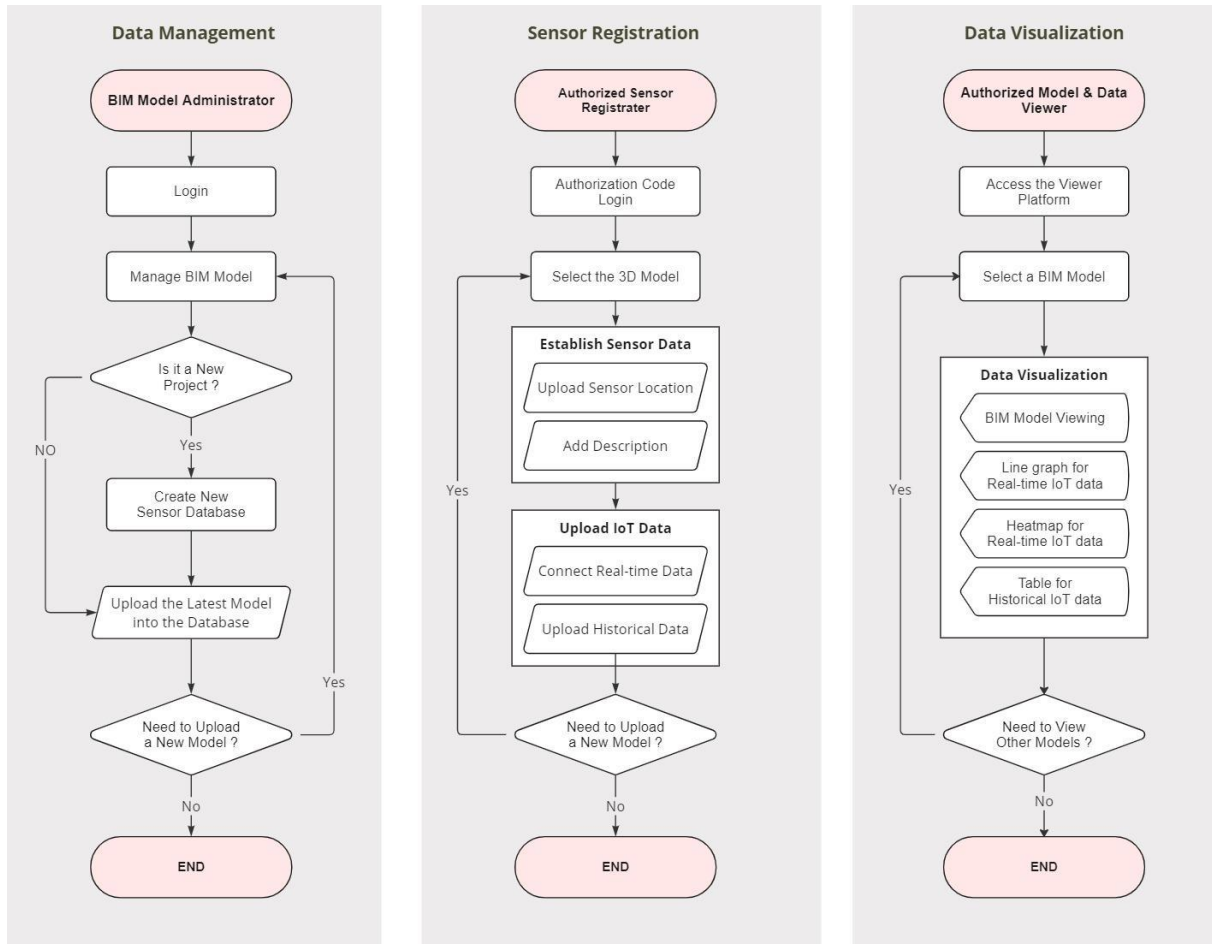


Figure 3. The flowchart in the proposed platform

3. RESULT & DISCUSSION: A CASE STUDY

This paper uses the Civil Engineering Research Building at National Taiwan University as a demonstration model, with sensors installed in Room 611 on the sixth floor to demonstrate the workflow described above. Figure 4 shows the platform's user interface, which is explained in three parts: (1) the building administrator uploading and managing the building model, (2) sensor data uploading by the sensor registrant, and (3) model data visualization for general users.

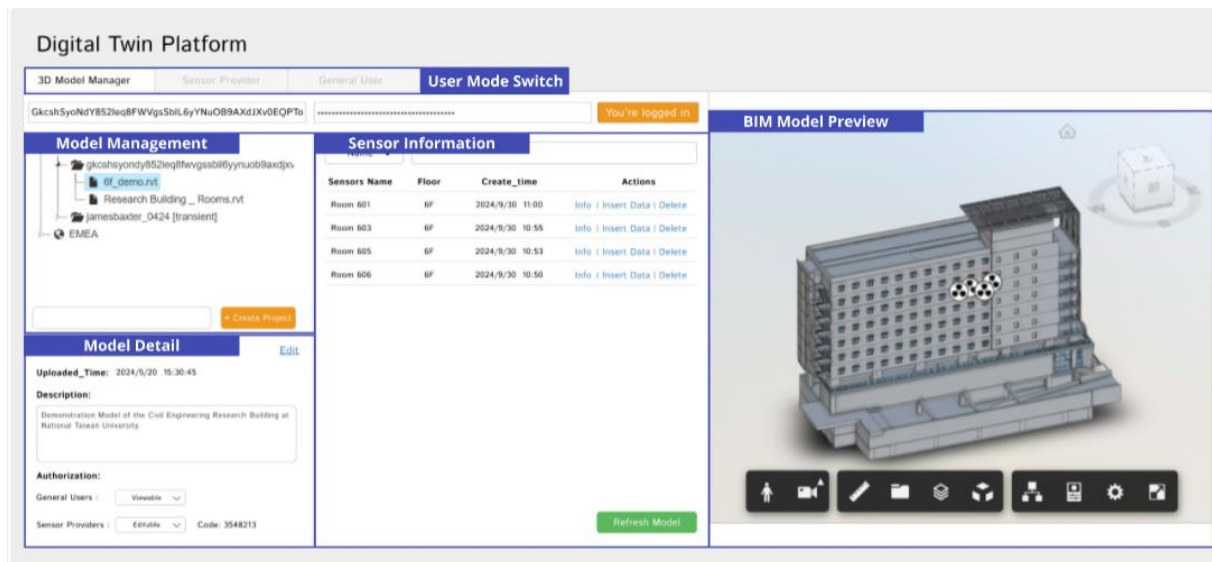


Figure 4. The user interface in the proposed platform

3.1 BIM Model Administrator

BIM Model managers can perform the following operations on the platform: (1) Model Upload and Management: Administrators can upload and manage multiple building models. (2) Sensor and IoT Data Management: Administrators can handle sensor data associated with the uploaded models.

(1) Model Upload and Management

First, the data manager must log in to the management system using their account. After selecting the regional server, they can browse all projects associated with their management account. As shown in Figure 5, to create a new project, the manager can use the 'Create new project' button and enter the project name to establish a new project. To upload a model, right-click on a specific project to upload a BIM model. Within each project, models are displayed in the order they were uploaded, facilitating efficient management of different versions. Finally, detailed information is displayed on the interface by clicking on a specific model, including model information, permission settings, BIM model preview, and a view of all IoT sensors within the model, as illustrated in Figure 6.

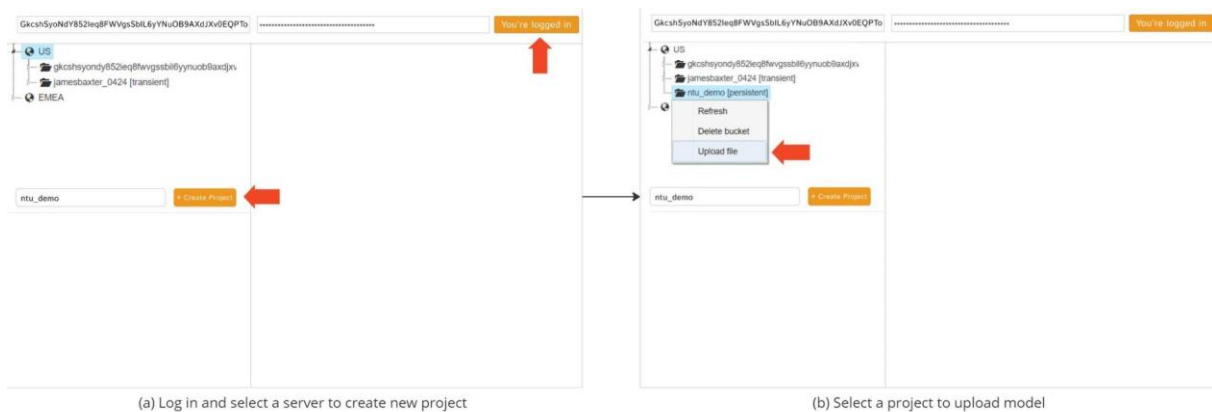


Figure 5. Key steps for creating projects and uploading models

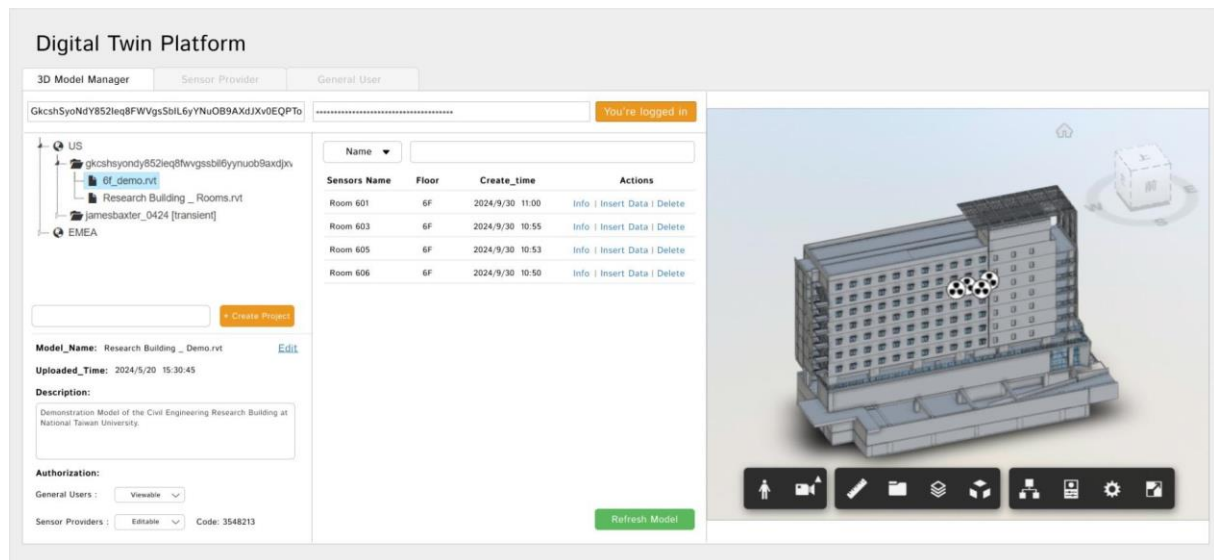


Figure 6. The user interface for the data manager

(2) Sensors and IoT Data Management

As shown in Figure 7, to manage and view IoT data, users can click on a specific sensor's "info" button in the platform. The sensor information will be synchronized and displayed in the BIM model viewer on the right side. Using the dropdown menu, users can edit the sensor's spatial coordinates within the model, modify historical sensor data, and check the status of real-time data reception, as illustrated in Figures 8 and 9.

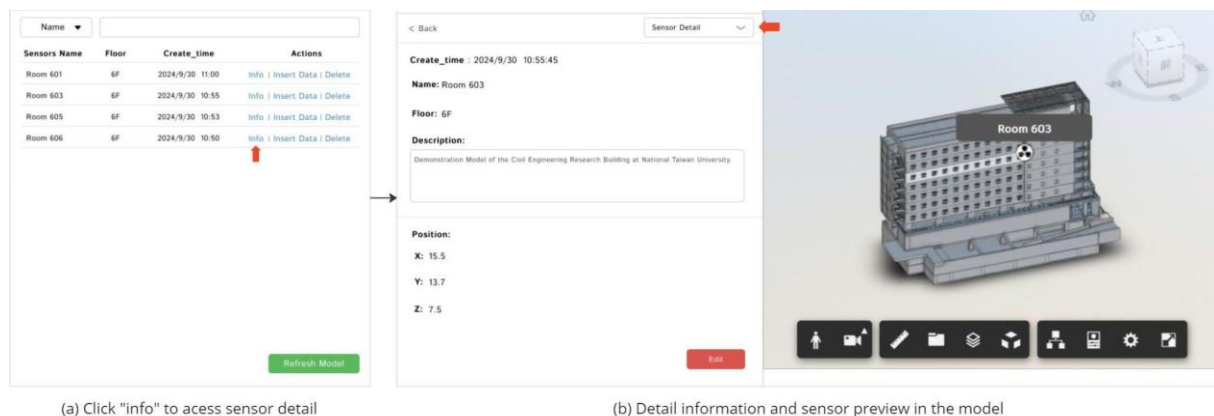


Figure 7. Key steps for viewing sensor detail information

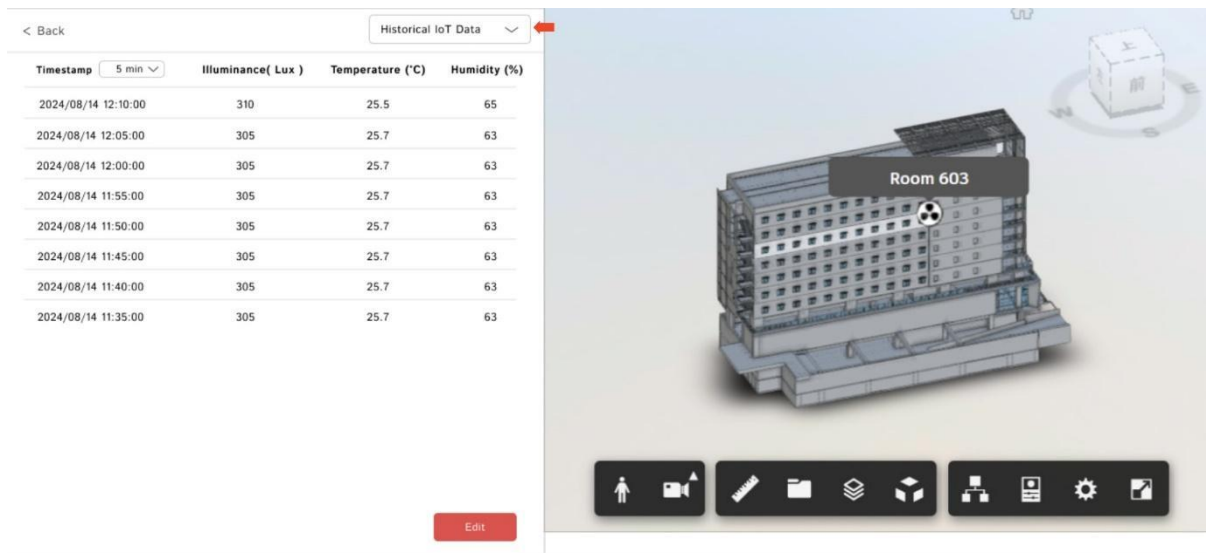


Figure 8. Historical IoT data viewing

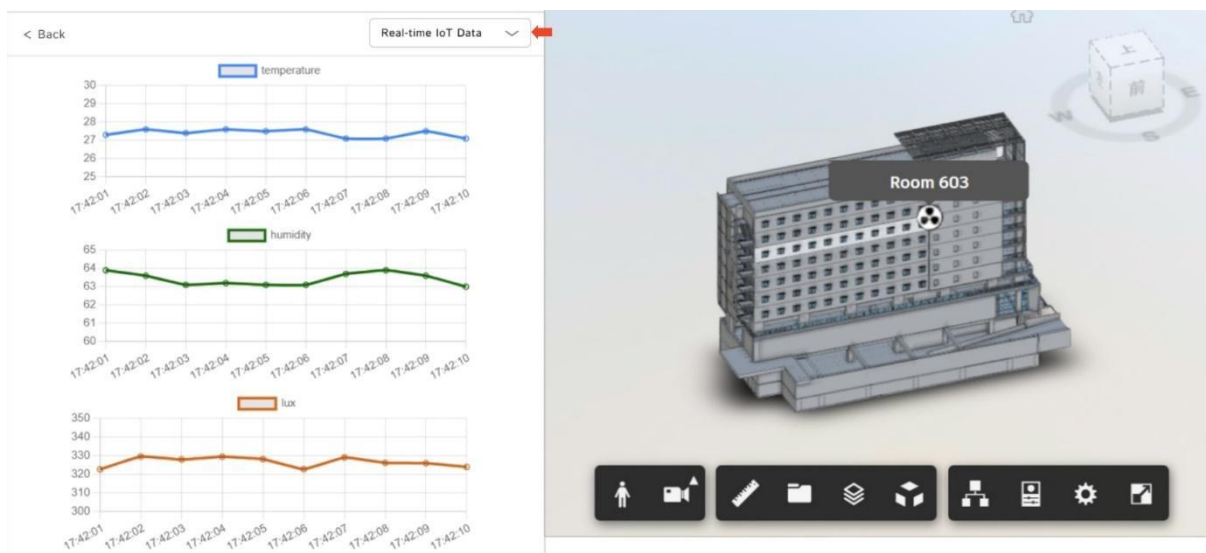


Figure 9. Real-time IoT data viewing

3.2 Sensor Registration

The sensor registrar can perform the following operations on the platform: (1) Place the new sensors in the model: set up and define the spatial coordinates for new sensors within the model. (2) Input historical and real-time sensor data: integrate historical sensor and real-time data into the database.

This registration method facilitates the quick and simple addition of sensor locations and data reception. Even without a basic understanding of BIM models or modeling software technologies, the general public can easily use this platform to set up sensor components and visualize information, promoting the adoption of Digital Twin technology. Moreover, since sensor and model data are stored separately, updates or adjustments to room names and interior spaces within a Revit model do not directly affect the sensor storage state. This approach enhances the convenience of separately maintaining model and sensor data during internal collaboration. It prevents issues such as version conflicts if both types of data are stored in the same Revit file and updated simultaneously by different users.

(1) Place New Sensors in The Model

The placement of sensors will be completed in several steps. First, as shown in Figure 10, users must enter an authorization code to access the model authorized by the data manager. At this stage, any previously established sensor information and the model can be viewed through the interface. Users can begin the sensor placing process after clicking the 'Create new Sensor' button. As shown in Figure 11, the placement of sensors involves several

steps. After switching the model floor, users can click the 'Get Coordinates' button and click on the model to obtain the coordinates within the model in a 3D environment.

Additionally, after clicking, a sphere will be created at the selected location in the model. Users can then use the transform controller to adjust the sphere, fine-tuning the desired coordinate position. Finally, the newly added sensor can be previewed in the model by pressing the 'Save' button.

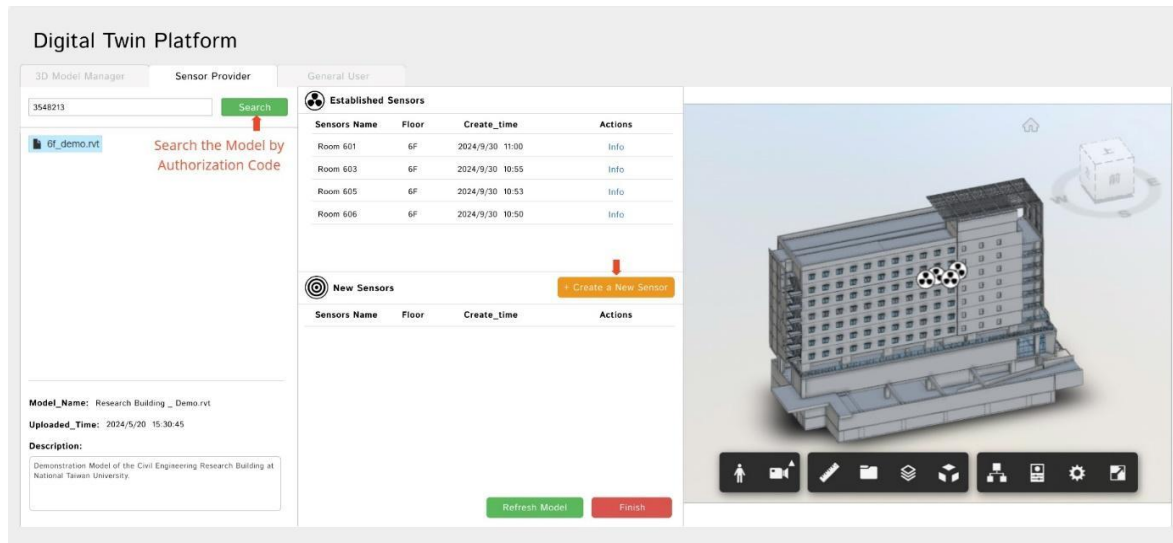


Figure 10. The user interface for sensor registration

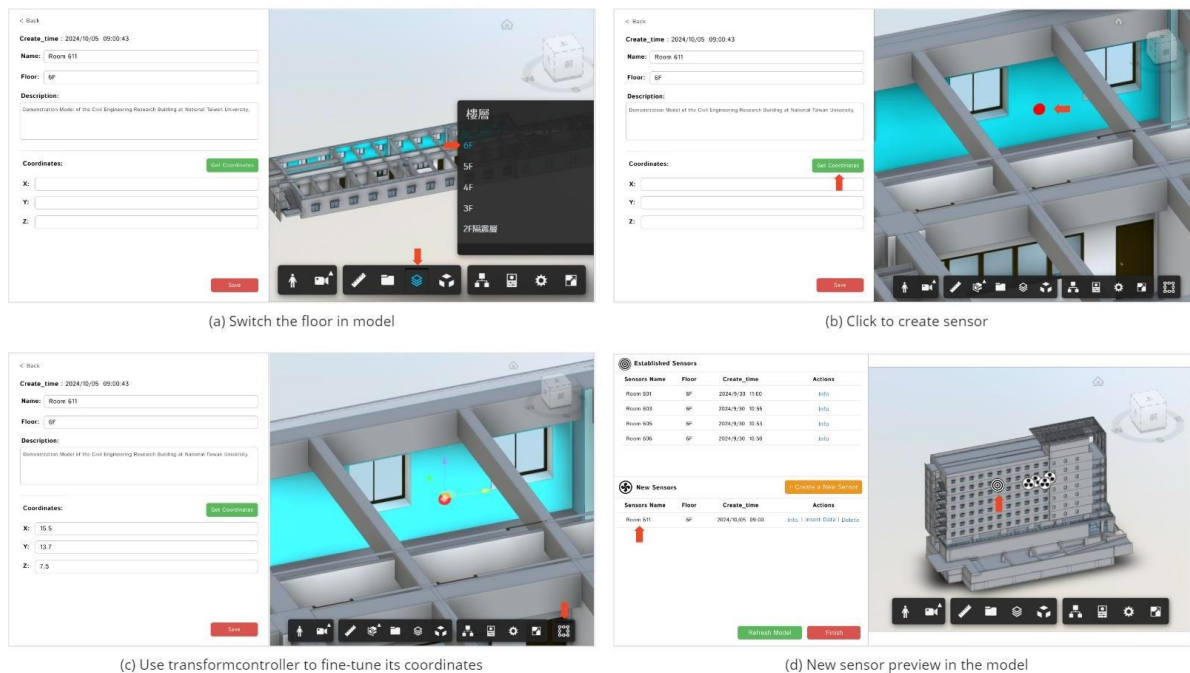


Figure 11. Key steps for registering a new sensor

(2) Upload Historical and Real-time Sensor Data

After placing a new sensor in the model, as shown in Figure 12, users can click to insert data into the sensor and switch to IoT data input via a dropdown menu. In this study, it is expected that all data collected by the sensors will first be transmitted to a database, from which the platform retrieves and displays the data. To facilitate real-time data collection from sensors, an API is provided for sensors to upload real-time data to the database, which the platform can then access and display. For sensors that have already collected data that needs to be integrated into this platform, an option to upload CSV files is provided to import the data into the database. Combining new and historical data through these two methods enables efficient data searching and viewing.

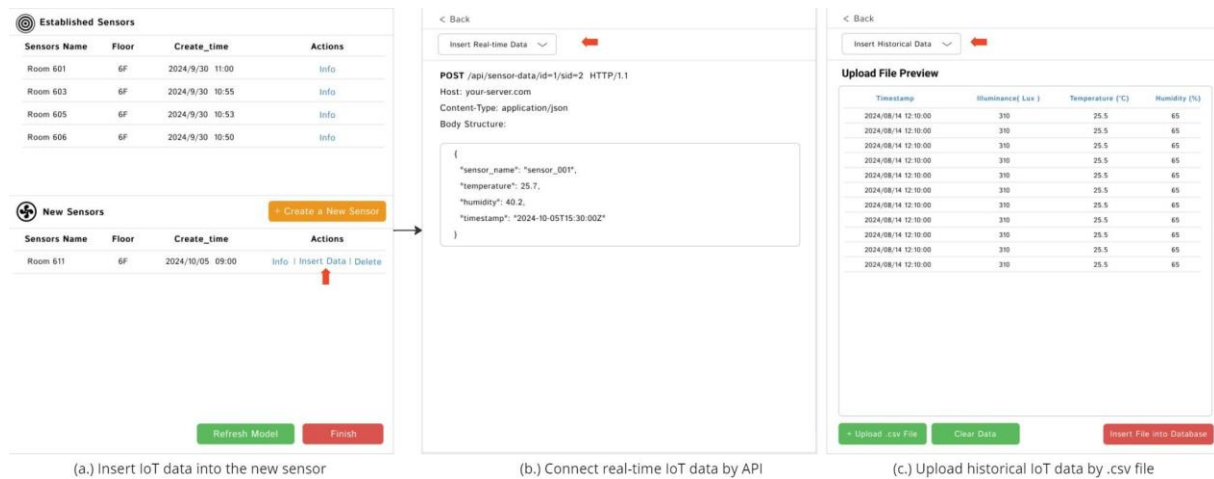


Figure 12. Key steps for uploading IoT sensor data

3.3 General User

As shown in Figure 13, a dropdown menu allows general users to view the models authorized by the manager. This includes accessing and analyzing historical energy data and viewing real-time sensor data from different spaces within the building through various visual representations, such as thermal maps, icons, and line charts, as depicted in Figures 14 and 15. This functionality helps users comprehensively understand the energy dynamics within the building.

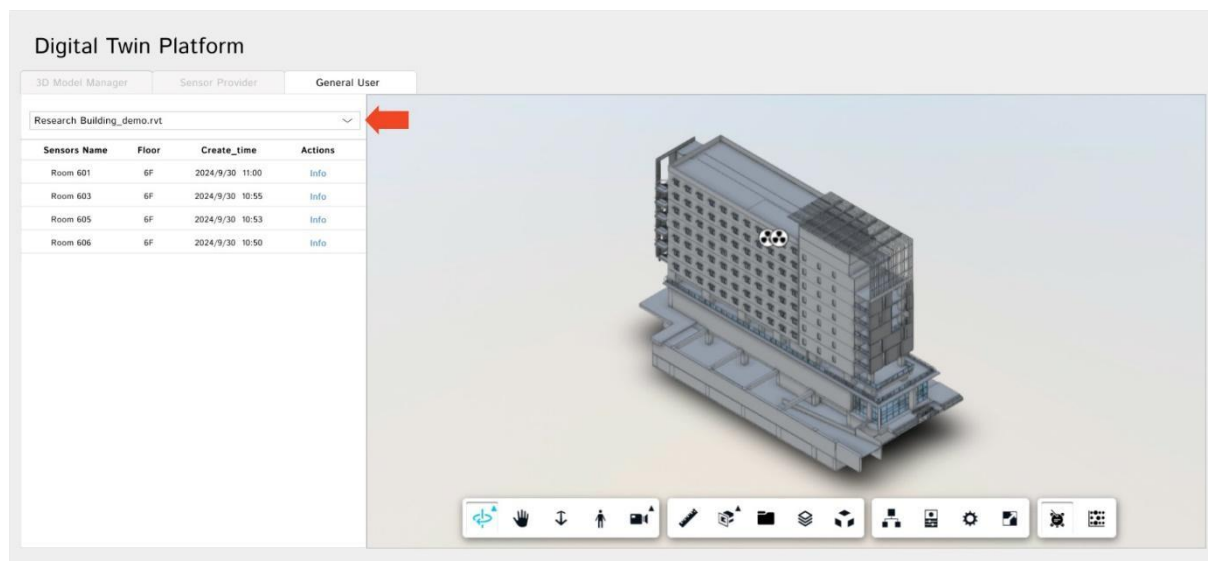


Figure 13. The user interface for general user

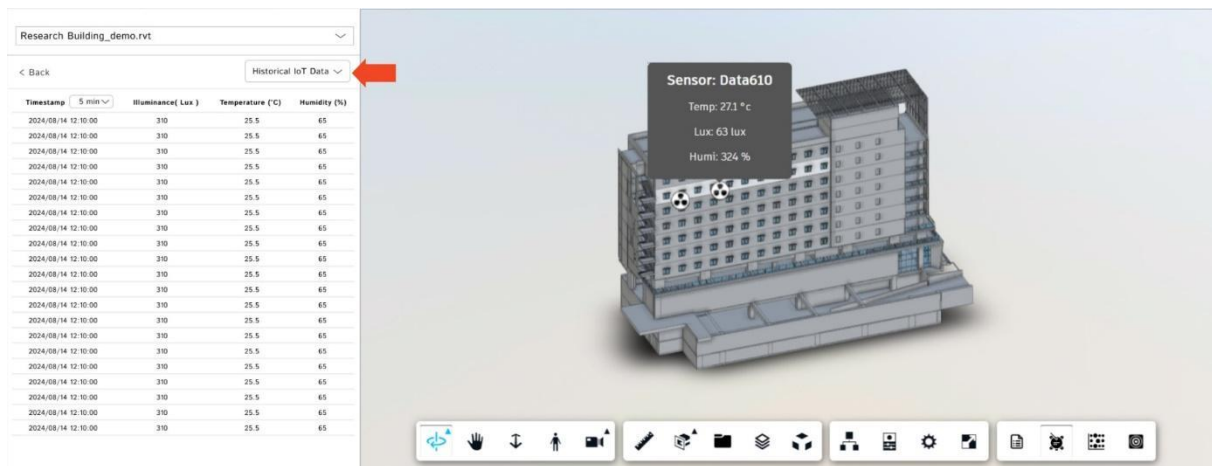


Figure 14. Historical IoT data viewing

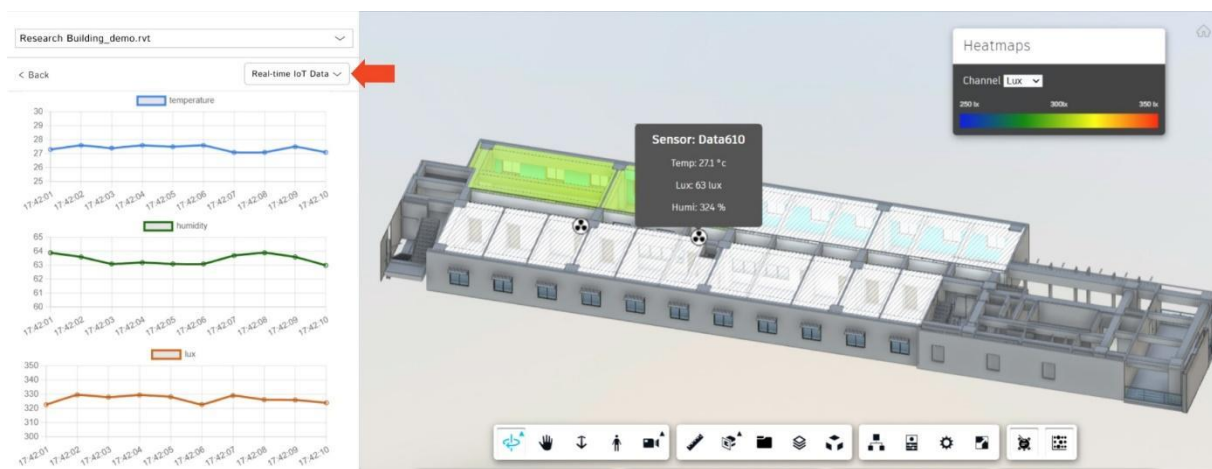


Figure 15. Heatmap and real-time IoT data viewing

3.4 Limitations

The workflow and methods for storing model data and sensor data mentioned in the article still need to be solved. One major issue lies in the accuracy of sensor coordinates. After installing sensors on-site, registering their coordinates on the platform is relatively simple, often leading to imprecise coordinate information. Additionally, their coordinates must be manually updated on the platform when repositioning sensors.

Another challenge arises when significant updates occur in the model space. The current data storage approach is designed to accommodate minor adjustments, such as the removal of partition walls, updates to equipment information, or reconfigurations of indoor spaces. However, sensor coordinates are recorded as absolute positions within the model space. Suppose the building model undergoes significant shifts (e.g., a large-scale repositioning of the entire structure within the model space). In that case, the sensor coordinates may become misaligned, requiring manual updates to restore visualization functionality.

Finally, the system displays the latest model version and the most recent sensor positions. However, it needs to effectively record and manage historical data, including earlier model versions and past sensor information. This limitation poses challenges for tracking changes over time and ensuring data consistency across different versions.

4. CONCLUSIONS & FUTURE WORK

This study streamlines the complex process of building a digital twin, achieving an integrated workflow through establishing a platform. Data storage is divided by purpose into model and sensor data storage, ensuring efficient and organized data management. The platform's integrated workflow meets the needs of three main types of users, each with specific workflows for effective data management and system utilization. Moreover, the framework facilitates efficient data management and maintenance by storing model and sensor data separately. In addition, through a user-friendly sensor registration method, the framework enables individuals without knowledge of BIM modeling or software operation to access digital twin technologies and visualize data. This

streamlined approach aims to foster a more cohesive and integrated process, ultimately reducing the complexity and cost of implementing and maintaining the digital twin.

In terms of future development, the study's primary objective is to establish a digital twin platform for the National Taiwan University campus. Through data visualization and environmental simulation, this platform aims to assist the university in addressing key issues, such as understanding the relationship between occupancy and energy consumption or the impact of environmental temperature and humidity on space energy efficiency. The platform will enable university management to make informed decisions. However, challenges arise during communication between campus administrators and academic departments, as some staff may need more technical background to operate complex systems. The simplified workflow proposed in this study addresses these challenges by establishing a clear system framework, categorized into three user types—administrators, data uploaders, and data viewers—to implement the platform and ultimately fulfill its goal of supporting campus management and decision-making through digital twin technology. Although the security of sensor data collection is not the primary focus of this research, it will be one of the key aspects to address in the future development of the proposed framework.

REFERENCES

- Adam E Nagy. (2017). *aps-hubs-tools*. Retrieved from GitHub website: <https://github.com/autodesk-platform-services/apshubs-tools>
- Cespedes-Cubides, A. S., & Jradi, M. (2024). A review of building digital twins to improve energy efficiency in the building operational stage. *Energy Informatics*, 7(1), 11
- Lee, C. M., Kuo, W. L., Tung, T. J., Huang, B. K., Hsu, S. H., & Hsieh, S. H. (2019, May). Government open data and sensing data integration framework for smart construction site management. In *Proceedings of the 36th International Symposium on Automation and Robotics in Construction, ISARC* (pp. 1261-1267).
- Kaewunruen, S., Rungskunroch, P., & Welsh, J. (2018). A digital-twin evaluation of net zero energy building for existing buildings. *Sustainability*, 11(1), 159.

LEVERAGING TRANSFORMERS FOR IMPROVED DECISION-MAKING IN ROAD MAINTENANCE

Rui Kang¹, Stephen Green², Ioannis Brilakis³

1) Ph.D. Candidate, Department of Engineering, University of Cambridge, United Kingdom. Email: rk703@cam.ac.uk

2) Ph.D., Department of Engineering, University of Cambridge, United Kingdom. Email: slg79@cam.ac.uk

3) Ph.D., Prof., Department of Engineering, University of Cambridge, United Kingdom. Email: ib340@cam.ac.uk

Abstract: Transportation performance is heavily influenced by the overall quality and the effectiveness of road maintenance. However, this remains an expert-dependent activity, despite recent efforts to digitalize road geometries and management processes. Road maintenance knowledge accumulated through addressing relevant enquiries is inexplicitly learned by experts and transferred into experience, which contributes very little to developing maintenance digitalization techniques and automated decision-making processes. In this case, fully utilizing historical maintenance records and turning them into computer-readable knowledge is a crucial task to be solved. This paper aims to extract key information from road maintenance request texts and then implement step-by-step thinking to make road maintenance decisions. This chain of thought is first proposed by reviewing the key elements and logical flow of road maintenance decision-making. Then, a cross-attention mechanism based on a transformer architecture is implemented on maintenance record texts and target knowledge element sequences. The result of this experiment overperforms on a pre-trained BERT model and demonstrates a valid performance on the text-knowledge alignment in road maintenance domain. The method proposed in this paper provides a solution for reliable and traceable decision-making and shows a promising application in domain-specific knowledge management.

Keywords: Transformer, Road maintenance, Decision-making, Natural Language Processing, Knowledge Management

1. INTRODUCTION

Roads are the main mode of transport for individuals and their maintenance has a great impact on the service quality of roads across the life cycle. For example, over 92% of the distance travelled by passengers in the UK is covered by roads, accounting for 595 billion passenger kilometers (Department for Transport, 2022). Over 60% of motorways in the UK were built during the 1960s and 1970s (Highways England, 2017). These aging road networks are facing problems caused by terrible road conditions. In England, about 39% of roads are marked with visible deterioration or in urgent need of maintenance (Asphalt Industry Alliance, 2023). Untimely maintenance will further increase costs and cause rapid deterioration of road assets. Hence, effective and reliable decision-making can have an important role in road maintenance management.

Road maintenance management processes can be divided into four types: planning, programming, preparation and operation (Robinson et al., 1998). Table 1 shows the spatial coverage, time horizons and staff involved in these processes. For the former two processes, long and medium-term systematic planning over a year-long is conducted by professionals taking policy and budgets into account. For the latter two processes, cyclical and reactive maintenance decisions are made by engineers and technicians to respond to regular inspections and reports. High-quality road maintenance decision-making is required to minimize traffic disruption, prevent safety issues, and improve cost effectiveness. However, decision-making in road maintenance remains heavily expert dependent. Despite the use of data management systems to record road maintenance activities, there are no effective measurements that use these records to support or automate decision-making processes. These decisions involve identifying road assets and defects, determining appropriate repair jobs, and prioritizing jobs based on asset importance, defect severity and job urgency. Such decisions still primarily rely on the accumulated experience of engineers.

Table 1. Road maintenance management process type

Management process	Spatial coverage	Time horizon	Staff concerned
Planning	Network-wide	Long term	Policy level professional
Programming	Network to sub-network	Medium term	Middle level professional
Preparation	Section or project	Budget year	Engineer/technician
Operation	Sub-section or activity	Actual duration	Technician/worker

Rule-based methods are commonly used to efficiently make use of human expertise in road maintenance decision-making. One of the most straightforward techniques is a decision tree. Since road defects are finite, engineers can manually construct decision trees to select appropriate treatments. There are also studies that use decision trees to select maintenance activities and support complicated decision-making processes (de Figueiredo et al., 2022; Hadjidemetriou et al., 2020). Since the capacity of addressing complex problems is limited by the

simplicity of decision trees, expert systems are proposed to resolve intricate decision-making problems. Tailored by experts, expert systems compile and formalize expert knowledge into rules and use an inference engine to handle knowledge (Tripathi, 2011). Compared to expert decisions, expert systems provide competent solutions for material selection (Salihudin et al., 2007) and task prioritization (Moazami et al., 2011) in the road maintenance domain. However, since rule-based methods are encoded from human knowledge and maintained by domain experts, they lack the capability to learn from historical data and flexibility to deal with any unforeseen changes.

Natural Language Processing (NLP) methods and Large Language Models (LLMs) are regarded as promising solutions to manage knowledge and support decision-making. (D’Orazio et al., 2023) investigated how machine learning methods are integrated in NLP and the automation the priority assignment of building maintenance tasks. This work verified the effectiveness of NLP in conducting a preliminary check of the most urgent requests. Similarly, NLP is used to rank the severity of maintenance requests through text mining and to support maintenance decision-making (D’Orazio et al., 2022). With the development of LLMs, pretrained models such as BERT (Devlin et al., 2019) and GPTs (OpenAI, 2024) outperform other NLP methods and are applied for multiple tasks, including decision-making. (Diemert and Weber, 2023) assessed ChatGPT responses on hazard analysis and suggested LLMs may be useful to assist human experts in this field. (Rasmy et al., 2021) used a pre-trained BERT and fine-tuned it for medical disease prediction. Such applications demonstrate the potential of LLMs for various decision-making tasks across different fields. However, applying them in specific domains requires delicate fine-tuning, as they were pre-trained on large common-sense corpora and consists of millions of parameters and their performance can be hindered when domain-specific data is insufficient. Additionally, responses from LLMs can be unreliable. Since their answers are generated by black-box processes, changes in prompts can lead to different outputs. In practice, a human expert is required to supervise the use of LLMs to ensure the correctness.

This paper proposes a transformer-based method which targets road maintenance decision-making processes and provides traceable and reliable decisions with step-by-step thinking. The decisions consist of four parts: the judgement of asset types, defect types, job types, and priority. The key difference between our approach and popular pre-trained LLMs such as BERT and GPTs is the attention mechanism (a method that enables the model to focus on the most important part of the input data). While BERT and GPTs primarily use self-attention mechanisms to focus on single source inputs only, our approach relies on a cross-attention mechanism, which enables connections between two different sources, to align road maintenance enquiry text (the description of road conditions requiring a maintenance activity) and maintenance knowledge. The contributions of this approach are summarized as follows:

- (1) Road maintenance knowledge is structured in logical chains, which imitate expert decision-making think path and provide visible and traceable steps for decision-making processes. The use of logical chains enables stepwise thinking within the model through the cross-attention mechanism.
- (2) The model is trained with real-world road maintenance records and achieves automatic maintenance priority rankings, together with asset, defect, and job information alignment.
- (3) This model, while using fewer parameters, outperforms pre-trained and fine-tuned BERT models, demonstrating the capacity of decision-making in road maintenance domain.

2. METHOD

Figure 1. provides a high-level illustration of the architecture (named RTransformer standing for the Road-Transformer), consisting of three stages: (1) the processing of inputs, including texts and logical chains, before the attention mechanism, which involves formatting the logical chains and mapping inputs into a continuous vector space as embeddings. Both the enquiry descriptions and the logical chains are extracted from real-world road maintenance records. The logical chain is then arranged in a specific order, with a special placeholder <SOS> (Start of Sequence), inserted at the beginning. The discussion of the logical chain is covered in subsection 2.1; (2) the attention mechanism, including self-attention, where input information is integrated and prepared for the next layer, and cross-attention, where the logical chains can focus on the text information. A mask is added in logical chain processing to prevent data leakage. The V stands for Value, K stands for Key and Q stands for Query. The attention mechanism is discussed in detail in subsection 2.2; (3) the decoding and prediction of the next position in the chain, which is achieved using a linear decoder and a softmax layer to obtain the probability distribution of positions.

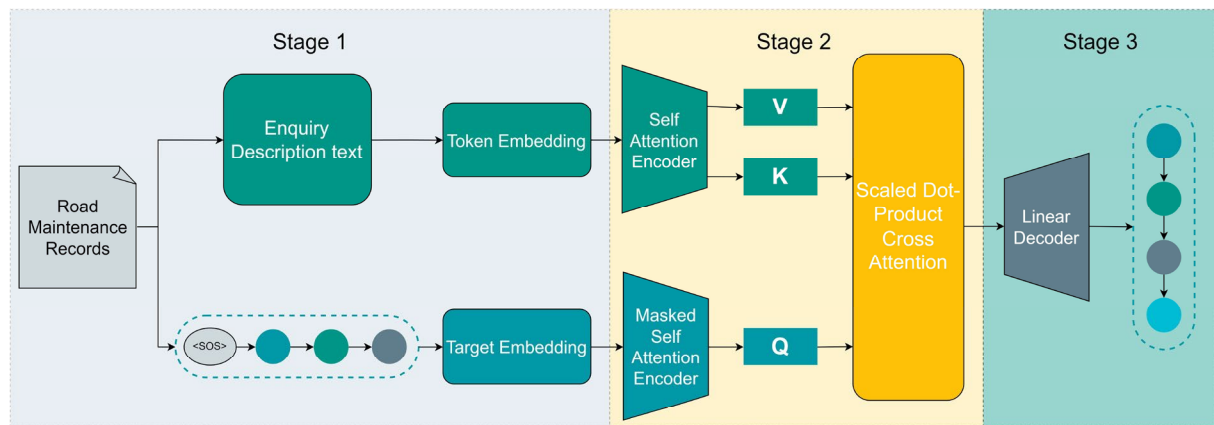


Figure 1. The architecture of RTransformer model.

2.1 Road maintenance decision-making logical chains

Road maintenance decision-making by human experts involves the intermediate judgement of maintenance-related factors and the final decision made on top of these factors. With their experience, experts can classify key factors before making maintenance decisions. Experts can intuitively identify the asset for which the maintenance enquiry is reported on, the defect which has occurred on the asset, and the job required to fix the defect. Given all this information, final decisions are based on the priority of maintenance tasks, which reflects their level of urgency and significance. Some examples of road maintenance decision-making key factors are listed in Table 1.

Table 1. Examples of key factors of road maintenance

Asset Type	Defect Type	Job Type	Priority
Single 2-Lane Carriageway	Pavement Pothole	Incident Response	2 hours
Gully	Drainage Defect	Incident Repair	24 hours
Safety Barrier (Steel)	Miscellaneous Accident damage	Fences, Walls etc. Repair	2 months
Lightning Column	Lighting Column Defect	Lightning Repair	7 days
Post (Signs)	Road Markings Wear	Traffic Sign/Road Marks Repair	28 days
Catch pit	Drainage Flooding	Drainage, Service Ducts Repair	Fix now
Dual 1-Lane	Pavement Debris in traffic lane	Sweeping, Cleaning	14 days
.....

NLP methods are available to extract key factors from natural language. It is common practice to treat NLP-based decision-making problems as classification tasks, directly predicting final decisions such as activity priority. Such methods typically involve locating key words for decision-making, such as Named Entity Recognition (NER). However, road maintenance enquiries can be vague and written in layman's language, as it describes road conditions based not only on inspections but also on feedback from customers. The following example demonstrates part of an enquiry raised by a customer:

"Same problem Very loud clattering noise every time traffic goes over it. Sound like something metal which is loose The noise every time a vehicle goes over it, is immense. I wouldn't be surprised also now that it could possible cause a serious accident....."

This enquiry describes a rattling manhole issue; however, no terms related to drainage are mentioned, and misleading words like "serious" and "accident" are used. In such cases, NLP methods that make direct decisions may overlook the complexity inherent in assessing multiple factors, leading to unexplainable and inaccurate outcomes.

To address this problem, we proposed a logical chain that manages knowledge in a structured format and mirrors the intuitive decision-making process of experts. An illustration is shown in Figure 2. The logical chain for road maintenance is designed in the order of Asset Type, Defect Type, Job Type and Priority, aligning with real-world practices. From the enquiry, the type of asset is first confirmed, and then the type of defect is reasoned based on both enquiry content and asset information. By repeating this process, the logical chain enables each

subsequent prediction to benefit from the additional information provided by all previous predictions until the final decision of priority is made. Taking the previous customer raised enquiry as an example, “Manhole” is first matched as the asset type, then “Drainage Defect” is reasoned as the defect type with additional asset information, followed by the job type as “Drainage, Service Ducts” follows. Finally, the priority is decided as “7 days”, considering all previously extracted information.

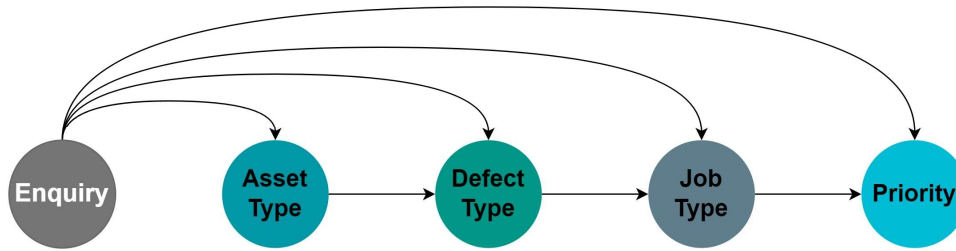


Figure 2. The logical chain of road maintenance

2.2 Attention mechanism

The attention mechanism used in this work is Scaled Dot-Product Attention (Vaswani et al., 2017) as used in the transformer model, which can be expressed as Equation (1). Q (for query), K (for key) and V (for value) are vectors that represent inputs in continuous space. The dimension of vectors is denoted as d , and \sqrt{d} is used as a scaling factor.

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (1)$$

The self-attention mechanism is wrapped in encoders to separately process inputs from the enquiry text source and the knowledge source. In the self-attention mechanism, Q, K and V are derived from the same source and contribute to the vector v that assembles all token features. In the self-attention encoder on the text side, the output v_{text} is subsequently used as K and V in the following cross-attention mechanism. On the knowledge side, additional masking is added to prevent data leakage. As the positions of the logical chain are the targets to predict, the current prediction depends on all known previous positions. The output $v_{knowledge}$ provides Q for the cross-attention mechanism. Hence, for cross attention mechanism, the original attention equation can be rewritten as Equation (2):

$$Attention(v_{knowledge}, v_{text}) = softmax\left(\frac{v_{knowledge}^i v_{text}^T}{\sqrt{d}}\right)v_{text} \quad (2)$$

Equation (2) is used in cross-attention for text-knowledge alignment, where i is the index of the position to predict and $i \in \{1, 2, 3, 4\}$. The data flow from self-attention to cross-attention is illustrated in Figure 3 using a real-world example. The enquiry description text “Road Traffic Collision CAR HIT LAMP COL” is first tokenized into separate tokens, and the token embedding layer generates embeddings for each token, represented as $[x_1, x_2, \dots, x_d]$. Through self-attention mechanism, v_{text} is generated based on all token embeddings. On the other hand, <SOS> is concatenated at the beginning of the logical chain as a placeholder for cross-attention. Through the target embedding layer, positions and <SOS> are embedded as $[y_1, y_2, \dots, y_d]$. Self-attention mechanism also processes the embeddings of positions less than i (because of masking) and outputs $v_{knowledge}^i$. Cross-attention mechanism, v_{text} attends to $v_{knowledge}^i$ of each position for a step-by-step prediction. For example, to predict Job Type position, masked self-attention accumulates data from <SOS> embedding, “Lighting Column” embedding at Asset Type position, and “Miscellaneous Accident damage” embedding at Defect Type position, and outputs $v_{knowledge}^3$ as the result of self-attention. Data from “Incident Response” and “30 Minutes” embeddings at position 3 and 4 are not considered. Together with v_{text} from self-attention for texts, the attention is computed by Equation (2) and then decoded by the linear decoder to predict the exact content at Job Type position.

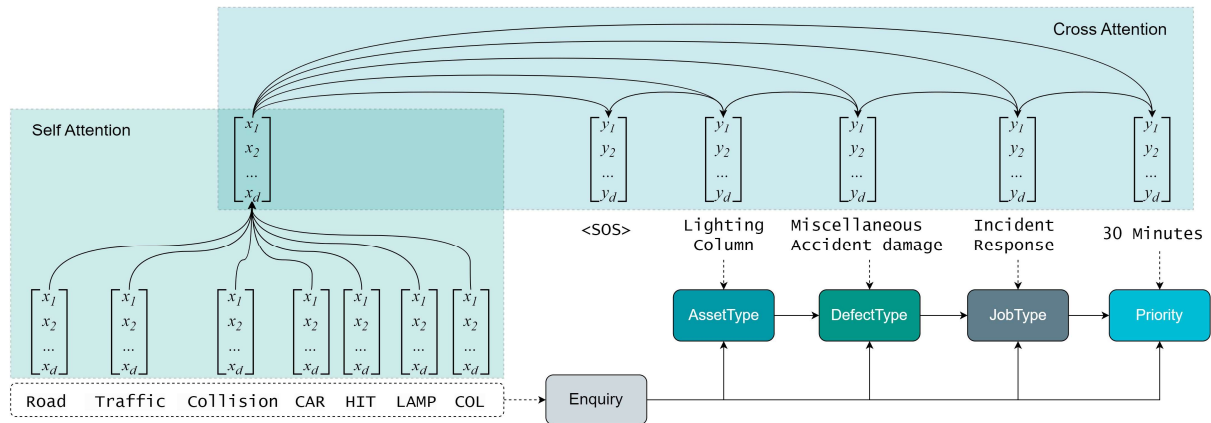


Figure 3. Data flow in the combined attention mechanism

2.3 Training and reasoning processes

In this section, we introduce the training and reasoning processes employed by our RTransformer. Both processes follow the path of stepwise prediction, but the strategy of training and reasoning is different. During training, a teacher forcing strategy is used to guide the model learning process with ground truth data. In the reasoning process, ground truth is not available, and the subsequent predictions depend on the previous predictions. These different strategies allow the model to be efficiently trained and to make predictions independently.

(1) Training process

The teacher forcing strategy is shown in Figure 4. The attention mechanism is employed among ground truth rather than predictions. During the training process, the Enquiry text, <SOS> placeholder and ground truth are the inputs to RTransformer model. To implement teacher forcing, each position of the logical chain is shifted by one. The last position is dropped, and the beginning of the chain is held by <SOS>. The first prediction is made from Enquiry and <SOS>. Instead of using the predicted position, the next prediction depends on the ground truth of the previous position, e.g. Defect Type prediction is made from Enquiry, <SOS> and ground truth Asset Type. In this way, the model always learns from the correct knowledge.

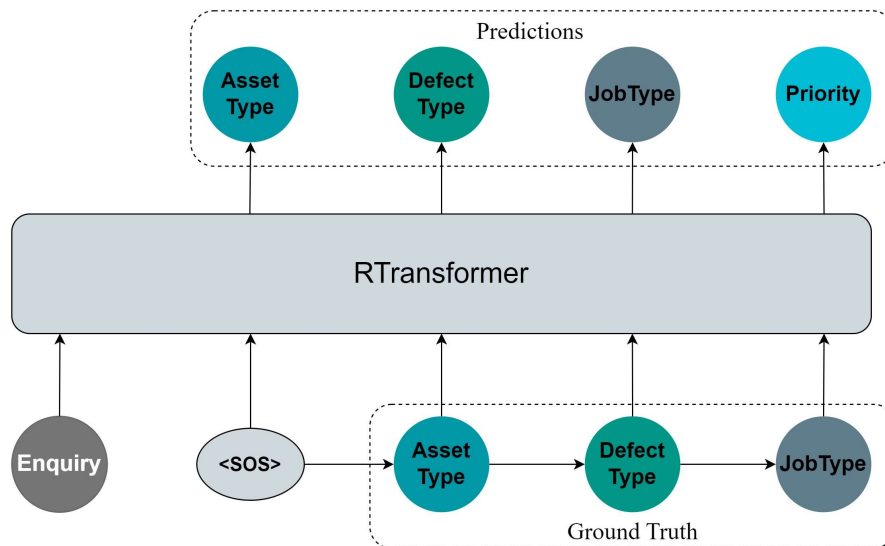


Figure 4. Teacher forcing in RTransformer training process

(2) Reasoning process

The reasoning process can be described in pseudocode as shown in Figure 5. The major difference is that the inputs to the RTransformer on the knowledge side are predictions rather than ground truth. The RTransformer model and <SOS> are pre-trained and their parameters are frozen during reasoning. The data flow on the Enquiry text side remains the same as in the training process. On the knowledge side, the input is initialized from the placeholder <SOS> and dynamically updated with the predictions in the reasoning loop. The reasoning loop stops

after four predictions, concluding the decision-making process. The predicted logical chain is the final output of RTransformer.

Algorithm 1 Reasoning Process

```

1:  Input: Enquiry text
2:  Retrieve the pre-trained model RTransformer, and the pre-trained Placeholder <SOS>
3:  Tokenize Enquiry, applying truncation or padding to achieve the predefined length
4:  Initialize Target using Placeholder <SOS>
5:  Initialize Position array
6:  for i from 0 to 3 do
7:      Calculate Logits using RTransformer(Enquiry, Target)
8:      Predict Positioni class from softmax(Logits)
9:      Concatenate the predicted Positioni to the end of Target to form the new Target
10: end for
11: Return Position
    
```

Figure 5. Pseudocode of Decision Reasoning Process

3. EXPERIMENTS

This section describes the RTransformer training process in detail, together with comparison to BERT variants as baseline models. This model is trained on a virtual machine with 1 NVIDIA Tesla V100 GPU.

3.1 Training data

The RTransformer and the baseline models are both trained on real-world road maintenance records from National Highways England (Highways, 2023). The original data was cleaned and balanced across classes for a better quality. The total amount of records used for training is 19,297, and each record contains enquiry description texts, the Asset Type, the Defect Type, the Job Type and the Priority. A logical chain is generated for each record. For enquiry texts, the sequence length is limited to 100 tokens, as over 99% sequences are shorter than this threshold. A special token <PAD> is used for padding. The shared vocabulary in the token embedding layer consists of 9,231 tokens. For the target positions to predict, there are 125 classes in total across all four positions.

3.2 Baseline model

We used BERT as our baseline model to compare the performance on the same training dataset. As BERT uses the encoder only architecture and focuses on self-attention, we simplified the decision-making problem as a last step classification task. In this way, the original step-by-step decision-making of 125 classes is turned into one final priority classification of 30 classes. With fewer classes, it is easier for pre-trained BERT to differentiate between classes. In our experiments, BertForSequenceClassification model from Hugging Face library (Wolf et al., 2020) is used, and the pre-trained weights are from different versions of BERT-based models with different sizes.

3.3 Optimizer and hyperparameters

We used the AdamW optimizer (Loshchilov and Hutter, 2017) instead of the Adam optimizer (Kingma and Ba, 2017) that is used in the original transformer. The change was made to address the overfitting issue caused by insufficient training data. The hyperparameters are set as follows: $\beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 10^{-8}, \text{weight_decay} = 10^{-4}, \text{learning_rate} = 10^{-4}$. A dropout rate of 0.1 is also applied in the model.

In this work, we also employ Multi-Head Attention to jointly attend to information in different aspects at different positions. A test of different head numbers and layers is also conducted to find an optimal model structure.

4. RESULTS

This section presents the results of experiments of testing RTransformer in different variations and the performance of baseline models.

4.1 Model variations

The experiments are conducted based on an early stop strategy. The whole dataset is divided into two subsets, with 70% of the data allocated for training and 30% reserved for validation. The patience parameter in the early stopping strategy is set to 5, allowing a wait of five epochs for any improvement in validation loss.

Variations of RTransformer have been tested in different aspects, including the change of layer numbers, batch sizes, number of heads and the dimension of models. We denote the attention layer number in encoders and cross-attention block as layer, the dimension of $v_{knowledge}$ and v_{text} before Multi-Head Attention application as d_{model} , the dimension of the feed forward layer in the linear decoder as d_{ff} . In Table 2, the params column records

the amount of all trainable parameters in RTransformer. The epoch corresponds to the last instance before the validation loss begins to increase, indicating potential overfitting. And the accuracies of the four positions are also listed in the table. All accuracy metrics are obtained in the validation dataset at the epoch corresponding to the minimum validation loss.

The base model architecture is configured to align with the original transformer base model settings. Due to the significantly smaller size of the training dataset compared to the original transformer, it is essential to adjust the number of layers and heads to identify the optimal configuration. A balance is also required between accuracy performance and the number of parameters, which directly impacts computational cost.

Table 2. Variations on the RTransformer architecture and experiment results

	layer	batch	d _{model}	d _{ff}	head	params ×10 ⁶	epoch	AssetType accuracy	DefectType accuracy	JobType accuracy	Priority accuracy
base	6	256	512	2048	8	51.4	10	0.453	0.674	0.803	0.610
(A)	3					29.4	10	0.465	0.697	0.810	0.626
	2					22.0	11	0.483	0.701	0.809	0.632
	1					14.6	11	0.473	0.673	0.790	0.610
(B)	2				1	22.0	12	0.459	0.672	0.785	0.610
	2				4	22.0	12	0.482	0.697	0.806	0.631
	2				16	22.0	12	0.511	0.727	0.827	0.650
	2				32	22.0	10	0.484	0.701	0.810	0.626
(C)	2		256			9.6	20	0.499	0.718	0.817	0.643
	2		768		16	37.5	8	0.476	0.699	0.805	0.627
(D)	2	64			16	22.0	7	0.478	0.683	0.805	0.616
	2	128			16	22.0	8	0.482	0.705	0.807	0.625

The result summary of model variations is listed in Table 2. Group (A) includes the test on how different numbers of layers influence the result. The best result is gained when two layers are used, with 22.0 million parameters. The model with more layers does not achieve better performance, which may result from the misalignment between the training data volume and the parameter volume. When the number of layers is reduced to one, the model does not fully utilize its potential to learn from the data. Group (B) explores the influence of heads. As heads allow the model to attend to different aspects of the data, the model with 16 heads achieves the best result, while using 32 heads leads to earlier overfitting. The test in group (C) examines the variations in the model dimensions. A dimension of 256 leads to the disqualification of the model, while an increase in dimensions results in early overfitting. The test in group (D) focuses on the batch size, which influences the stability of the training. A bigger batch size improves the performance with fewer learning steps in each epoch. However, limited by the memory of the GPU, the biggest available batch size is 256.

The results presented in Table 2 demonstrate a trend of overfitting. Although the vanilla Transformer achieves the best performance with 213 million parameters (Vaswani et al., 2017) in their experiments, the bigger the better does not apply to domain-specific applications. One of the primary reasons is the deficiency of training data both in terms of quantity and diversity, which easily triggers the early stop strategy to prevent overfitting.

4.2 Comparison to baseline models

BERT-based pretrained models are used as baseline models to compare performance on priority prediction. As these models are pretrained on lots of various books and Wikipedia, fine-tuning is required to adapt them for the road maintenance domain. This fine-tuning dataset is the same dataset used in RTransformer training.

BERT variants of different sizes are tested on the same dataset. The results are listed in Table 3. BERT_{BASE} is the original BERT model from (Devlin et al., 2019), the DistilBERT model leverages knowledge distillation and reduces the model size to 65.8 million (Sanh et al., 2019). To compare performance across various model sizes, BERT_{medium} and BERT_{small} (Turc et al., 2019) are also tested.

Due to the limitation of GPU memory, the training batch size is unified as 128.

Table 3. Comparison among BERT variants and RTransformer

	layer	batch	d _{model}	d _{ff}	head	params ×10 ⁶	epoch	Priority accuracy
ours	2	128	512	2048	16	22.0	8	0.625
BERT _{BASE}	12	128	768	3072	12	110	4	0.580
DistilBERT	6	128	768	3072	12	65.8	5	0.476
BERT _{medium}	8	128	512	2048	8	41.7	4	0.582
BERT _{small}	4	128	512	2048	8	29.1	5	0.583

From Table 3, we can see that our model achieves the best performance with the smallest model size. Additionally, the accuracy of our model is gained on the base of 125 classes of all positions, while other BERT variants are tested with 30 classes from the priority position only. Apart from DistilBERT, which utilizes a different mechanism, the size reduction of BERT models has little impact on accuracy results. This may be attributed to the quality and quantity of fine-tuning data.

5. CONCLUSIONS

This paper leverages a transformer to conduct decision-making in road maintenance. The cross-attention mechanism is applied to align text and domain-specific knowledge. By integrating the cross-attention mechanism with the logical chain design, this approach achieves a stepwise decision-making process. The results of experiments on real-world data demonstrate that RTransformer outperforms BERT models that utilize self-attention alone.

In domain specific tasks such as road maintenance decision-making, the limited availability of data restricts the performance of models, often leading to overfitting during training and fine-tuning. In this context, a restrained model architecture is preferable. Although pretrained LLMs are considered suitable solutions for knowledge management, they may not be optimal choices in domain-specific tasks, especially when considering data availability and computational costs.

The method proposed in this paper demonstrates the potential of utilizing the Transformer architecture for decision-making tasks in specific domains. It provides a feasible solution to guide the learning process of the model by designing logical chains to allow stepwise thinking.

Despite the promising results achieved in this study, several research limitations remain. Although the stepwise thinking introduced by the logical chain improves the performance of our model, it relies on semantic relations among elements and overlooks the structural features of knowledge. Additionally, the format of logical chains lacks compatibility. Future work will aim to organize knowledge in a more structured manner and incorporate structural features into the decision-making process.

ACKNOWLEDGMENTS

This work was supported by the UK Engineering and Physical Sciences Research Council (EPSRC) [grant number EP/V056441/1] and Trimble, the first author is sponsored by the China Scholarship Council and Cambridge Trust for PhD study at the University of Cambridge. The authors of this paper would like to thank Lavindra de Silva, Alix Marie d'Avigneau and Haritha Jayasinghe for their comments and inspiration.

REFERENCES

- Asphalt Industry Alliance, 2023. Annual Local Authority Road Maintenance Survey Report 2023. UK.
- de Figueiredo, B.H., dos Santos, M., Fávero, L.P.L., Moreira, M.Â.L., Costa, I.P. de A., 2022. Analysis of maintenance activities in Urban Pavement Management Systems based on Decision Tree Algorithm. *Procedia Computer Science*, 9th International Conference on Information Technology and Quantitative Management 214, 712–719. <https://doi.org/10.1016/j.procs.2022.11.233>
- Department for Transport, 2022. Transport Statistics Great Britain (No. TSGB01). UK.
- Devlin, J., Chang, M.-W., Lee, K., Toutanova, K., 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. <https://doi.org/10.48550/arXiv.1810.04805>
- Diemert, S., Weber, J.H., 2023. Can Large Language Models Assist in Hazard Analysis?, in: Guiochet, J., Tonetta, S., Schoitsch, E., Roy, M., Bitsch, F. (Eds.), *Computer Safety, Reliability, and Security. SAFECOMP 2023 Workshops*. Springer Nature Switzerland, Cham, pp. 410–422. https://doi.org/10.1007/978-3-031-40953-0_35
- D'Orazio, M., Bernardini, G., Di Giuseppe, E., 2023. Automated Priority Assignment of Building Maintenance Tasks Using Natural Language Processing and Machine Learning. *Journal of Architectural Engineering* 29, 04023027. <https://doi.org/10.1061/JAEIED.AEENG-1516>
- D'Orazio, M., Di Giuseppe, E., Bernardini, G., 2022. Automatic detection of maintenance requests: Comparison of Human Manual Annotation and Sentiment Analysis techniques. *Automation in Construction* 134, 104068. <https://doi.org/10.1016/j.autcon.2021.104068>
- Hadjidemetriou, G.M., Masino, J., Christodoulou, S.E., Gauterin, F., Brilakis, I., 2020. Comprehensive Decision Support System for Managing Asphalt Pavements. *Journal of Transportation Engineering, Part B: Pavements* 146, 06020001. <https://doi.org/10.1061/JPEODX.0000189>
- Highways England, 2017. Highways England Strategic Road Network Initial Report - Overview.
- Highways, N., 2023. National Highways - National Highways [WWW Document]. URL <https://nationalhighways.co.uk/> (accessed 10.4.24).

- Kingma, D.P., Ba, J., 2017. Adam: A Method for Stochastic Optimization. <https://doi.org/10.48550/arXiv.1412.6980>
- Loshchilov, I., Hutter, F., 2017. Decoupled Weight Decay Regularization.
- Moazami, D., Behbahani, H., Muniandy, R., 2011. Pavement rehabilitation and maintenance prioritization of urban roads using fuzzy logic. *Expert Systems with Applications* 38, 12869–12879. <https://doi.org/10.1016/j.eswa.2011.04.079>
- OpenAI, 2024. ChatGPT [WWW Document]. URL <https://chatgpt.com> (accessed 9.24.24).
- Rasmy, L., Xiang, Y., Xie, Z., Tao, C., Zhi, D., 2021. Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. *npj Digit. Med.* 4, 1–13. <https://doi.org/10.1038/s41746-021-00455-y>
- Robinson, R., Danielson, U., Snaith, M., 1998. Road Maintenance Management. Macmillan Education UK, London. <https://doi.org/10.1007/978-1-349-14676-5>
- Salihudin, H., Teh, K.T., Muniandy, R., Omar, H., Hassan, A., 2007. A Prototype Expert System for the Selection of Road Construction Materials. *The Journal of Engineering Research [TJER]* 4, 1. <https://doi.org/10.24200/tjer.vol4iss1pp1-10>
- Sanh, V., Debut, L., Chaumond, J., Wolf, T., 2019. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter.
- Tripathi, K.P., 2011. A Review on Knowledge-based Expert System: Concept and Architecture. *Artificial Intelligence Techniques*.
- Turc, I., Chang, M.-W., Lee, K., Toutanova, K., 2019. Well-Read Students Learn Better: On the Importance of Pre-training Compact Models. <https://doi.org/10.48550/arXiv.1908.08962>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is All you Need, in: *Advances in Neural Information Processing Systems*. Curran Associates, Inc.
- Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T.L., Gugger, S., Drame, M., Lhoest, Q., Rush, A.M., 2020. HuggingFace's Transformers: State-of-the-art Natural Language Processing. <https://doi.org/10.48550/arXiv.1910.03771>

EFFICIENT ACCESSIBILITY INTEGRATED MULTI-OCCUPANCY BUILDING LAYOUT GENERATION USING CONSTRAINED DIFFUSION MODELS

Haolan Zhang¹, Ruichuan Zhang²

1) Ph.D. Candidate, Myers-Lawson School of Construction, Virginia Polytechnic Institute and State University, Blacksburg, VA, US., Email: haolanz@vt.edu

2) Ph.D., Asst. Prof., Myers-Lawson School of Construction, Virginia Polytechnic Institute and State University, Blacksburg, VA, US., Email: ruichuanz@vt.edu

Abstract: Designing complex, multi-occupancy building layouts requires considering complicated spatial arrangements, design constraints, and strict accessibility requirements. Despite advancements in automatic building layout generation, existing methods struggle to address the complexities of these layouts and often overlook critical accessibility features. This paper presents a novel deep learning-based approach for generating complex, multi-occupancy building layouts that meet both architectural and accessibility standards. To improve the efficiency of training, non-corridor rooms are approximated by minimum rotation rectangles, while a graph neural network (GNN) predicts the number of corners for corridors. To address the quadratic complexity of transformers, we incorporate FlashAttention, to enhance computational efficiency. Accessibility features are integrated into the model by enforcing geometric requirements, including room size ratios and maximum corner distances to account for travel distance to egress. Additionally, a distance penalty is introduced in the loss function to ensure compliance with wheelchair clearance requirements. Experimental results show that our approach outperforms baseline models in generating realistic, complex layouts while ensuring compliance with design and accessibility constraints, making it a robust solution for generating multi-occupancy building layouts.

Keywords: Generative Design, Diffusion Model, Graph Neural Network, Multi-occupancy Building Layout

1. INTRODUCTION

Designing building layouts is a critical, resource-intensive aspect of building design, involving iterative collaboration between designers and clients to balance client needs and regulatory constraints such as accessibility requirements. Automatic building layout generation has the potential to revolutionize architectural design by enabling architects and designers to produce layouts that meet specific criteria using algorithms and machine learning models.

Early efforts in this domain relied heavily on rule-based systems and optimization algorithms to generate layouts (Dincer et al., 2014; Nisztuk and Myszkowski, 2019), which, while effective for simpler designs, struggled to scale up for more complex architectural demands. Recent advances in machine learning, especially generative models such as Generative Adversarial Networks (GANs), have provided more sophisticated tools for tackling these challenges. GANs and other deep learning models can generate layouts by learning patterns from large datasets (Nauata et al., 2020; Aalaei et al., 2023; Luo and Huang, 2022). These methods excel at creating diverse and realistic layouts but have primarily been limited to small or single-unit layouts.

However, real-world designs, especially multi-occupancy building layouts, are significantly more complex than single-unit layouts. Multi-occupancy building layouts contain more rooms, more complex spatial arrangements, and stricter accessibility requirements. These accessibility requirements, such as ensuring adequate clearance for wheelchair, and adhering to egress distance requirements, add an additional layer of complexity to the design process. Such constraints are essential to meet building codes, such as the Americans with Disabilities Act (ADA), and to ensure inclusivity in the built environment. These complexities present unique challenges that existing models struggle to address and require more scalable model designs and the integration of constraints throughout the generative process.

To address these challenges, this paper introduces a novel approach for generating accessible multi-occupancy building layouts using a combination of graph neural network (GNN) and a constrained diffusion model. To handle the increased size and complexity of multi-occupancy building layouts, we optimized the input encoding and incorporated more advanced attention mechanisms to improve learning efficiency. We also introduced additional inputs, including the maximum travel distance within each room, and modified the model's loss function to account for wheelchair accessibility requirements. The proposed approach consists of three key steps: encoding building layout and input requirements, predicting the number of corners for corridors, and generating building layouts through a constrained diffusion process. We trained and validated our model using the Modified Swiss Dwellings (MSD) dataset, an open-source collection of multi-apartment building layouts. Our test results demonstrate that the proposed method outperforms existing approaches in generating complex, accessible multi-occupancy building layouts.

2. BACKGROUND

Generative design of building layouts leverages algorithms, ranging from rule-based, optimization, and artificial intelligence (AI) and machine learning, to automatically generate and optimize building layouts. Early methods capitalized on rule-based methods and optimization methods. For example, Dincer et al. (2014) encoded the adjacency of spaces as neighboring rules and used cellular automata to generate layouts within 8 by 8 grids. Nisztuk and Myszkowski (2019) utilized evolutionary algorithms to generate designs that adhere the topology requirements. These methods rely on extensive hand-crafted features and predefined rules, and they lack flexibility in adapting changes in rules. comes to change of rules.

Recently, machine learning methods has become more popular due to its better adaptability and performance in creating more realistic building layouts. These methods primarily focus on single-unit residential floor plan layout, and use datasets such as RPLAN (Wu et al., 2019), LIFULL (LIFULL, 2015). For instance, Wu et al. (2019) employed an encoder-decoder model that can generate convincing layouts. To include room connectivity requirements, Nauata et al. (2020) and Nauata et al. (2021) combined graph neural networks with GANs to generate layouts from bubble diagrams, using LIFULL and RPLAN to develop their model. Shabani et al. (2023) proposed using a constrained diffusion model to generate layouts based on bubble diagrams. Relying solely on bubble diagrams to condition the GAN model often results in arbitrary boundaries, which are not sufficiently practical for real-world applications. To control the output boundary of layouts, Zheng et al. (2020), Aalaei et al. (2023), Luo and Huang (2022), Hu et al. (2020) develop their models using RPLAN dataset and incorporated layout boundaries into the GAN model, constraining the output's boundary shape.

There are also a few research focusing on multi-occupancy building layouts generation. For instance, Wang et al. (2023) used a U-Net with a spatial attention mechanism to generate building layouts, given boundary constraints. They mainly focus on higher-level layout and omit the layout within each apartment in the building layout. Kuhn (2024) used diffusion model to generate building layouts given structural walls and room connectivity, however, the generated layouts remain overly simplified because all rooms are represented as rectangles, and their visual quality is low (Van Engelenburg et al., 2024).

While these methods have succeeded in generating realistic building layouts and incorporating basic requirements such as room connectivity and boundary constraints, they still face two significant limitations. First, existing models struggles with generating complex geometries and topologies in building layouts. Figure 1 compares the multi-apartment building layout dataset MSD (Van Engelenburg et al., 2024) with the RPLAN and LIFULL datasets, which are single-unit floor plan layouts. It is evident that multi-apartment building layout are considerably more complex than single-unit plans, as they involve a larger number of rooms, increased spatial complexity, and more intricate connectivity requirements between units. The sheer volume of rooms and connections requires more efficient encoding mechanisms and a more scalable model architecture capable of managing the additional complexity. Second, current methods largely overlook the integration of accessibility requirements, which are critical to ensuring both regulatory compliance and inclusivity in building design. For example, the International Building Code (IBC) and ADA mandate that the travel distance to an egress in a building with a single exit point must not exceed 22.86 meters. Additionally, wheelchair accessibility standards require that corridors have a minimum width of 1.5 meters to accommodate wheelchair users. These interconnected requirements add further complexity to the building layout generation process, as they must be considered holistically to ensure that the resulting designs are functional, compliant, and accessible. Addressing these two gaps is essential for making automated building layout designs more applicable in real-world architectural projects.

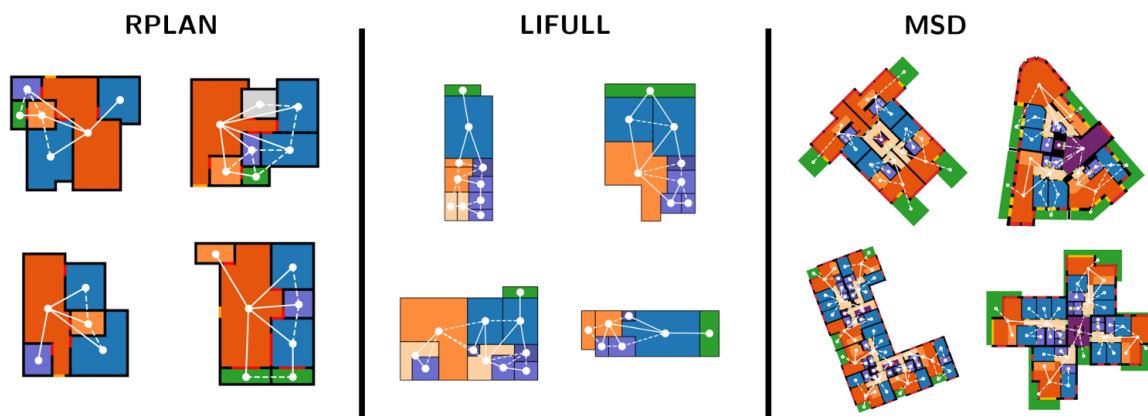


Figure 1. Comparison between floor plan layout datasets (Van Engelenburg et al. 2024)

3. METHOD

Our proposed approach for multi-occupancy building layout generation employs a GNN and a constrained diffusion model, consisting of three primary steps as depicted in Figure 2.

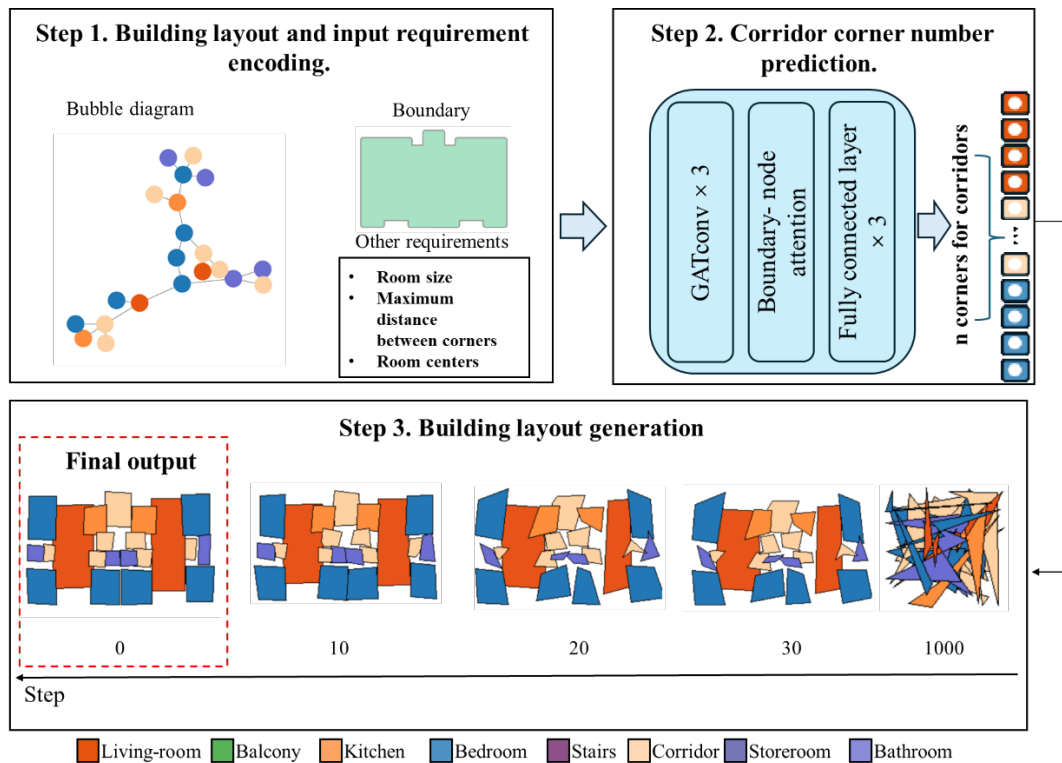


Figure 2. Proposed approach

Step 1. Building layout and input requirement encoding. Each room in a building layout is represented as a polygon, with non-corridor rooms simplified as rectangles with a fixed number of four corners to ensure computational efficiency. The corner number of corridors will be predicted using GNN in Step 2. The input requirement includes a bubble diagram to defines room connectivity and types, building layout's boundary, and size requirements including room sizes and the maximum distances between corners for accessibility compliance. Room center coordinates are also encoded to control the spatial positioning of the generated rooms.

Step 2. Corridor corner number prediction. Corridors, unlike other rooms, require a more flexible representation because their shape depend on the arrangement of adjacent rooms. The corner number for corridors is variable and must be predicted. To handle this, we employ a GNN that predicts the number of corners for each corridor based on the spatial context of the surrounding rooms. The inputs to the GNN include the bubble diagram, room sizes, and room center coordinates, allowing the model to adaptively determine the appropriate corridor corner number, ensuring that it fits seamlessly within the overall building layout.

Step 3. Building layout generation using diffusion model. We use a constrained diffusion model to generate the building layout by progressively refining it through denoising steps. Conditioned on the input encoded in the first step, the model employs a transformer architecture optimized for handling the complexity of multi-occupancy building layouts. Several attention mechanisms are incorporated to allow the transformer to learn the intricate relationships between connected rooms, boundaries, and corner coordinates within each room. Furthermore, a modified loss function is introduced to ensuring wheelchair clearance requirements.

3.1 Building layout and Input Requirement Encoding

(1) Building Layout Encoding

Given a building layout F , Let $F = \{R_1, R_2, R_3, \dots, R_N\}$, where R_N is the room in F . R_N is represented as a polygon, which are a sequence of 2D coordinates. For each R_i in F , $R_i = \{c_{i,1}, c_{i,2}, c_{i,3}, \dots, c_{i,N}\}$. $c_{i,N}$ is the 2D coordinates and N is the number of corners. The challenge for such encoding lies in determining the number of corners N . A simple approach would be to fix N as the maximum number of corners found in the dataset. While this might work for single-unit layouts, it becomes highly inefficient for complex layouts. For instance, a 40-room layout could require over 1000 corner coordinates, making transformer-based models, with their quadratic complexity, impractical to train. Some of the previous studies (Nauata et al., 2018; Van Engelenburg et al., 2024) treat each room as a rectangle, but this simplification results in a significant loss of detail. To balance

efficiency and detail retention, we simplify non-corridor rooms by approximating them using a minimum bounding rectangle, as most non-corridor rooms are close to rectangular. However, corridors, with their irregular and variable shapes, depend heavily on the configuration of surrounding rooms and cannot be effectively approximated as rectangles. For these, we use a GNN to predict the number of corners. This approach reduces the representation to fewer than 200 corner coordinates for a 40-room layout, making it computationally feasible while preserving essential detail. The GNN model is discussed in detail in Section 3.2. Figure 3 illustrates how non-corridor rooms are simplified into rectangles while preserving the original shapes of the corridors (beige color).

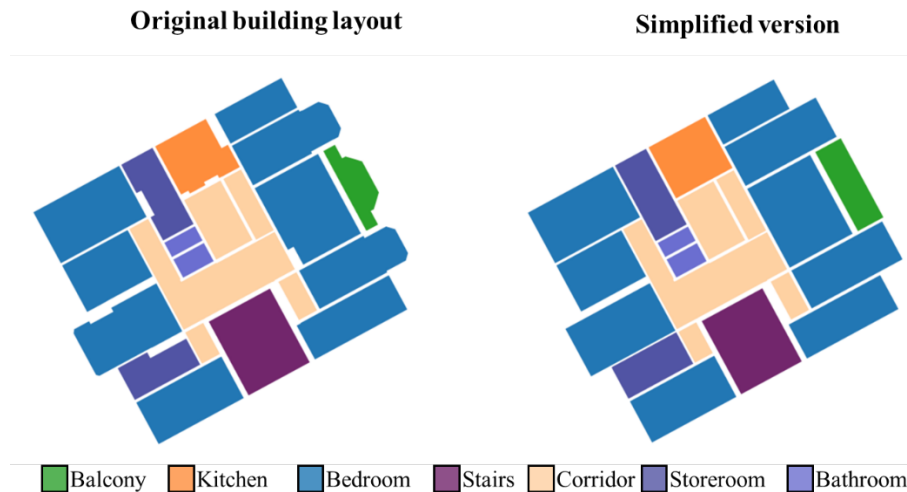


Figure 3. Example of simplified building layout

(2) Boundary Encoding

The building layout boundary defines the overall shape of the layout. Similar to building layout, the boundary is represented as a polygon formed by a sequence of 2D corner coordinates, which can be expressed as $bd = \{C_1, C_2, C_3, \dots, C_N\}$, where C_N represents the 2D coordinates of the boundary corners.

(3) Bubble Diagram Encoding

The bubble diagram represents room types and their connections, encoded as an adjacency matrix A , where each element A_{ij} being defined by as Equation (1)

$$A_{ij} = \begin{cases} 0 & \text{if room } i \text{ is connected to room } j \\ 1 & \text{otherwise} \end{cases} \quad (1)$$

The adjacency matrix also serves as an attention mask in the transformer model, directing the network's focus during training. A '0' in the matrix indicates a direct connection between two rooms.

(4) Room Size and Room Center Encoding

Room size and center encodings capture additional accessibility and spatial requirements. The room size requirement includes two components: the room size ratio, represented as the ratio of the room area to the total building layout area, ensuring each room meets the specified size constraints, and the maximum distance between corners, which represents the longest possible travel distance within a room. This distance is crucial for calculating the travel distance to the nearest egress and ensuring compliance with egress requirements. The room center is encoded as 2D coordinates, which control the spatial positioning of the generated rooms within the building layout. These encodings ensure that the model maintains spatial coherence and adheres to critical size and accessibility requirements during the generation process.

3.2 Corridor Corner Number Prediction

Since corridors heavily depend on the spatial arrangement of surrounding rooms, we employ a GNN to predict the number of corners for corridors dynamically. To achieve this, we use a Graph Attention Network (GAT) due to its ability to efficiently model the complex relationships between rooms. This is particularly beneficial for corridor corner prediction, as the shape and size of corridors are influenced by internal room connectivity. The node features in the graph are comprised of the room type, room size, and room center coordinates. The edges in the graph represent whether two rooms are connected, based on the adjacency matrix defined in the bubble diagram. As shown in Figure 2, we use three graph attention convolutional (GATConv) layers in our model.

Additionally, an extra input for the building layout boundary is incorporated into the model as the boundary also influence the shape of corridor. The boundary input is processed using a linear layer to align its features with the room-based graph. After this, an attention layer combines the boundary features and the graph features by computing their relevance, which helps the model attend to important boundary information when

predicting corridor corner numbers.

3.3 Diffusion Model for Building Layout Generation

(1) Forward and Backward Processes

During the forward diffusion process, Gaussian noise is progressively added to these corner coordinates, gradually transforming the original building layout into random noise. This process follows Equation (2). where γ_t is a noise schedule that transitioning from 1 to 0, and $\epsilon \sim N(0, I)$ represents Gaussian noise. This process transforms x_0 into random noise x_t

$$x_t = \sqrt{\gamma_t}x_0 + \sqrt{(1 - \gamma_t)}\epsilon \quad (2)$$

The reverse process leverages a transformer model to iteratively remove the noise, gradually reconstructing the building layout from its noisy state. The denoising is conditioned on inputs discussed in Section 3.1 ensuring that the generated building layout adheres to design constraints and accessibility requirements. This iterative refinement allows the model to generate accurate and compliant building layouts.

(2) Feature Embedding

Figure 4 shows the transformer architecture. The model utilized four embeddings: building layout, condition, boundary, and time embeddings. All embeddings are projected into a 512-dimensional feature space for consistency.

Building Layout Embedding. The building layout coordinates are projected into a 512-dimensional space using a linear layer.

Condition Embedding. Room features, including index, type, corner index, size, maximum corner distance, and room center coordinates, are encoded. The index and type are represented as one-hot vectors, room center coordinates are encoded as 2D coordinates, and size and corner distance are encoded as scalars. These features are then concatenated and expanded to 512 dimensions using a linear layer.

Boundary Embedding. The input boundary is encoded and projected into a 512-dimensional vector using a linear layer, ensuring boundary constraints are incorporated during generation.

Time Embedding. The diffusion step is encoded as a scalar and projected into 512 dimensions, allowing the model to track the generation process over time.

Finally, the building layout, condition, and time embeddings are concatenated to form the input for the model.

(3) Transformer Architecture

The transformer model in our approach utilizes multiple new attention mechanisms (Dao et al., 2022), replacing standard attention mechanisms to address the quadratic complexity issue in typical transformer architectures. FlashAttention significantly reduces both computational cost and memory requirements, enabling efficient processing of longer sequences and allowing us to handle complex, multi-occupancy building layouts with greater scalability. This optimization ensure that our model can generate intricate layouts without the prohibitive computational overhead. In our experiments, training time was reduced by half when using FlashAttention compared to traditional attention mechanisms.

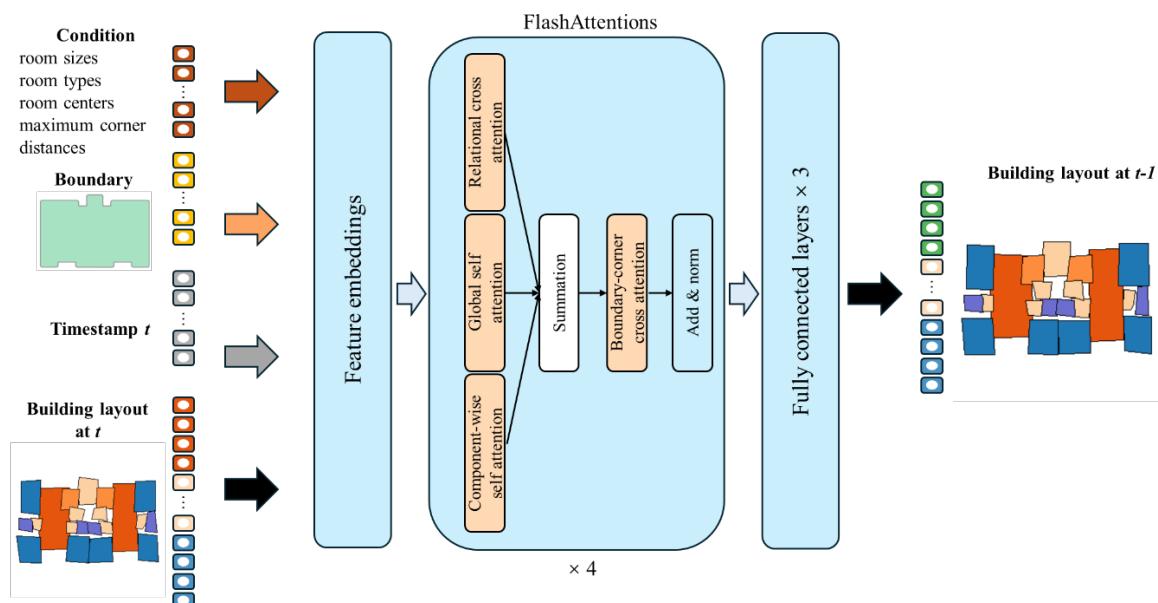


Figure 4. Transformer architecture

We first integrate three attention mechanisms adapted from Shabani et al. (2023): component-wise self-attention within rooms, global self-attention across all corners, and relational cross-attention between connected rooms.

To incorporate the building layout boundary, we use a boundary-to-corner cross-attention module (Equation (3)) to capture interactions between boundary corners and room corners, ensuring proper alignment between the room layout and the external boundary.

$$Attention_{boundary-to-corner}(Q, K_b, V_b) = softmax\left(\frac{QK_b^T}{\sqrt{d_b}}\right)V_b \quad (3)$$

In this equation, Q is derived from a concatenation of the condition embedding, time embedding, and room corner embedding features. K_b and V_b are the key and value matrices from the boundary embeddings, and d_b is the dimensionality of the boundary key vectors.

Additionally, we modified the loss function by incorporating a distance penalty (Equation (4)), which enforces a minimum distance between room corners. This penalty ensures compliance with accessibility requirements by maintaining wheelchair clearance.

$$penalty_{distance} = \sum_{i=1}^n max(0, d_{min} - d(p_i, p_{i+1})) \quad (4)$$

Where n is the corner number of room, p_i and p_{i+1} represent consecutive corner points of a room. $d(p_i, p_{i+1})$ is the Euclidean distance between two consecutive corner points. d_{min} is the minimum required distance, in our experiment, we set d_{min} as the smallest minimum distance between room corners in ground truth. The penalty works by penalizing cases where the distance between consecutive corners falls below the specified minimum d_{min} . This encourages the model to generate rooms that maintain sufficient space for wheelchair clearance between their corners.

The final loss function combines the corner coordinate mean squared error (MSE) loss with the distance penalty to balance the precision of corner positioning and adherence to accessibility constraints. As shown in Equation (5).

$$Loss = (1-\lambda) MSE_{corners} + \lambda \cdot penalty_{distance} \quad (5)$$

Where λ is a hyperparameter controlling the influence of the distance penalty relative to the corner MSE loss. In our experiment, we set $\lambda = 0.5$.

4. EXPERIMENTS AND RESULTS

4.1 Dataset Preparation

For our experiments, we employed the MSD dataset, which contains over 5,300 building layouts of medium- to large-scale building complexes, encompassing more than 18,900 distinct apartments. Each building layout includes annotations for rooms, structural components, and bubble diagrams, offering a diverse range of room shapes. The dataset provides 13 room type categories, essential for capturing the diverse spatial arrangements and functional requirements of real-world designs.

For the cleaning process, we removed building layouts with an excessive number of rooms to ensure consistency and manageability in training. This filtering step helped eliminate overly complex layouts that could introduce noise, ensuring the dataset remained representative of typical multi-apartment designs while maintaining computational efficiency. Additionally, we observed that the original dataset was not axis-aligned. To address this, we rotated the building layouts to align them with the axes, which also served as a data augmentation technique to improve model generalization. Both the original, non-axis-aligned and the newly axis-aligned building layouts were used during training. In total, we used 11,166 building layouts for training, 300 for validation, and 712 for testing.

4.2 Model Training

The model was trained using PyTorch with the Adam optimizer and a batch size of 128. Training lasted for 84,000 steps, starting with a learning rate of 1e-3, which was reduced by a factor of 10 after every 50,000 steps. The diffusion process was structured over 1,000 steps with uniformly sampled time steps.

4.3 Evaluation Metrics

Graph compatibility. This metric evaluates how well the generated building layout matches the input bubble diagram in terms of room connectivity. The graph compatibility is defined in Equation (6).

$$Compatibility(Q, K) = \frac{1}{S} \sum_{i=1}^S \frac{1}{|E_i^K|} \sum_{e \in E_i^K} 1[e \in E_i^Q] \quad (6)$$

where S represents the total number of test samples, E_i^K represents the set of edges in the graph extracted from the generated building layout K , and E_i^Q represents the edges from the input bubble diagram Q . $1[e \in E_i^Q]$ is an indicator function that equals 1 if the edge e exists in both E_i^K and E_i^Q , and 0 otherwise.

Boundary mean intersection over union (MIOU). This metric evaluates the overlap between the

input and generated building layout boundaries. The MIoU is defined in Equation (7).

$$\text{MIoU}_{\text{boundary}}(B_{\text{in}}, B_{\text{out}}) = \frac{1}{S} \sum_{i=1}^S \frac{|B_{\text{in}} \cap B_{\text{out}}|}{|B_{\text{in}} \cup B_{\text{out}}|} \quad (7)$$

where B_{in} represents the input boundary, B_{out} represents the output boundary, and the IoU measures the overlap between the input and generated boundaries. S is the number of samples.

Size requirement MSE. The MSE is used to measure the combined difference between the predicted room sizes and the maximum distances between room corners, compared to the ground truth, as shown in Equation (8).

$$\text{MSE}_{\text{size}} = \frac{1}{S} \sum_{i=1}^S \left[(\hat{s}_i - s_i)^2 + (\hat{d}_{\text{max},i} - d_{\text{max},i})^2 \right] \quad (8)$$

where s_i is the actual room size, \hat{s}_i is the room size calculated from generated building layouts, $d_{\text{max},i}$ is the actual maximum corner distance, and $\hat{d}_{\text{max},i}$ is the maximum corner distance calculated from generated building layouts. S is the number of samples.

Wheelchair clearance compliance percentage. We measure the percentage of rooms that comply with the required wheelchair clearance. This metric is defined in Equation (9).

$$\text{Compliance} = \frac{\text{Number of compliant rooms}}{\text{Total number of rooms}} \times 100\% \quad (9)$$

This metric calculates the proportion of rooms that meet the accessibility requirements for wheelchair clearance.

4.4 Experimental Results and Discussion.

Figure 5 showcases sample output building layouts alongside the ground truth, demonstrating the performance of our proposed approach. The generated plans closely adhere to the input requirements, with room locations and sizes highly similar to the ground truth. Although boundary alignment is not always perfect, the model consistently produces layouts that are remarkably close to the ground truth boundaries.

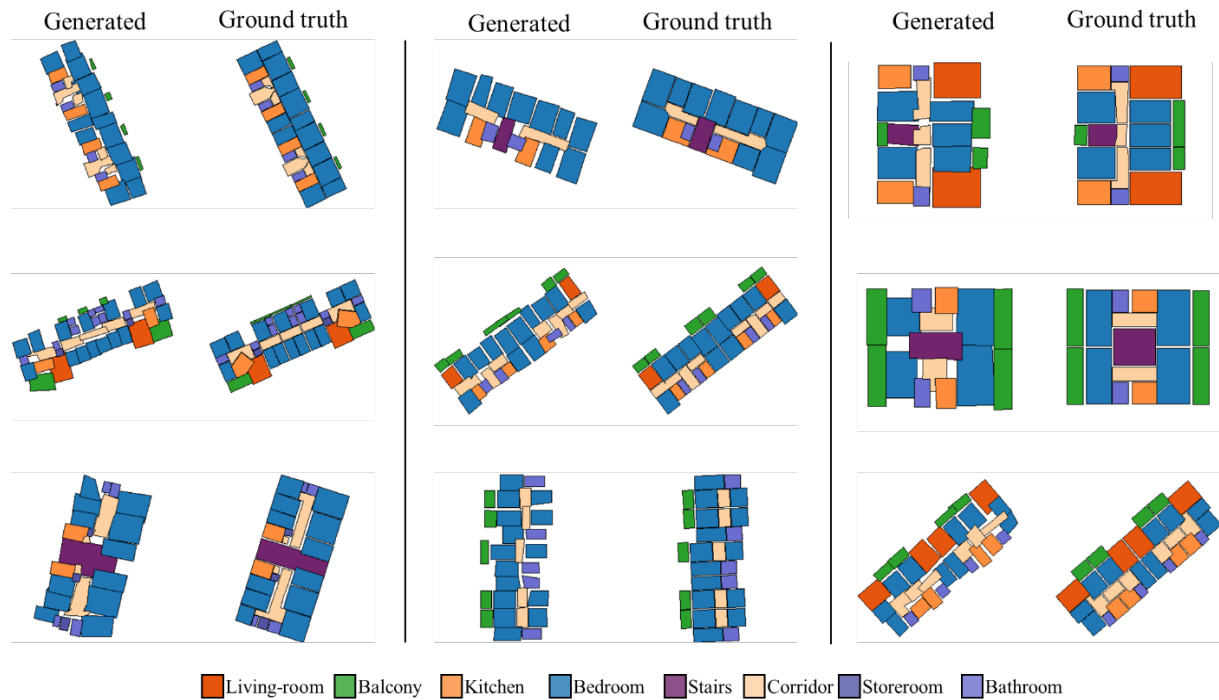


Figure 5. Generated samples

Table 1 presents the graph compatibility and boundary MIoU for our proposed approach compared to the baseline model across different room count intervals. The baseline model is a simplified version that does not incorporate data augmentation or boundary inputs. In terms of graph compatibility, our approach consistently outperforms the baseline in graph compatibility across all room count intervals, with an average improvement of 3%. These results demonstrate the effectiveness of data augmentation in capturing the room connectivity relationships. Regarding boundary MIoU, our approach surpasses the baseline by an average of 16%. These improvements suggest that the proposed approach better maintains alignment between the generated and input

boundaries, benefiting from the inclusion of boundary input in the training process.

Table 1. Graph compatibility and boundary MIoU test results

Room count interval	Proposed approach		Baseline	
	Graph compatibility ↑	Boundary MIoU ↑	Graph compatibility ↑	Boundary MIoU ↑
10-20	0.78	0.73	0.74	0.58
20-30	0.79	0.74	0.76	0.59
30-40	0.82	0.73	0.8	0.58
40-50	0.82	0.74	0.79	0.56

Note: bold=best performance; ↑ indicates that higher values are better. ↓ indicates that lower values are better.

Table 2 compares the Size requirement MSE and wheelchair clearance compliance percentage for our proposed approach and the baseline model across different room count intervals. The baseline model does not include the maximum length input or the distance penalty in the loss function, which impacts its ability to enforce accessibility constraints. For the room size and corner distance MSE, the baseline has comparable results in the lower room count intervals. However, as the room count increases, our proposed approach shows improvement, achieving 0.01 of improvements. For wheelchair clearance compliance percentage, our proposed approach outperforms the baseline in the latter three room count intervals and with an average improvement of 1.7%. This demonstrates that the inclusion of the maximum length input and distance penalty leads to better accessibility compliance, especially in larger and more complex layouts.

Table 2. Size requirement MSE and wheelchair clearance compliance percentage

Room count interval	Proposed approach		Baseline	
	Size requirement MSE ↓	Wheelchair clearance compliance percentage ↑	Size requirement MSE ↓	Wheelchair clearance compliance percentage ↑
10-20	0.07	0.94	0.07	0.94
20-30	0.06	0.96	0.06	0.95
30-40	0.05	0.97	0.06	0.95
40-50	0.04	0.97	0.05	0.95

Note: bold=best performance; ↑ indicates that higher values are better. ↓ indicates that lower values are better.

5. CONCLUSIONS

In this paper, we introduced a novel approach for generating complex multi-occupancy building layouts by combining a graph neural network (GNN) with a constrained diffusion model. To improve the efficiency of training on complex building layouts, we optimized the building layout encoding by approximating non-corridor rooms with minimum bounding rectangles and using a GNN to predict the number of corners for corridors. Additionally, to address the quadratic complexity of transformers, we employed FlashAttention, which significantly reduces computational cost and memory usage, making the model more efficient. To incorporate accessibility features, we introduced room size requirements, including room size ratios and maximum corner distances, alongside fundamental constraints such as layout boundaries and bubble diagrams. We also designed a distance penalty to ensure compliance with wheelchair clearance regulations.

Our model significantly outperforms the baseline across several key metrics, including graph compatibility, boundary alignment, and wheelchair clearance compliance. Specifically, the proposed approach improves graph compatibility by an average of 3%, boundary alignment by 16%, room size and corner distance MSE by 0.01, and wheelchair clearance compliance by 1.7%. These results demonstrate the model's capability to generate building layouts that not only adhere closely to design constraints but also meet important accessibility requirements.

Looking forward, we identified three key areas for improvement. (1) *Refining corridor shapes*. The current approach approximates corridors but does not fully capture their complex and irregular geometries. Further refining the model to better represent corridor shapes will enhance the practicality of the generated building layouts. (2) *Addressing overlapping and gaps*. Although our model performs well overall, issues with room overlaps and gaps between rooms still arise in more complex layouts. Developing more robust mechanisms to prevent overlaps and close gaps will improve the usability of the building layouts. (3) *Introducing more accessibility features*. We will expand the scope to include additional accessibility features such as door widths, accessible bathroom layouts and will increase the model's applicability for universal design and compliance with more comprehensive accessibility standards.

REFERENCES

- Aalaei, M., Saadi, M., Rahbar, M., and Ekhlassi, A. (2023). Architectural layout generation using a graph-constrained conditional generative adversarial network (GAN). *Automation in Construction*, 155, 105053.
- Dao, T., Fu, D., Ermon, S., Rudra, A., and Ré, C. (2022). Flashattention: Fast and memory-efficient exact attention with io-awareness. *Advances in Neural Information Processing Systems*, 35, 16344-16359.
- Dincer, A. E., Cagdas, G., and Tong, H. (2014). A digital tool for customized mass housing design. *Fusion, Proceedings of the 32nd eCAADe Conference, Department of Architecture and Built Environment*, pp. 10-12.
- Hu, R., Huang, Z., Tang, Y., Van Kaick, O., Zhang, H., and Huang, H. (2020). Graph2plan: Learning floorplan generation from layout graphs. *ACM Transactions on Graphics (TOG)*, 39(4), 118-1.
- Kuhn, E. (2023). Adapting HouseDiffusion for conditional Floor Plan generation on Modified Swiss Dwellings dataset. *arXiv preprint arXiv:2312.03938*.
- LIFULL Co., Ltd. (2015). LIFULL HOME'S Snapshot Data of Rentals, Informatics Research Data Repository, *National Institute of informatics*.
- Luo, Z., and Huang, W. (2022). FloorplanGAN: Vector residential floorplan adversarial generation. *Automation in Construction*, 142, 104470.
- Nauata, N., Chang, K. H., Cheng, C. Y., Mori, G., and Furukawa, Y. (2020). House-gan: Relational generative adversarial networks for graph-constrained house layout generation. *Computer Vision-ECCV 2020: 16th European Conference*, Glasgow, UK, pp. 162-177.
- Nauata, N., Hosseini, S., Chang, K. H., Chu, H., Cheng, C. Y., and Furukawa, Y. (2021). House-gan++: Generative adversarial layout refinement network towards intelligent computational agent for professional architects. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13632-13641.
- Nisztuk, M., and Myszkowski, P. B. (2019). Hybrid evolutionary algorithm applied to automated floor plan generation. *International Journal of Architectural Computing*, 17(3), 260-283.
- Shabani, M. A., Hosseini, S., & Furukawa, Y. (2023). Housediffusion: Vector floorplan generation via a diffusion model with discrete and continuous denoising. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5466-5475.
- Van Engelenburg, C., Mostafavi, F., Kuhn, E., Jeon, Y., Franzen, M., Standfest, M., ... and Khademi, S. (2024). MSD: A Benchmark Dataset for Floor Plan Generation of Building Complexes. *arXiv preprint arXiv:2407.10121*.
- Wang, L., Liu, J., Zeng, Y., Cheng, G., Hu, H., Hu, J., and Huang, X. (2023). Automated building layout generation using deep learning and graph algorithms. *Automation in Construction*, 154, 105036.
- Wu, W., Fu, X. M., Tang, R., Wang, Y., Qi, Y. H., and Liu, L. (2019). Data-driven interior plan generation for residential buildings. *ACM Transactions on Graphics (TOG)*, 38(6), 1-12.
- Zheng, H., Keyao, A. N., Jingxuan, W. E. I., and Yue, R. E. N. (2020). Apartment floor plans generation via generative adversarial networks. *25th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA 2020): RE: Anthropocene, Design in the Age of Humans*, pp. 601-610.

MULTI-STORY FLOOR PLAN GENERATION FROM BUILDING VOLUMETRIC DESIGN USING GRAPH NEURAL NETWORK

Atsuya Ogura¹, Tomohiro Fukuda², Nobuyoshi Yabuki³, and Taro Narahara⁴

1) M.S. Student., Division of Sustainable Energy and Environmental Engineering, Graduate School of Engineering, Osaka University, Japan. Email: aogura0207yabuki@gmail.com

2) Ph.D., Assoc. Prof., Division of Sustainable Energy and Environmental Engineering, Graduate School of Engineering, Osaka University, Japan. Email: fukuda.tomohiro.see.eng@osaka-u.ac.jp

3) Ph.D., Prof., Division of Sustainable Energy and Environmental Engineering, Graduate School of Engineering, Osaka University, Japan. Email: yabuki@see.eng.osaka-u.ac.jp

4) Ph.D., Assoc. Prof., Hillier College of Architecture and Design (HCAD), New Jersey Institute of Technology (NJIT), Newark, NJ, USA. Email: taro.narahara@njit.edu

Abstract: The automatic design of architectural floor plans using deep learning has been widely studied to assist architectural design. Traditionally, floor plans generated by deep learning have been limited to single floors. Recently, research has been developing the use of graph neural networks (GNNs), which applies deep learning to graph data to generate building volumes that consider the in-building spatial use. Although these studies aim to generate new building volumes, practical architectural design often requires the generation of floor plans within predefined building outlines, constrained by various legal and regulatory requirements. This study proposes a method for generating multi-story floor plans in a given building volume by using a graph convolutional network (GCN), which adapts convolutional operations to graph data representing the given building volume. The implemented GCN model successfully predicted, with an accuracy of 74.66%, the spatial use class for each node within the graph representing the building. This enables the generation of detailed floorplans across multiple floors. This research contributes to the design support of multi-story floor plans in a given building volume. Moreover, when integrated with the latest 3D generative AI technologies, this approach promises to advance the automatic creation of 3D building models with comprehensive interior designs, starting from scratch in volumes initially devoid of any interior information.

Keywords: Floorplan generation, 3D building layout, Graph neural network, Deep learning, Generative design

1. INTRODUCTION

The design of building volumes and floor plans is conducted at the early stages of architectural design. Therefore, it is a crucial design process that significantly influences subsequent detailed architectural, equipment, and structural design (Zhong et al. 2023). Automated design of building volumes and floor plans enables the rapid presentation of diverse design alternatives to design decision makers. Additionally, from the diverse automatically generated designs, the optimal design can be selected based on various objective functions such as livability, environmental impact, building operability, and structural stability, which are significant motivations for automating the design of building volumes and floor plans (Weber et al. 2022).

The attempt to automate floor plan design using computers began in the early 1970s (Friedman 1971). Since then, there have been three main types of methods for automatic floorplan generation: 1) Bottom-up methods, 2) Top-down methods, and 3) Referential methods (Weber et al. 2022). Of these, Bottom-up methods and Top-down methods are antithetical methodologies, while Referential methods widely refer to methods that use machine learning and deep learning, and are sometimes applied to Bottom-up methods and Top-down methods.

Bottom-up methods generate floor plans by taking required space uses, such as offices and break rooms, as elements, connecting them, and aggregating these elements. The generated floor plans are evaluated based on the relationships between the rooms (Merrell et al. 2010) and environmental factors such as sunlight conditions (Yi and Yi 2014), thereby producing the most suitable floor plan for the intended objectives.

Top-down methods are used when designing within a predefined building envelope, particularly when architectural shapes are strongly constrained by building regulations or regional treaties, or when renovating existing structures. This method subdivides the input building envelope, such as building volume and floorplan outline, to generate a floor plan that best meets desired objectives (Baušys and Pankrašovaitė 2005; Rodrigues et al. 2014).

Referential methods generate floor plans by referencing databases of past floor plans, similar to how architectural designers learn from past architectural works. Deep learning used for referential methods include models that learn image data of floor plans (Wu et al. 2019; Chaillou 2020) and models that learn graph data (Hu et al. 2020; Nauata et al. 2021). However, deep learning models that learn and generate image data have difficulty generating floor plans for multiple floors. This is because the dataset of floorplan image used for deep learning is a set of single floorplan images (Weber 2022). On the other hand, deep learning methods that learn graph data, which are called graph neural networks (GNNs), have prior studies, Building-GNN (Zhong et al. 2023) and Building-GAN (Chang et al. 2022), which generate multi-story floorplans.

Graph data refers to data consisting of a set of nodes and edges. On this graph data, GNNs perform tasks such as node classification, edge link prediction, and graph generation. Many types of data, including molecular structures, power networks, and social networks, can be described as graph data, and GNNs have been applied in various fields such as chemistry, pharmacology, and transportation (Zhou et al. 2020). Their application is also anticipated in the construction industry (Jia et al. 2023).

Both Building-GNN (Zhong et al. 2023) and Building-GAN (Chang et al. 2022), represent building volumes with spatial use information as graph data and use GNN to generate unknown building volumes. This is a study applying the Referential method to the Bottom-up method. However, no study has been proposed to apply GNNs to the Top-down method, which generates multiple floor plans from predefined building volumes.

Therefore, the objective of this study is to generate multi-story floorplans in a given building volume. This study proposed and implemented a method to predict the spatial use in a given building volume and generate floor plans for multiple floors by developing a GNN, specifically graph convolutional network (GCN), which extends convolutional operations to graph data (Kipf and Welling 2017) and predicts which spatial use a node in the graph belongs to.

This research contributes to design support for designing floor plans that span multiple floors within a given building volume. In addition, the proposed method of adding floor plans as a starting point for interior design to a building 3D model that does not have design information in the building volume is expected to contribute to the generation of building 3D models with detailed interior design when combined with the recently developed 3D generation AI.

2. RELATED WORK

2.1 Architectural Floorplan Generation Using GNNs

Previous studies applying deep learning to generate floorplans can be broadly classified into two types: methods that learn from image data and methods that use graph data as constraints.

In the methods of generating floorplans by training image data to deep learning, it is first required to unify the floorplan images of the dataset to be trained, such as 128×128 pixels or 256×256 pixels (Wu et al. 2019). This limits the scale of the floorplan to a relatively small scale. Furthermore, since the training images are single-floor plans, the generated floor plans are also restricted to being single-floor (Weber et al. 2022).

On the other hand, the method of generating floor plans by providing graph data to deep learning as a constraint is to first represent rooms as nodes and connect movable rooms with edges to represent the relationship between rooms in a graph (bubble diagram). The floorplan is then generated by evaluating the generated floorplan with the ground truth floorplan while maintaining the constraints between rooms in the bubble diagram, and then sequentially improving the floorplan to make it more similar to the ground truth floorplan. The floorplans can be generated bubble diagram, making it easier for users to generate the desired floorplans (Hu et al. 2020; Nauata et al. 2021).

In all these cases, there is no direct way to generate multiple-story floor plans. However, there are studies that use GNNs to learn the connection and arrangement of spatial uses in a 3D building model and generate building volumes that consider the connection and arrangement of spatial uses. In these studies, the generated building volume is cut horizontally, and this cut plane is used as a floor plan to create floor plans for multiple floors. The method proposed in this study also predicts the interior spatial use of a given building volume and generates a multi-story floor plan by cutting the building volume horizontally.

2.2 Building Volume to Multi-Story Floorplans

Common representation methods for building 3D models include voxel representation, point clouds, mesh representation, and neural fields (Wang et al. 2023). Of these, the representation method used by previous studies that apply GNNs to generate building 3D models is based on voxel representation.

Building-GNN (Zhong et al. 2023) generates the overall building volume by combining GNN and recurrent neural network, successively adding new building volumes based on the partial building volume initially provided by the designer. Both the initial building volume and the additional building volumes contain class information about the spatial use in the building. As a result, the final generated building model retains this internal usage information. By taking horizontal cross-sections of this building volume with usage information at each floor level, it is possible to obtain floor plans that span multiple floors.

Building-GAN (Chang et al. 2022) proposes a method to generate a new building volume by taking as input a bubble diagram of a floor plan spanning multiple floors and the designable area of the building. Building-GAN, like Building-GNN, generates a building volume with information on the building's internal spatial use, so that by obtaining a cut plane for each floor of this volume, a floor plan for multiple floors can be obtained. However, both Building-GNN and Building-GAN applied GNNs to bottom-up methods for floorplan generation, and there

is no research on the application of GNNs to top-down methods for generating floorplans for multiple floors within a predefined building volume.

3. PROPOSED METHOD

This study proposes a method to automatically generate multi-story floor plans in a given building volume. The input is a building volume divided horizontally by reference lines and vertically by floor heights, and the output is multiple floor plans within the building volume. In the framework of this automatic floorplan generation, the main process consists of the three steps shown in Figure 1: Step 1. Conversion from Input Building Volume to Unclassified Voxel Graph; Step 2. Conversion from Unclassified Voxel Graph to Classified Voxel Graph using GNN; Step 3. Conversion from Classified Voxel Graph to Building Volume with Spatial Usage Class. By going through these three steps, the given building volume is converted into a building volume with spatial usage class. Multi-story floor plans are obtained by creating a sectional view of this building volume with classes of in-building spatial use.

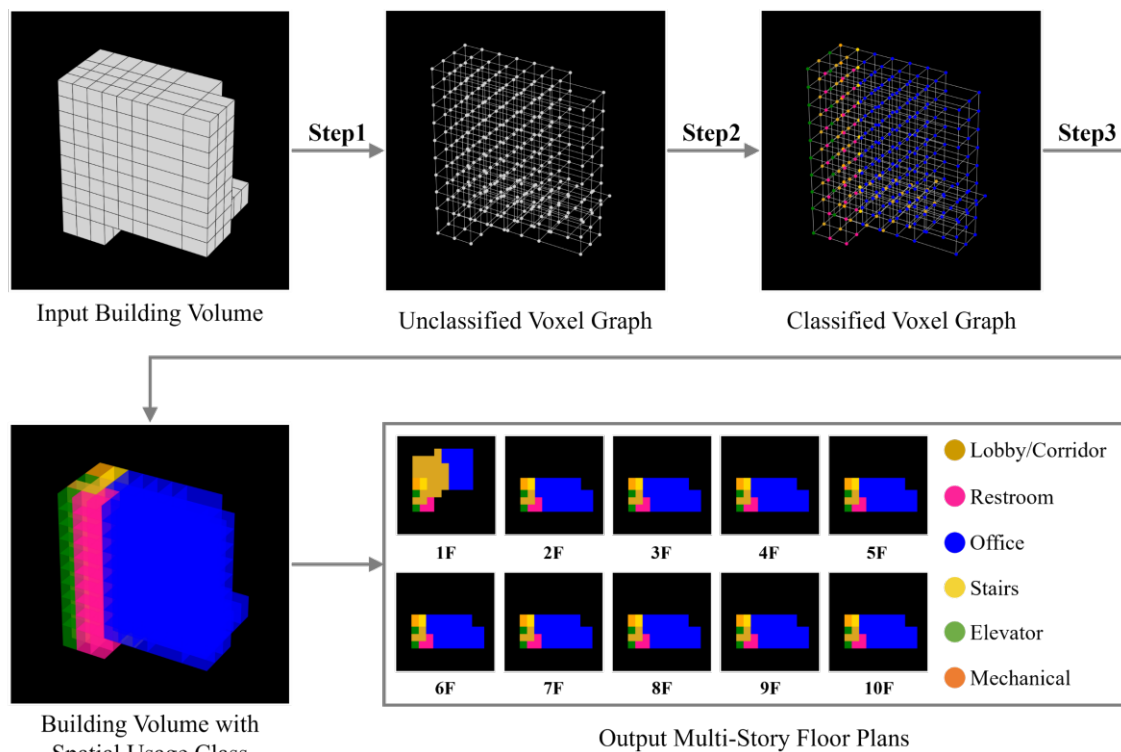


Figure 1. Overview of Proposed Method

3.1. Voxel Graph

Voxel graph is a method proposed by Building-GAN to represent building volumes as a 3D lattice graph (Chang et al. 2022). This 3D lattice graph is set up as follows. First, the voxel-represented building volume is divided by the reference lines used for design. The original building volume is then represented as a set of cuboids, as shown in Figure 2 (a). A 3D lattice graph is obtained by taking the center points of these cuboids and connecting the lattice points with edges when the cuboids share a face with a neighboring cuboid. Voxel graph is a graph in which these grid points are nodes, and the node's attribute information includes the node's coordinates, the node's class information (the class that represents the spatial use in the building), and the height, width, and height of the building voxel that a node represents. This study used the dataset of voxel graph provided by Chang et al. (2022), which is a dataset of 120,000 office buildings generated according to patterns and rules provided by professional architects. Figure 2 (b) shows a voxel graph where each node has either six classes information (Lobby/Corridor, Restroom, Office, Stairs, Elevator, and Mechanical).

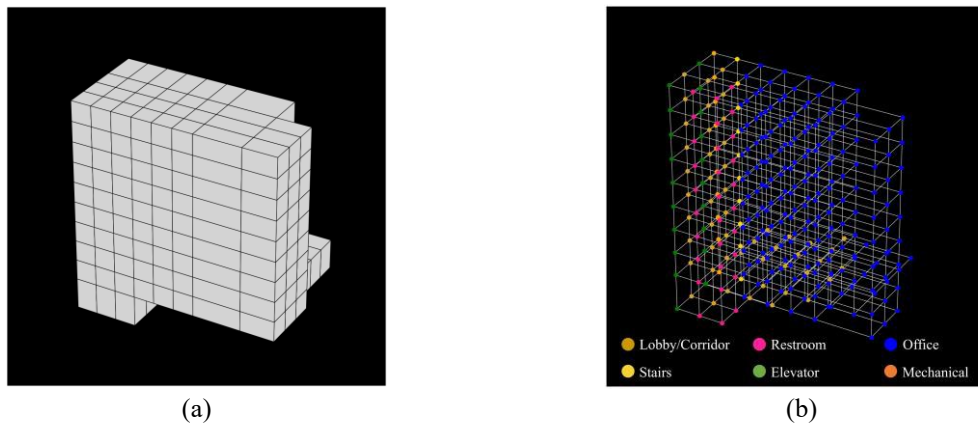


Figure 2. (a) Building Volume. (b) Voxel Graph.

3.2. Conversion from Input Building Volume to Unclassified Voxel Graph

As shown in Figure 1, Step 1 converts an architectural volume into an unclassified voxel graph.

(1) Unclassified Voxel Graph and Classified Voxel Graph

In this study, a voxel graph whose nodes do not have in-building spatial use class information is called an unclassified voxel graph (Figure 3 (a)), and a voxel graph whose nodes have in-building spatial use class information is called a classified voxel graph.

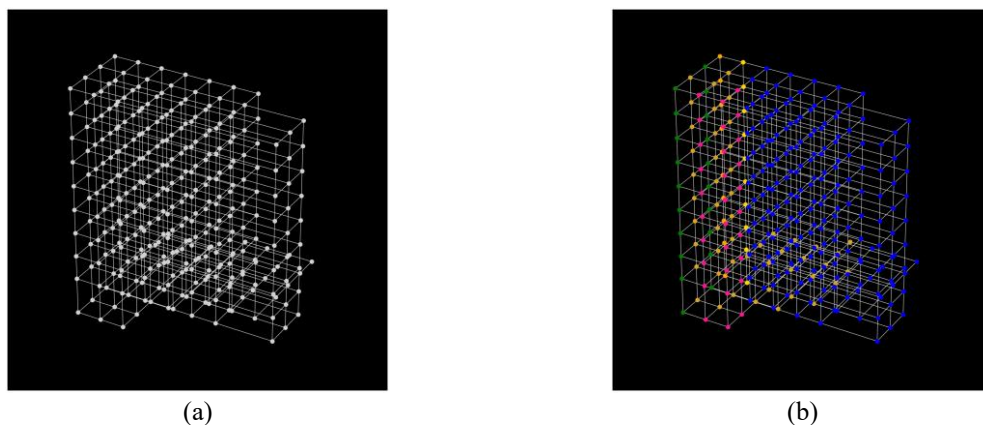


Figure 3. (a) Unclassified Voxel Graph. (b) Classified Voxel Graph.

(2) Building Volume with Any Representation Methods to Unclassified Voxel Graph

In the method proposed in this study, a given building volume refers to the volume and shape of the space occupied by a building. The representation method of the building volume is not specified. The method of representation can be voxel representation, mesh representation of the volume surface, volume rendering, or any other arbitrary method. When a building volume created using any representation method is divided by the design baseline, the building volume is divided into a set of cuboids, regardless of the method of representation. By obtaining nodes at the center points of these cuboids, an unclassified voxel graph can be obtained.

3.3. Conversion from Unclassified Voxel Graph to Classified Voxel Graph using GNN

In Step 2, the unclassified voxel graph is converted to a classified voxel graph using GNN. Step 2 is a process to predict which class of spatial use a node without class information in the unclassified voxel graph belongs to and is a node class classification problem in GNN.

(1) Dataset

For GNN to learn the voxel graph as a node classification problem, this study created a voxel graph dataset with the configuration shown in Table 1. The number of voxel graphs used for training data is 10,000. The number of voxel graphs used for validation data and test data is both 1,000. A single voxel graph has three types of data: 1) adjacency matrix, 2) node features, and 3) node's ground truth class label.

Table 1. Structure of Voxel Graph Dataset

The number of training data	10,000
The number of validation data	1,000
The number of test data	1,000
Data in one voxel graph	Adjacency matrix
	Node feature (Coordinate or Signal)
	Ground truth class label for each node

For the node features, the following two cases were tested: 1) the X, Y, and Z coordinates were used as node features (Coordinate Feature); 2) some nodes were given the ground truth class label (Signal Feature), which was used as node features. Each of these cases can be interpreted in the voxel graph, as shown in Figure 4.

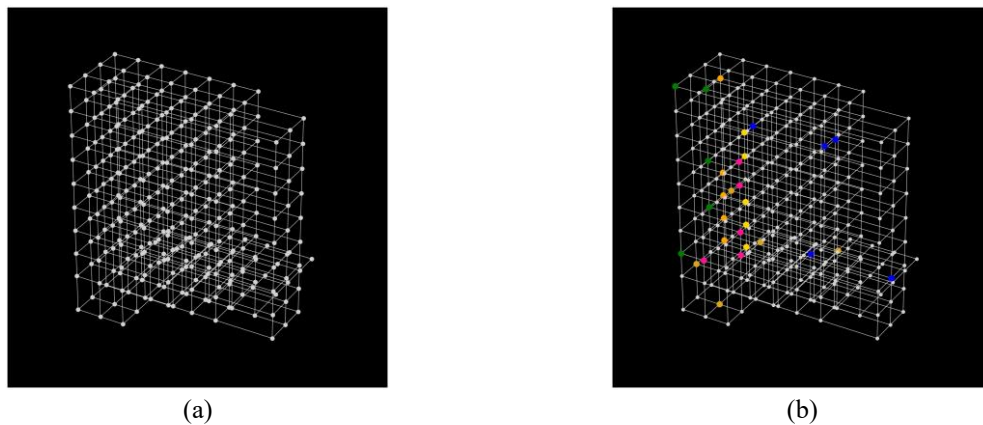


Figure 4. (a) Coordinate Feature.

(b) Signal Feature. 5 nodes for each class are given ground truth labels in this case (5 signals).

When GNN is trained with X, Y, and Z coordinates as node features, the GNN will predict the class of in-building spatial use (e.g., office, facility room, etc.) from the given building volume alone. When some of the nodes are given the ground truth class label, and GNN is trained, the GNN will classify the remaining nodes that were not given the ground truth label. This is equivalent to the designer specifying the spatial use of a part of the building volume and using it as a constraint, and the GNN automatically designs the rest of the in-building spatial use.

(2) Graph Neural Network Architecture

The basic idea of GNNs is to update the feature vector of each node with the feature vectors of its neighboring nodes, thereby enriching the node's features to features that reflect the structure of the graph. Equation (1) is the update mechanism for a typical GNNs.

$$\mathbf{h}_i = \sum_{j \in \mathcal{N}_i} x_j W^T \quad (1)$$

Where x_j is the feature vector for node j . \mathbf{h}_i is the updated vector for node i . W is the weight vector. \mathcal{N}_i is the set of node i and its neighboring nodes.

This study used GCN (Kipf and Welling 2017) to classify nodes. General GNN models have the problem that nodes with large degree (the degree of a node is the number of edges that incident to it) have extremely large influence compared to nodes with small degree due to the difference in the degree of the nodes. As shown in Equation (2), GCN normalizes the imbalance that general GNNs suffer from by dividing the updated features of a node by degrees when updating the features of the node.

$$\mathbf{h}_i = \frac{1}{\deg(i)} \sum_{j \in \mathcal{N}_i} x_j W^T \quad (2)$$

Where $\deg(i)$ is the degree of node i .

The GNN that learns the voxel graph dataset and performs node class classification has the configuration shown in Figure 5: Node features of the voxel graph are embedded, passed through the 4-layer GCN, and connected to all coupling layers to perform classification for each node.

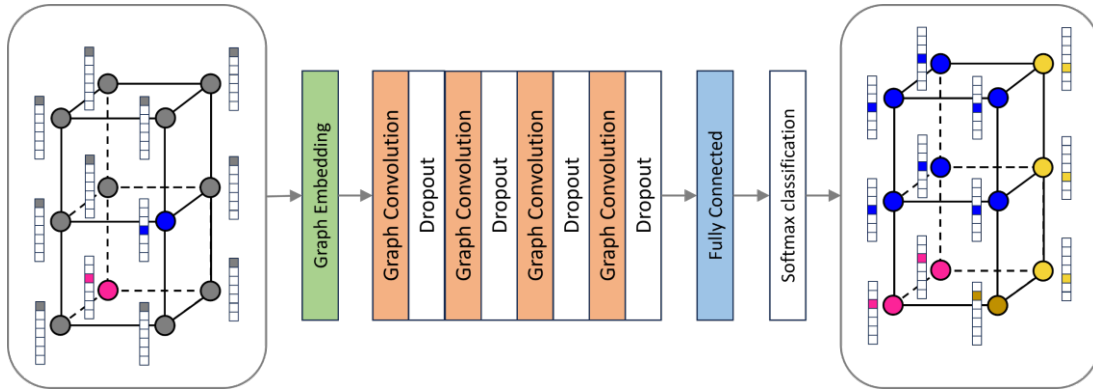


Figure 5. Structure of Graph Neural Network for Voxel Graph Dataset

3.4. Conversion from Classified Voxel Graph to Building Volume with Spatial Usage Class

In Step 3 in Figure 1, the classified voxel graph obtained by Step 2 is converted into a building volume with class information on in-building spatial use. In Step 3, the classified voxel graph can be converted into a building volume by voxelizing each node of the classified voxel graph according to the design baseline, centered at the node's coordinates, in the reverse of Step 1. Finally, the building volume with in-building spatial use is cut horizontally according to the reference lines to output floor plans for each floor.

4. RESULTS

4.1. Implementation Details

The GNN used in this study was implemented with Pytorch and Deep Graph Library (Wang et al. 2019), which provides a Python package necessary for GNN implementation. The Optimizer is AdamW, the initial value of the learning rate is 5×10^{-5} , and if the loss function decreases for 25 epochs, the learning rate is halved, and when the learning rate falls below 1×10^{-6} , the GNN training is completed, and the node classification accuracy at this time is defined as classification accuracy of the GNN implemented in this study.

4.2. Definition of Node Classification Accuracy

The classification of voxel graph node by GNN was performed using 10,000 voxel graphs as training data, with coordinate feature and signal feature. The overall node classification accuracy is the average accuracy for each class. The accuracy for one class is obtained by dividing the number of nodes whose ground truth label was correctly estimated by GNN by the number of nodes belonging to that class. The node classification accuracy for each class is expressed in Equation (3), and the overall node classification accuracy is expressed in Equation (4).

$$\text{Accuracy for Class } k (A_k) = \frac{N_{Pred(k)}}{N_{GT(k)}} \quad (3)$$

$N_{GT(k)}$ is the number of nodes whose ground truth label is class k . $N_{Pred(k)}$ is the number of nodes that is predicted to belong to class k by GNN.

$$\text{Accuracy} = \frac{\sum_{k=1}^{nb_{classes}} A_k}{nb_{classes}} \quad (4)$$

4.3. Node Classification Accuracy

The relationship between node classification accuracy and the epochs on the training data is illustrated in Figure 6. The "5 signals" scenario represents the case where five nodes per class are randomly selected in advance, and their ground-truth classes are provided as node features. Similarly, the "10 signals" scenario corresponds to the case where ten nodes per class are randomly selected, with their ground-truth classes used as node features. The node classification accuracy for both the training and test datasets is summarized in Table 2. For the test dataset, when the node features were XYZ coordinates, the accuracy was 22.8%. In the "5 signals" scenario, the accuracy increased to 56.67%, and in the "10 signals" scenario, it further improved to 74.66%.

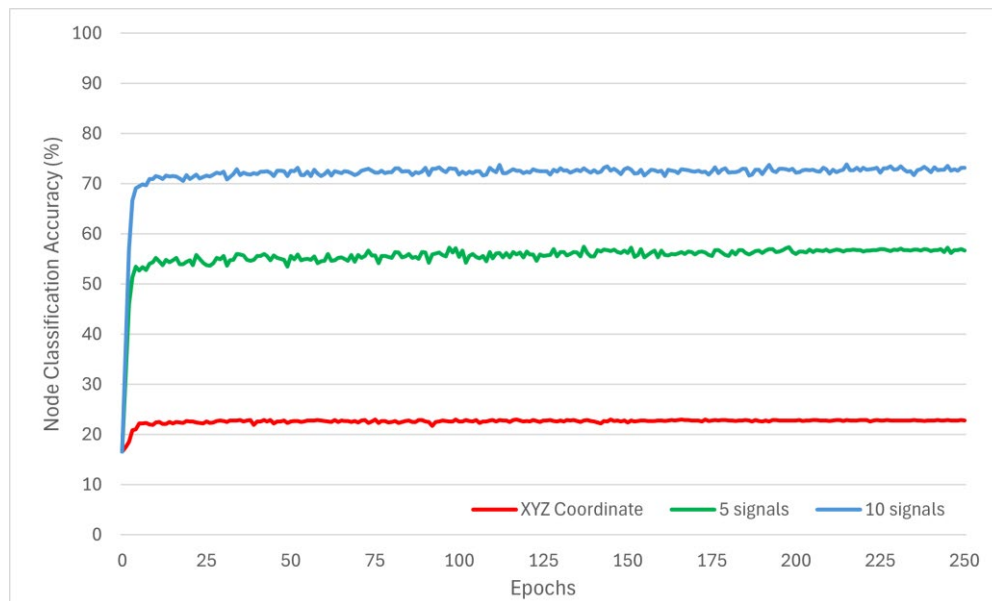


Figure 6. The Relationship between Node Classification Accuracy and the Epochs on the Training Data

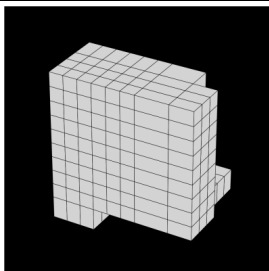
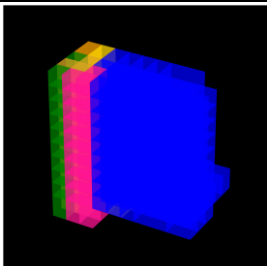
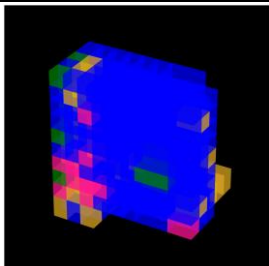
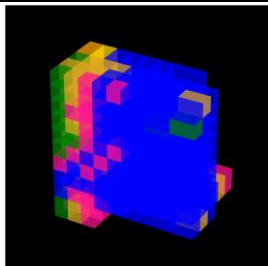
Table 2. Node Classification Accuracy on Test Data and Training Data

	XYZ Coordinate	5 Signals	10 Signals
Test Data	22.81%	56.67%	74.66%
Training Data	22.35%	56.37%	76.01%

The multi-story floor plans are directly obtained from the building volume that is voxelized from the classified voxel graph by Step 3 in Figure 1. Therefore, this GNN node classification accuracy, which indicates the accuracy of Step 2 of generating the classified voxel graph, is an indicator of how well the floorplan generated predicts the ground truth floorplan of the input building volume.

4.4. Output Floorplans

Figure 7 shows the results of floorplan generation by integrating the GNN trained by 5 signals and 10 signals, respectively, into the proposed method. The input building volume in Figure 7 is an example of a building volume randomly extracted from the test data of Voxel Graph Dataset.

	Ground Truth	Generated Floorplans with 5 Signals	Generated Floorplans with 10 Signals
Input Building Volume			<ul style="list-style-type: none"> Lobby/Corridor Restroom Office Stairs Elevator Mechanical
Node Classification Accuracy		56.67 %	74.66 %
Generated Building Volume			

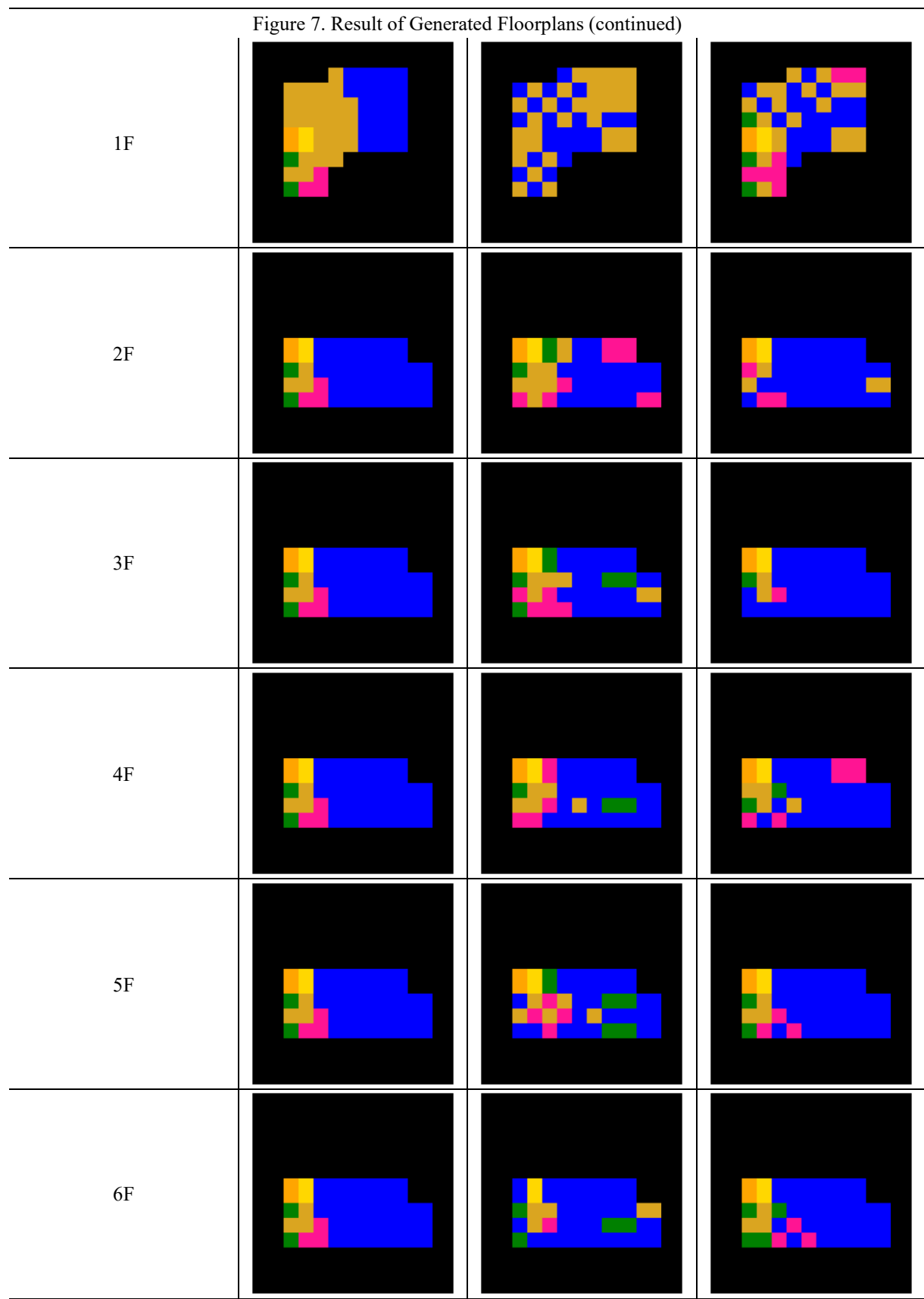


Figure 7. Result of Generated Floorplans (continued)

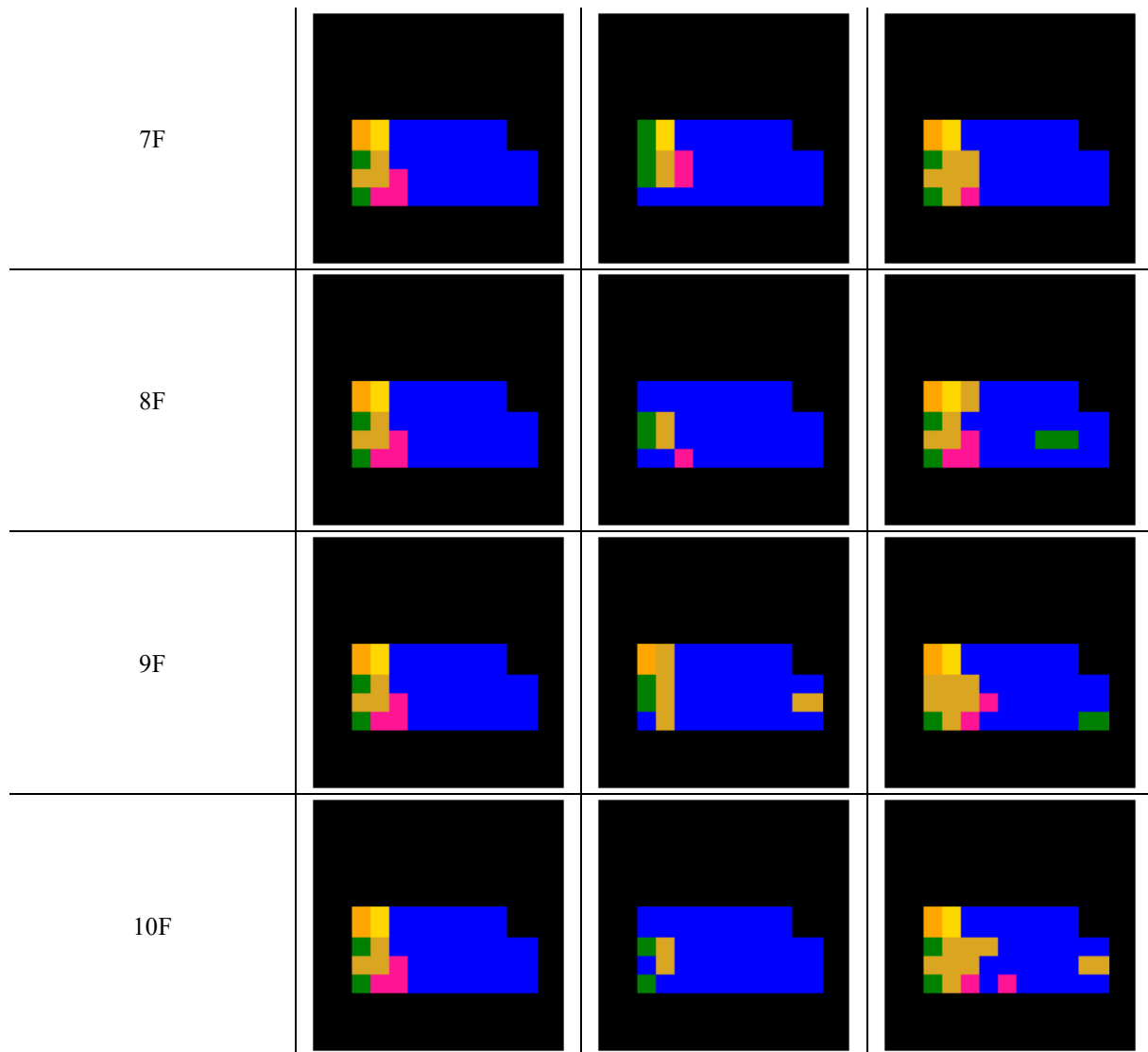


Figure 7. Result of Generated Floorplans

5. DISCUSSION

Comparing the generated floor plans with the Ground Truth floor plans, the generated first floor plans do not replicate the Ground Truth floor plans as well as the floor plans on the second and higher floors. This can be attributed to the fact that the Ground Truth floor plans differ significantly between the first floor and the second floor and above. This suggests that GNN was able to learn the features of similar floor plans on the second and higher floors of Ground Truth but was unable to learn the rapidly changing floor plan on the first floor. In addition, when looking at the floor plans generated when five signal features were given, there is a horizontal series of green areas indicating elevators. This may be improved by adding coordinate features to the five signal features and training the GNN. In addition, the voxel graph dataset used in this study was designed for office buildings, so the generation of floorplans by this study is limited to office buildings. Different datasets should be prepared for the design of hotels, high-rise residential buildings, and commercial facilities.

6. CONCLUSION

This study proposed a novel method to generate multi-story floor plans in a given building volume by using GCN on voxel graph, which represents building volumes with spatial use information as graph data. The GNN implemented in this study predicts the remaining volume's spatial uses from a small number of spatial uses given as initial design conditions. By specifying 5 nodes per class for each spatial use in the voxel graph, the remaining nodes are classified with 56.67% accuracy. If 10 nodes per class for each spatial use are specified within the voxel graph, the remaining nodes are classified with 74.66% accuracy.

The contributions of this study are as follows.

- A novel method was proposed to generate multi-story floor plans within a given building volume.
- It has been demonstrated that a GNN can be implemented to predict the spatial use of the remaining building volume by specifying the spatial use for a part of the building volume.

The proposed method generates a floor plan, serving as the starting point for subsequent interior design within a 3D building model that initially contains no internal details. Future work will aim to integrate this research with 3D generative AI that creates building volumes, potentially enabling the generation of 3D building models with detailed interior designs.

ACKNOWLEDGMENTS

This work was partially supported by JSPS KAKENHI Grant Number 23K11724.

REFERENCES

- Baušys, R., & Pankrašvaitė, I. 2005. "Optimization of architectural layout by the improved genetic algorithm." *Journal of Civil Engineering and Management*, 11(1).
- Chang, K.-H., Cheng, C.-Y., Luo, J., Murata, S., Nourbakhsh, M., & Tsuji, Y. 2022. "Building-GAN: Graph-Conditioned Architectural Volumetric Design Generation." *IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, QC, Canada, 2021, pp. 11936-11945.
- Friedman, Y. 1971. "Flatwriter: choice by computer." *Progressive Architecture* 03 89–101.
- Hu, R., Huang, Z., Tang, Y., van Kaick, O., Zhang, H., & Huang, H. 2020. "Graph2Plan: Learning Floorplan Generation from Layout Graphs." *ACM Transactions on Graphics*, 39(4).
- Jia, Y., Wang, J., Shou, W., Hosseini, M. R., & Bai, Y. 2023. "Graph neural networks for construction applications." *Automation in Construction* (Vol. 154).
- Kipf, T. N., & Welling, M. 2017. "Semi-supervised classification with graph convolutional networks." 5th *International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*.
- Merrell, P., Schkufza, E., & Koltun, V. 2010. "Computer-Generated Residential Building Layouts." *ACM Transactions on Graphics*, 29(6).
- Nauata, N., Hosseini, S., Chang, K. H., Chu, H., Cheng, C. Y., & Furukawa, Y. 2021. "House-GAN++: Generative adversarial layout refinement network towards intelligent computational agent for professional architects." *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.
- Rodrigues, E., Gaspar, A. R., & Gomes, Á. 2014. "Automated approach for design generation and thermal assessment of alternative floor plans." *Energy and Buildings*, 81.
- Wang, A., Dong, J., Lee, L., Shen, J., & Hui, P. 2023. "Towards AI-Architecture Liberty: A Comprehensive Survey on Designing and Collaborating Virtual Architecture by Deep Learning in the Metaverse." *ArXiv*.
- Wang, M., Zheng, D., Ye, Z., Gan, Q., Li, M., Song, X., Zhou, J., Ma, C., Yu, L., Gai, Y., Xiao, T., He, T., Karypis, G., Li, J., & Zhang, Z. (2019). Deep Graph Library: A Graph-Centric, Highly-Performant Package for Graph Neural Networks. *ArXiv*. /abs/1909.01315
- Weber, R. E., Mueller, C., & Reinhart, C. 2022. "Automated floorplan generation in architectural design: A review of methods and applications." *Automation in Construction* (Vol. 140).
- Wu, W., Fu, X. M., Tang, R., Wang, Y., Qi, Y. H., & Liu, L. 2019. "Data-driven interior plan generation for residential buildings." *ACM Transactions on Graphics (TOG)*, 38(6), 1-12.
- Yi, H., & Yi, Y. K. 2014. "Performance based architectural design optimization: Automated 3D space layout using simulated annealing." 2014 *ASHRAE/IBPSA-USA Building Simulation Conference*.
- Zhong, X., Koh, I., & Fricker, P. 2023. "Building-GNN: Exploring a co-design framework for generating controllable 3D building prototypes by graph and recurrent neural networks." *Digital Design Reconsidered: Proceedings of the 41st Conference on Education and Research in Computer Aided Architectural Design in Europe (eCAADe 2023)* (Vol. 2, pp. 431-440).
- Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., & Sun, M. 2020. "Graph neural networks: A review of methods and applications." *AI Open* (Vol. 1).
- Zhou, P., & Chang, Y. 2021. "Automated classification of building structures for urban built environment identification using machine learning." *Journal of Building Engineering*, 43, 103008.

IFC-BASED INDOOR SPACE EXTRACTION WITH TOPOLOGICAL GRAPH MODELING

Zheng Zhao¹, Weiya Chen², Zhiyuan Guo³, Wenming Jiang^{4*}

1) Master's Student, School of Civil and Hydraulic Engineering, Huazhong University of Science and Technology, National Center of Technology Innovation for Digital Construction, International Joint Research Laboratory of Smart Construction, Wuhan, China. Email: m202271570@hust.edu.cn

2) Ph.D., Assoc. Prof., School of Civil and Hydraulic Engineering, Huazhong University of Science and Technology, National Center of Technology Innovation for Digital Construction, International Joint Research Laboratory of Smart Construction, Wuhan, China. Email: weiya_chen@hust.edu.cn

3) Ph.D. Candidate, School of Civil and Hydraulic Engineering, Huazhong University of Science and Technology, National Center of Technology Innovation for Digital Construction, International Joint Research Laboratory of Smart Construction, Wuhan, China. Email: d202180541@hust.edu.cn

4) Ph.D. Candidate, School of Civil and Hydraulic Engineering, Huazhong University of Science and Technology, Wuhan, China. China Academy of Building Research, Beijing, China. Email: jiangwenming@cabrttech.com

Abstract: The semantic enrichment of building information models (BIMs) has been widely explored, with various approaches utilizing graph neural networks (GNNs) to infer the types of indoor spaces. However, there is a gap in the intermediate process that translates the building information model into a graph model suitable for GNNs. To address this problem, we propose a structured graph model designed to represent the attributes and topological relationships of indoor spaces for space type classification. Based on the Industry Foundation Classes (IFC) file format, we define the concept of indoor space and propose an automated method for its extraction. The extracted interior subspaces are analyzed based on their geometric and topological properties, focusing on their relationship to the overall layout of the building's interior spaces. The subspaces are represented as nodes in the graph model, and the edges between the nodes are defined according to the topological relationships between the subspaces. A result was carried out to demonstrate the effect of the proposed indoor space extraction algorithm, which provided the basis for the inference of indoor spatial semantic types in building information models.

Keywords: Building information modeling, IFC, Spatial analysis, Graph theory

1. INTRODUCTION

BIM is a technology and a series of processes used to produce, communicate, and analyze building models, widely used in architecture, engineering, and construction (AEC) (Eastman et al., 2011). The adoption of Building Information Modeling (BIM) in the architecture, engineering, and construction industry has brought many benefits, but it has also introduced interoperability issues among different stakeholders and their BIM platforms (Olofsson et al., 2008). The Industry Foundation Classes (IFC) schema was created by Building Smart to overcome this issue (buildingSMART, 2013). However, when IFC model files are exchanged between different platforms in practice, there are issues of information loss or distortion (Pazlar et al., 2008). In practical modeling, space-related entities and information in BIM models are often overlooked. Therefore, space extraction and type classification are gradually becoming key tasks for the semantic enrichment of BIM models.

A space can be defined in a few different ways, based on its physical separation from other spaces, or based on the users' required function for the space (Ekholm et al., 2000). Thus there is not an only way to define and classify space type. A graph is a data structure composed of nodes and edges, where the nodes are connected by edges, and each node has zero or more associated edges (West, 2001). Graph representation is a traditional approach to designing building layouts. For example, a classic graph-based method used in space syntax is the justified plan graph, which analyzes building layouts by considering the relative size, shape, location, and orientation of rooms (Ostwald & Michael, 2011). With the introduction of graph neural networks (GNNs), a new field of graph structure learning has emerged in recent years, offering novel approaches to identifying indoor space types and enriching BIM semantics. Relational graph convolutional networks (RGCN) and traditional machine learning are used to infer the usage of rooms in public buildings, with the floor plan represented as a graph (Hu et al., 2020). Two methods are proposed and compared for predicting both the topology and the categories of rooms given a partial map. The map is represented as a graph, with rooms as nodes and their connectivity as edges (Aydemir et al., 2012). An improved algorithm based on Graph-SAGE has been proposed, which takes into account the features of edges representing the topological relationships between rooms and has achieved good results in the room type classification task (Wang et al., 2021). A few other papers focus on room type classification using different GNN models and graph representations, with tests conducted on the open-source database House-GAN (Verma & Jadeja, 2023; Paudel et al., 2021; Nauata et al., 2020).

The establishment of indoor spaces can be categorized into two main approaches. The first approach involves manually modeling space entities using different software during the design phase. For example, in

Autodesk Revit, this is typically achieved by creating and defining rooms with the "Room" tool, which automatically assigns spatial boundaries based on walls, floors, and ceilings. The second approach involves automatically extracting indoor spaces from BIM models using geometric, topological, and semantic attributes. This method supplements spatial information that may be missing during the design phase. These methods can be classified based on the different types of model files used. A method to analyze the geometry and semantic information of 2D vector floor plans and automatically reconstruct the corresponding 3D building models has been proposed. This approach involves preprocessing wall lines using an odd-even-based method and employs the loop search method to extract indoor spaces (Zhu et al., 2014). Several studies have focused on indoor space reconstruction and subspace partitioning based on point cloud data (Tang et al., 2024; Yue et al., 2024; Patrick et al., 2021). The ArcGIS platform is chosen to integrate BIM data, and a method for extracting indoor space information in complex buildings is proposed. The connectivity between upper and lower floor spaces is determined by recognizing the polylines of floor slab hole symbols, and the indoor multi-floor spaces are extracted accordingly (Pang et al., 2018). However, there is a lack of discussion on indoor space extraction methods based on IFC files. Our approach builds upon and improves the method proposed by Pang et al. (2018), enhancing its effectiveness in extracting indoor spaces by incorporating IFC-based geometric data and enriching the semantic information of the spaces. Similarly, the space extraction task in the field of building energy modeling is also discussed and can be broadly categorized into the following types: graph- and rule-based algorithms, geometry computing-based algorithms, domain knowledge-based algorithms, hybrid algorithms, and 2D to 3D algorithms. The building topology graph (BTG) is introduced as a graph data structure where nodes represent building elements and edges connect adjacent building elements. A space recognition algorithm based on this graph structure and local features is proposed to support performance-oriented design in the early design stage (Chen et al., 2017).

Therefore, it is worth investigating methods for indoor space analysis in BIM models using IFC as input for semantic enrichment. Similarly, the application of GNNs for identifying indoor space functions relies on pre-existing graph models. It is worth studying how to automatically extract information from building information models to form graph models that can support GNN algorithms for semantic enrichment.

In this paper, we define indoor spaces and present an algorithm for their automatic extraction from IFC files. This algorithm involves analyzing and processing the geometric representations of BIM components within the IFC schema. We then introduce a graph model designed to integrate geometric and topological features of the extracted indoor spaces. The model serves as a bridge to transform the complex geometric information within IFC-based BIM models into a spatial graph structure that can effectively be processed by GNNs.

2. METHOD

2.1 Descriptions of Indoor Spaces

Building components are the basic elements that constitute a building, either individually or in combination with each other to fulfill the structural and functional requirements of the building. Common examples of building components include columns, slabs, walls, stairs, and windows. A space is defined as a portion of the building enclosed by the interconnection of these components. The elements that contribute to this enclosure are termed boundary components. These components define the indoor space by separating the interior of the building from its external environment. Moreover, boundary components subdivide the indoor space into smaller, functional subspaces, such as rooms, corridors, or stairwells, to satisfy diverse requirements within the building. After specifying the definition of the indoor space, it becomes essential to predefine the boundary components utilized in the space extraction algorithm. Table 1 details the semantic descriptions of all boundary component types, along with their respective entity names as specified in the IFC file.

Table 1. Semantic descriptions of boundary components

Component Name	Description	IFC Entity Name
Wall	A primary vertical component that serves structural, enclosing, and partitioning functions. It is a key element in forming the boundaries of indoor spaces.	IfcWall
Curtain Wall	A non-load-bearing facade element, typically used on building exteriors to form the boundary between indoor spaces and the external environment.	IfcCurtainWall
Column	A primary vertical component that typically does not directly serve as a space partitioning element but influences space segmentation.	IfcColumn
Slab	A primary horizontal component that bears vertical loads. It defines floors in vertical space and acts as the upper or lower boundary of a room.	IfcSlab

processing steps.

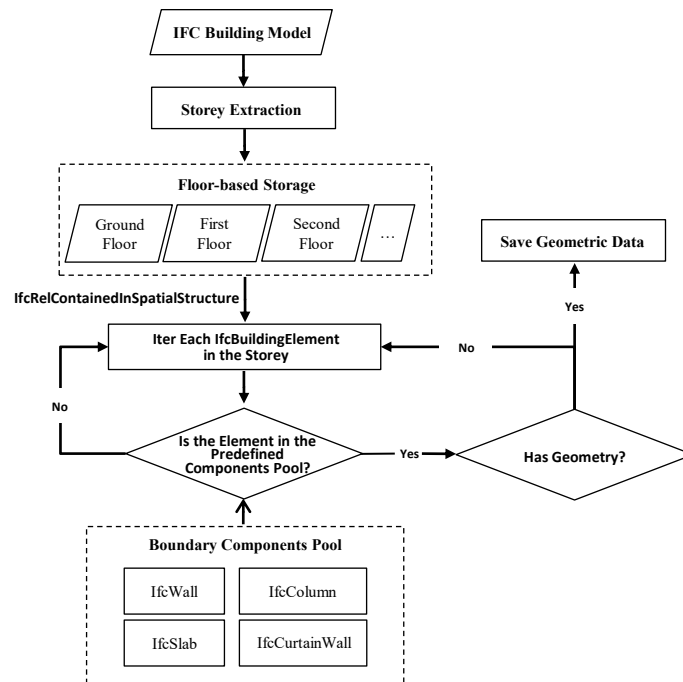


Figure 2. The workflow for extracting geometric data from IFC files.

Simplifying the model to a two-dimensional plane can effectively improve computational efficiency. Since the upper and lower boundaries of the indoor space are defined by the ceiling and the floor, a customized horizontal cutting plane is introduced between them. After obtaining the geometric information of all boundary components of the storey, these components are projected onto the cutting plane. At this stage, each boundary component B_i in the set $B = \{B_1, B_2, \dots, B_n\}$ corresponds to a projected polygon P_i . Subsequently, all the polygons P_i are merged through a union operation to form a unified geometric object P , as shown in Figure 3(a). The convex hull C of geometric shape P is computed, as shown in Figure 3(b). A Boolean difference operation is performed between the convex hull C and the shape P , where C is subtracted from P , resulting in the uncovered region set S , which represents the area not covered by the boundary components, as depicted in Figure 3(c). The set S consists of several regions represented by polygons, which can be classified into two categories based on their interaction with the boundary of the convex hull. The first category includes spaces that are separated from the outdoor environment by the outermost continuous building elements, such as walls and a few columns. These spaces are further subdivided by internal building components, resulting in distinct indoor subspaces. These subspaces represent the target spaces we need to extract, as shown by the blue areas in Figure 3(d). The other category consists of spaces located between the outermost components and the convex hull boundary, which are discarded as non-target outdoor spaces, indicated by the red areas in Figure 3(d).

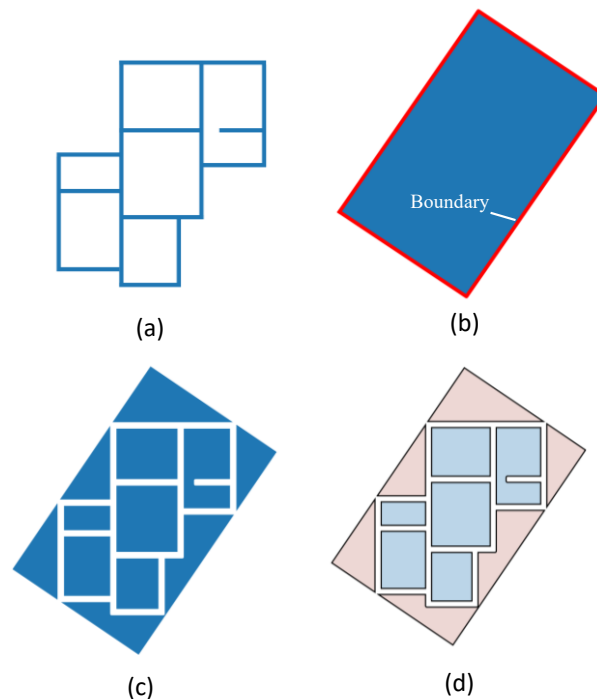


Figure 3. Visualization of the indoor space extraction process. (a) Projection of boundary components (b) Convex hull and its boundary (c) Boolean difference operation between the convex hull and boundary components (d) Classification of spaces into target indoor subspaces (blue areas) and outdoor spaces (pink areas).

2.3 The ISTAG Model

To represent the characteristics of indoor spaces within a building information model, this paper constructs the indoor space topology and attribute graph (ISTAG) model to describe the attributes and relationships of the spaces. This model is designed to support subsequent graph-based spatial analysis applications, enabling more efficient space utilization and functional analysis within the built environment.

$$G = (V, E) \quad (2)$$

The composition of the graph is represented as shown in Equation (2), where $V = \{v_1, v_2, \dots, v_n\}$ represents a finite set of nodes, with n denoting the number of nodes in the set V , each of which corresponds to an indoor subspace. Each subspace node has two types of attributes: geometric properties and topological information of the space. $E = \{e_1, e_2, \dots, e_m\}$ is the set of undirected edges, representing the connection relationships between two indoor subspaces. An edge $e = (v_i, v_j)$, where $1 \leq i, j \leq n$ and $i \neq j$, indicates a connection between nodes v_i and v_j . The edge attributes include connectivity and adjacency relations.

(1) Geometric Attribute of Indoor Spaces

In this section, the geometric features of indoor spaces in building information models are extracted and processed to construct the node feature information of ISTAG models. These geometric attributes include both basic geometric information (such as aspect ratio) and processed measurement values, which are used to more effectively describe spatial characteristics. A detailed description of each geometric attribute is given below:

Proportion of room area relative to total indoor floor area. Area is one of the most fundamental geometric attributes of a room. The function of a room is closely related to its size, and the ratio of a room's area to the total indoor floor area serves as a key distinguishing feature for different room types. For example, smaller spaces are typically used for storage or utility purposes, whereas larger spaces are more likely to serve as living rooms.

Aspect Ratio. The aspect ratio, defined as the ratio of a room's length to its width, is a key indicator of a room shape. Spaces with distinct functions tend to exhibit varying aspect ratios. For example, corridors, which serve as connecting spaces between rooms, typically have larger aspect ratios, whereas rooms with an aspect ratio close to 1:1 are more commonly associated with living spaces such as kitchens or living rooms.

Compactness. Compactness is measured as the ratio of a room's area to its perimeter, indicating the degree of geometric regularity. Spaces with high compactness often exhibit regular geometric shapes, whereas spaces with low compactness tend to have irregular or elongated forms.

Number of Boundary Lines. The number of boundary lines represents the count of edges that define a space's boundaries. Rooms with fewer boundary lines generally correspond to simpler geometric shapes, such as squares or rectangles, which are commonly found in primary functional spaces like living rooms, bedrooms, or offices. In contrast, spaces with more boundary lines typically correspond to complex or irregular shapes, or areas designed with specific functions. This attribute serves as a valuable reference for tasks such as space classification and spatial region identification.

Distance Between Subspace Center and Floor Center. This value measures the distance between the geometric centroid of a room and the center of the minimum oriented bounding box (OBB) that encompasses the entire floor space. It reflects the spatial positioning of the room within the floor layout. Rooms situated in the core areas of a building are often functionally more significant compared to those located at the edges or corners.

(2) Topological Attribute of Indoor Spaces

This paper introduces two types of topological relationships for indoor spaces: connectivity and adjacency. To support the establishment of edges in the ISTAG model, we propose a method for constructing indoor topology based on the IFC building information model. By extracting topological information between spaces, the logical structure of the building layout can be more effectively characterized, laying a solid foundation for advanced spatial analysis and functional inference.

Connectivity refers to whether different enclosed spaces can connect through openings (such as doors or windows), thereby enabling the movement of people or objects between spaces. In this study, the determination of spatial connectivity is primarily based on doors as critical components. Specifically, two enclosed spaces are considered connected if they are directly linked by at least one door. To evaluate the connectivity between spaces, the geometric representations of all doors (IfcDoor) on a given floor are extracted during the parsing of the IFC file. These geometries, along with other boundary components, are projected onto a two-dimensional plane. Figure 4(a) illustrates the results of an indoor space extraction algorithm, showing several subspaces $\{S_1 \sim S_6\}$.

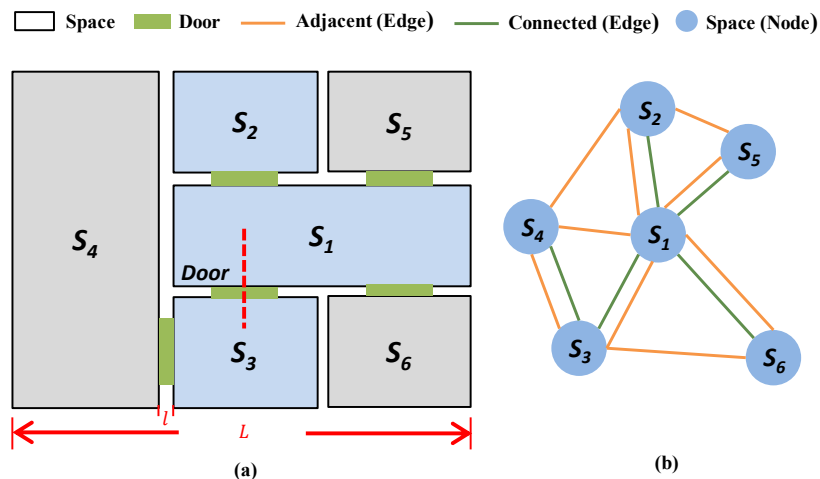


Figure 4. Topological analysis of indoor spaces for the ISTAG model.

Each door (represented as green polygons) naturally connects two indoor subspaces. If a door is associated with only one subspace, it is considered as leading to the exterior. To ensure precise identification of connected spaces, a line segment perpendicular to each door element is drawn, as illustrated by the red dashed lines in Figure 4(a). In this study, the length of this segment is set equal to the door's length. This fixed length ensures that a single door does not erroneously associate with multiple spaces. For example, as shown in the blue-highlighted area in the Figure 4(a), without this constraint, the door element could incorrectly connect both Space S_1 and Space S_2 . Connectivity emphasizes the accessibility between spaces, whereas adjacency focuses more on the proximity of two spaces within the layout. If two spaces share one or more boundary components (typically walls), they are considered adjacent spaces. Due to the presence of gaps between indoor subspaces, resulting from the Boolean difference operation applied to wall components during the indoor space extraction process, adjacency cannot be determined by whether spaces share a boundary. Therefore, we define a threshold to determine if two spaces are "in contact." For example, in Figure 4(a), if the distance l between two spaces is less than 5% of the floor plan length L , the two spaces are considered adjacent. In this case, an edge is created between the corresponding nodes in the ISTAG model to represent the adjacency relationship.

In the final ISTAG model, as shown in Figure 4(b), blue nodes represent indoor subspaces $\{S_1 \sim S_6\}$. The adjacency and connectivity relationships are differentiated by different edge types. Adjacency relationships,

represented by orange edges, while connectivity relationships, represented by green edges.

3. RESULTS

In this section, the visualization results are presented to validate the indoor space extraction algorithm based on the IFC model and the ISTAG model representation. First, individual floors were extracted from the building information model, and the results are shown in the left panel of Figure 5. The yellow components are extracted as boundary components enclosing the indoor space, so the red lines in the right picture represent the indoor space. The blue part corresponds to the outdoor space, whose boundary is defined by the outer railing (IfcRailing), which is not a predefined boundary component.

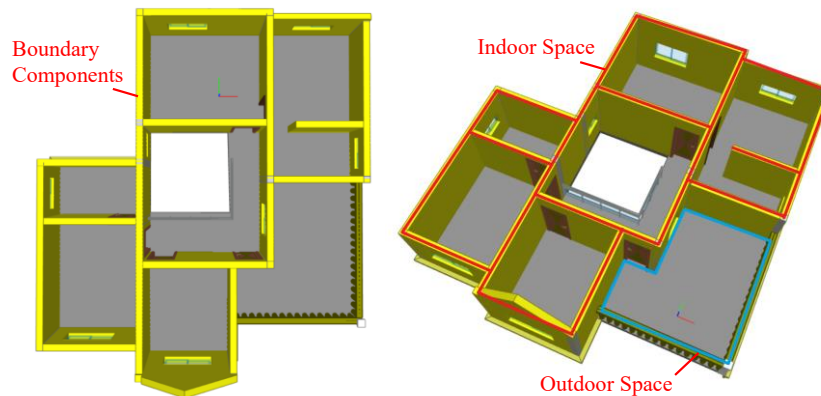


Figure 5. Boundary components extracted from IFC files

After the operation of the indoor space extraction algorithm, the indoor space is successfully extracted as shown in Figure 6. The geometric attributes are then computed and extracted for each indoor subspace, and the output is shown on the right. We take each indoor subspace as a node in the ISTAG model and the geometric attributes of the space as node features. It is worth noting that when a door is connected to the only indoor space, this door is considered to be connected to the outdoor space. Then this door is not considered when performing the topological information extraction of the indoor space.

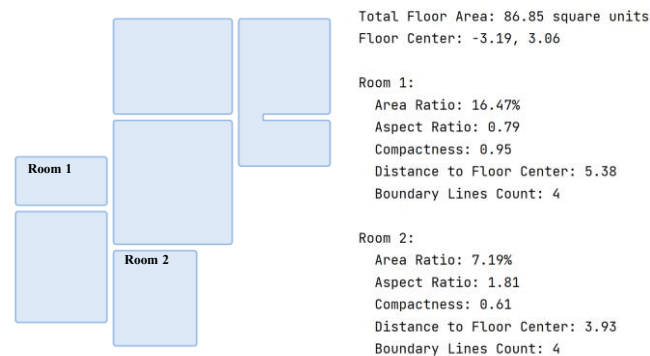


Figure 6. Geometric attributes of indoor spaces analysis results.

The final results of the indoor space connectivity and adjacency visualization built according to the method are shown in Figure 7.

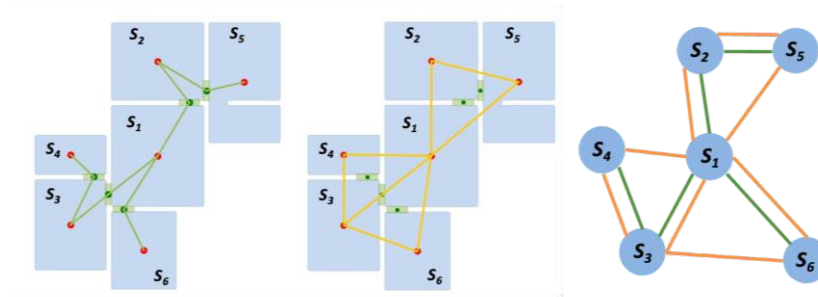


Figure 7. Visualization of ISTAG model construction. The green line indicates the connectivity established between spaces through doors. The orange line is used to indicate the adjacency of two spaces. On the right side is the ISTAG model established by the method

4. DISCUSSION

In this paper, an indoor space extraction method is designed for models in IFC file format to supplement the lack of spatial semantic information. On this basis, the interior subspace is analyzed to construct the ISTAG model. The research gap from raw BIM model files to the input of the GNNs algorithm is effectively filled. However, the current study still has some limitations, such as the inability to automatically identify virtual boundaries, and the degree of spatial connectivity or adjacency is not discussed yet. These issues are our next step work.

5. CONCLUSIONS

In conclusion, this paper proposed an indoor space extraction algorithm based on the IFC files, focusing on accurately processing BIM models using boundary components like walls, slabs, and columns. The algorithm effectively extracts indoor spaces from building models. The ISTAG model plays a crucial role in bridging the gap between BIM data and advanced spatial analysis using GNNs. The current spatial extraction algorithm is not able to deal with the virtual boundaries of indoor spaces, and the degree of spatial connectivity is not quantified in the ISTAG model. These aspects present valuable opportunities for future research.

ACKNOWLEDGMENTS

This study was supported by the National Key Research and Development Program of China (No. 2022YFC3802100) and the National Natural Science Foundation of China (Grant No. 72001086).

REFERENCES

- Aydemir, A., Jensfelt, P., & Folkesson, J. (2012). What can we learn from 38,000 rooms? Reasoning about unexplored space in indoor environments. *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 4675–4682. <https://doi.org/10.1109/IROS.2012.6386110>
- buildingSMART. (2013). *IFC4 Release Candidate 4*. Retrieved from buildingSMART website: <http://www.buildingsmart-tech.org/ifc/IFC2x4/rc4/html/index.htm>
- Chen, Hongzhong & Li, Ziwei & Wang, Xiran & Lin, Borong. (2017). A graph- and feature-based building space recognition algorithm for performance simulation in the early design stage, *Building Simulation*, 11. 10.1007/s12273-017-0412-x.
- Eastman, C., Teicholz, P., Sacks, R., and Liston, K. (2011). *BIM Handbook: A Guide to Building Information Modeling for Owners, Managers, Designers, Engineers, and Contractors* (2nd ed.). John Wiley & Sons.
- Ekholm, Anders & Fridqvist, A. (2000). Concept of space for building classification, product modelling, and design, *Automation in Construction*, 9 (3) (2000) pp. 315-328. 9. 10.1016/S0926-5805(99)00013-8.
- Hu, Xuke & Fan, Hongchao & Noskov, Alexey & Wang, Zhiyong & Zipf, Alexander & Gu, Fuqiang & Shang, Jianga. (2020). Room semantics inference using random forest and relational graph convolutional networks: A case study of research building, *Transactions in GIS*, 25. 10.1111/tgis.12664.
- Hübner, P., Weinmann, M., Wursthorn, S., & Hinz, S. (2021). Automatic voxel-based 3D indoor reconstruction and room partitioning from triangle meshes. *ISPRS Journal of Photogrammetry and Remote Sensing*, 181, 254–278.
- Nauata, Nelson & Chang, Kai-Hung & Cheng, Chin-Yi & Mori, Greg & Furukawa, Yasutaka. (2020). *House-GAN: Relational Generative Adversarial Networks for Graph-Constrained House Layout Generation*, 10.1007/978-3-030-58452-8_10.
- Olofsson, Thomas & Lee, Ghang & Eastman, Charles. (2008). *Editorial - Case studies of bim in use*, 13.

- 244-245.
- Ostwald, Michael. (2011). The Mathematics of Spatial Configuration: Revisiting, Revising and Critiquing Justified Plan Graph Theory, *Nexus Network Journal*, 13. 445-470. 10.1007/s00004-011-0075-3.
- Pang, Yueyong & Zhang, Chi & Zhou, Liangcheng & Lin, Bingxian & Lü, Guonian. (2018). Extracting Indoor Space Information in Complex Building Environments, *ISPRS International Journal of Geo-Information*, 7. 321. 10.3390/ijgi7080321.
- Paudel, Abhishek & Dhakal, Roshan & Bhattarai, Sakshat. (2021). *Room Classification on Floor Plan Graphs using Graph Neural Networks*, 10.48550/arXiv.2108.05947.
- Pazlar, Tomaž & Turk, Žiga. (2008). Interoperability in practice: Geometric data exchange using the IFC standard, *Electronic Journal of Information Technology in Construction*, 13. 362-380.
- Tang, S., Huang, J., Cai, B., Du, H., Zhou, B., Zhao, Z., Li, Y., Wang, W., & Guo, R. (2024). Back to geometry: Efficient indoor space segmentation from point clouds by 2D–3D geometry constrains. *International Journal of Applied Earth Observation and Geoinformation*, 135, 104265.
- Verma, Atul Kumar & Jadeja, Mahipal. (2023). CB-SAGE: A novel centrality based graph neural network for floor plan classification. *Engineering Applications of Artificial Intelligence*, 126. 107121. 10.1016/j.engappai.2023.107121.
- Wang, Zijian & Sacks, Rafael & Yeung, Timson. (2021). Exploring Graph Neural Networks for Semantic Enrichment: Room Type Classification, *Automation in Construction*, 134. 10.1016/j.autcon.2021.104039.
- West, D.B. (2001) Introduction to Graph Theory. 2nd Edition, Prentice-Hall, Inc., *Upper Saddle River*, 82-83.
- Yue, H., Wu, H., Lehtola, V., Wei, J., & Liu, C. (2024). Indoor functional subspace division from point clouds based on graph neural network. *International Journal of Applied Earth Observation and Geoinformation*, 127, 103656.
- Zhu, J., Zhang, H., & Wen, Y. (2014). A New Reconstruction Method for 3D Buildings from 2D Vector Floor Plan. *Computer-Aided Design and Applications*, 11(6), 704–714.

THE ROLE OF DIGITAL ECONOMY IN THE HIGH-QUALITY DEVELOPMENT OF THE CONSTRUCTION INDUSTRY: EVIDENCE FROM CHINA

Aoqi Huang¹, Hongjuan Wu², Taozhi Zhuang³

1) Master Candidate, School of Management Science and Real Estate, Chongqing University, Chongqing 400044, China, E-mail: aoqihuang@stu.cqu.edu.cn

2) Ph.D., Assoc. Prof., School of Management Science and Real Estate, Chongqing University, Chongqing 400044, China, E-mail: hongjuanwu@cqu.edu.cn

3) Ph.D., Assoc. Prof., School of Management Science and Real Estate, Chongqing University, Chongqing 400044, China, E-mail: tz.zhuang@cqu.edu.cn

Abstract: Taking into account the challenges posed by complex business management processes and low levels of digitalization, the emergence of the digital economy presents significant opportunities for fostering high-quality development in the construction industry. This study develops a theoretical framework to examine how the digital economy influences the high-quality development of the construction industry. Using panel data from 30 provinces in China spanning from 2012 to 2021, the projection pursuit model and spatial Durbin model were applied to examine the impact of the digital economy on the high-quality development of the construction industry. The findings reveal a general decline in both the digital economy and the high-quality development of the construction industry from the southeastern coastal provinces to the northwestern inland regions. The digital economy directly promotes the high-quality development of the construction industry and exhibits spatial spillover effects. Further analysis indicates that the impact of the digital economy varies by region. Moreover, government transparency and marketization processes are effective mechanisms through which the digital economy fosters high-quality development in the construction industry. This study provides valuable insights for integrating the digital economy into the construction on a broader scale.

Keywords: High quality development of the construction, Digital economy, Spatial spillover, Mediating effect

1. INTRODUCTION

Amid the ongoing advancements in technological innovation and industrial transformation, China's economy has shifted from a phase of rapid growth to one of high-quality development (Wen et al., 2024; Wang et al., 2023). As a pillar of the national economy, the construction industry has made significant contributions to job creation and infrastructure improvement (Golpîra, 2020). However, the industry also faces numerous obstacles, including slowed growth, low labor productivity, and challenges in information sharing. According to the McKinsey Global Institute's report *Imagining Construction's Digital Future*, the construction industry ranks just above agriculture in global digitalization levels. In light of these challenges, the construction industry urgently needs to change the extensive development model and shift toward a high-quality development path (Ding et al., 2024; Wang et al., 2021). Consequently, finding innovative solutions to promote high-quality development of the construction (HDCI) has become a critical priority.

As a new economic form encompassing the internet, big data, and cloud computing, the digital economy (DIG) has provided new momentum for the transformation of traditional industries (Rodrigues & Franco, 2021; Zhou et al., 2024). The China Digital Research and Development Report (2024) reveals that, by 2023, the scale of China's digital economy surpassed 55 trillion yuan, accounting for over 40% of the country's GDP. Unlike the traditional economy, data serves as the key driving force in the digital economy, which can overcome the limitations of space and time (Cong et al., 2021). Construction processes involve vast amounts of data, and a comprehensive data analysis system can enhance productivity (Trevisan et al., 2023). More importantly, the inherent characteristics of the digital economy, such as green sharing and lower transaction costs (Nambisan et al., 2017), aligns with the principles of high-quality development.

In recent years, policies related to empowering the high-quality development of the construction industry through the digital economy have often been included in key national plans of the Chinese government. Constructing a comprehensive and scientific evaluation system for measuring DIG and HDCI is essential. Analyzing the relationships between DIG and HDCI helps scientifically describe the driving pathways and improvement strategies. Research increasingly focuses on evaluating HDCI and the impact of DIG on it.

Firstly, since the concept of "high-quality development" was introduced, scholars have conducted extensive research on measuring the level of HDCI. At present, there are two main methods. First method adopts a single indicator to measure construction industry's high-quality development, such as green TFP (Yang et al., 2024). Another approach is to establish an evaluation indicator system for comprehensive measurement (Wang & Cheng, 2024). Besides, researchers have employed methods such as the entropy-weight TOPSIS method (Wang et al., 2019) and the matter-element extension model (Gao & Li, 2021) to evaluate the level of HDCI. Research on HDCI has been continuously expanding. However, there remains a lack of a clear measurement indicator system and a theoretical framework for the driving factors behind HDCI.

Secondly, the relationship between the digital economy and high-quality development has gained increasing attention in academia. Research in this area spans multiple industries, including the economy (Ding et al., 2021), manufacturing (Li et al., 2024), agriculture (Feng & Wang, 2024), and energy (Wang et al., 2022). These studies often use panel data to examine the impacts of DIG on high-quality development. Most use methods like fixed-effects regression and spatial Durbin models (Hong et al., 2022; Liu et al., 2023; Lei et al., 2024). Current studies have also identified factors that may affect the relationship between DIG and high-quality development, such as government policies (Wan et al., 2023) and market (Lyu et al., 2024). Although the impact of the digital economy on sectors like manufacturing and agriculture has been widely explored, academic research on its effects within the construction industry remains relatively limited.

In summary, previous research has failed to provide a unified framework to answer the question of how the digital economy promotes the high-quality development of the construction industry. To bridge this gap, this study utilizes panel data from 2012 to 2021 and applies the spatial Durbin model, verifying whether spatial spillover effects are present. Moreover, the influencing mechanisms of DIG on HDICI are analyzed by considering mediating variables such as government transparency and marketization.

This study makes three main contributions. First, from the perspective of research content, this study examines the direct impacts and spatial effects of the deep integration of the digital economy and the construction industry within the strategic context of high-quality growth, addressing existing research gaps. Second, in terms of research pathways, the paper employs a multidimensional analytical framework that incorporates government transparency and market development level as mediating variables. This framework offers detailed theoretical support for industry practices and policy formulation. Lastly, this paper examines the heterogeneous role of the digital economy in the process of high-quality development in the construction industry. It offers targeted policy recommendations to promote balanced progress in the construction industry.

The remainder of this study is organized as follows. Section 2 presents the theoretical framework and hypotheses. Section 3 details the data, variables, and methods. Section 4 presents the empirical results and discusses the findings. Finally, Section 5 summarizes the study conclusions, policy recommendations, and limitations.

2. THEORETICAL FRAMEWORK AND HYPOTHESES

2.1 Impact of the Digital Economy on the High-Quality Development of the Construction Industry

The dual drivers of data elements and digital technologies represent the core mechanism through which the digital economy fosters the development of traditional sectors, including the construction industry (Ren et al., 2023). First, as a major catalyst for technological innovation, DIG has transformed the traditional production model of the construction industry, promoting the overall standardization and modernization of the industry (Ammar et al., 2021; Alizadehsalehi et al., 2020). Digital tools like Building Information Modeling (BIM) and Augmented Reality (AR) have increased the efficiency of project management and enhanced the precision of construction processes (Khalesi et al., 2020; Schiavi et al., 2022). Furthermore, institutional constraints and information asymmetry often result in high transaction costs in construction projects. Digital platforms improve the efficiency of information exchange and resource matching, thereby reducing transaction costs, overcoming temporal and spatial barriers, and enhancing overall transaction efficiency (Bartelsman et al., 2013; Li et al., 2023). Based on this, we propose:

Hypothesis 1: DIG directly drives the HDICI.

2.2 Influence mechanism

To achieve high-quality development, 'a capable government' and 'an effective market' are indispensable (Wei, 2021). Based on this, this research considers that DIG drives HDICI through the efforts of two participants: government and market.

(1) Mediating Effect of Government Transparency

In a highly competitive environment, government transparency is essential for improving the effectiveness and accuracy of information, thereby reducing transaction costs (Li et al., 2017). According to information asymmetry theory, discrepancies in information between trading parties can lead to decision-making errors and efficiency losses. Improvements in government transparency unleashes the potential of the digital economy (Shahbaz et al., 2022). On the one hand, the integration of digital technology with government governance helps build a multi-sector collaborative digital governance system, promoting government transparency (Wang & Guo, 2024). The public availability of government information optimizes the regulatory and evaluation mechanisms in the construction industry. Government transparency improves the regulatory and assessment mechanisms of the construction industry and promotes its digital transformation through clear policy directions (Babalola & Harinarain, 2021; Del Río Castro et al., 2021). On the other hand, A transparent government environment enables firms to quickly adjust their strategies in response to policy changes, thereby enhancing their competitiveness (Wang et al., 2022; Djibo et al., 2022). Based on this, we propose:

Hypothesis 2: DIG enhances HDICI by increasing government transparency.

(2) Mediating Effect of Marketization Process

The digital economy which relies on the inherent openness of market economies. The industrial organization model of the digital economy with the network platform as the core has changed the market structure and reshaped the traditional market pattern (Liang et al., 2024). The degree of marketization enhances the efficiency of resource allocation, ultimately driving high-quality development in industries (Jiang et al., 2024). The use of data and digital technologies enables businesses to dynamically allocate tasks according to market demands, reducing production costs and improving production efficiency (Mubarak et al., 2021). Moreover, as a key element of the digital economy, data can overcome space and time barriers (Cheng et al., 2023). This stimulates market vitality and creativity, facilitating the free flow of production factors (such as capital and labor) across regions and industries (Xue et al., 2022). Based on this, we propose:

Hypothesis 3: DIG enhances HDI by promoting the marketization process.

Based on existing theoretical findings, we propose a theoretical framework as illustrated in Figure 1.

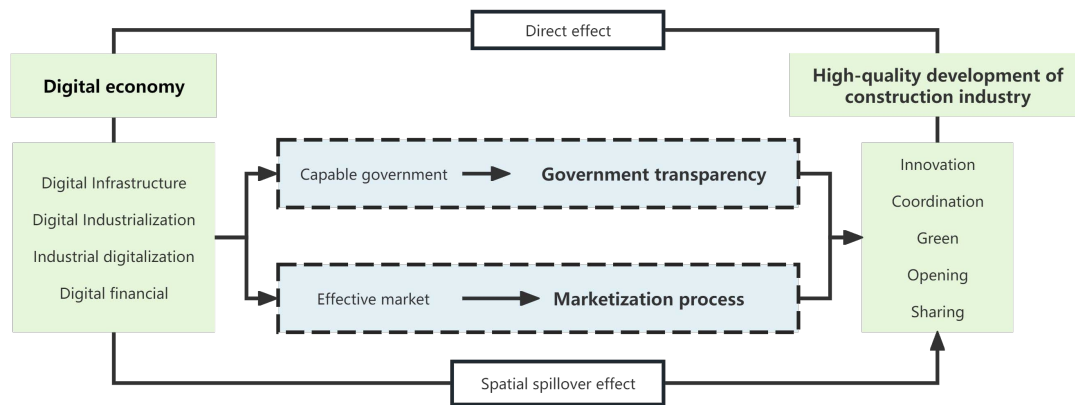


Figure 1 Theoretical mechanism

3. RESEARCH DESIGN

3.1 Models

(1) Projection pursuit model

The projection pursuit model has a strong ability to reduce the dimensionality of multivariate data. Considering the nonlinear and non-normal multidimensional characteristics of the data in the study of HDI and DIG, this paper employs the Projection Pursuit Model to assign weights to each indicator. The specific steps are outlined as follows:

Step 1: Standardize the indicators to eliminate the effects of different scales or units.

$$x'_{ij} = \begin{cases} \frac{x_{ij} - \min(x_{ij})}{\max(x_{ij}) - \min(x_{ij})}, & x_{ij} \text{ is Positive indicators} \\ \frac{\max(x_{ij}) - x_{ij}}{\max(x_{ij}) - \min(x_{ij})}, & x_{ij} \text{ is Negative indicators} \end{cases} \quad (1)$$

Specifically, where x_{ij} represents the j -th indicator of the i -th province.

Step 2: Construct the projection index function

The projection pursuit model combines p -dimensional data into a one-dimensional projection value using the projection direction $a = \{a(1), a(2), \dots, a(p)\}$:

$$Z(i) = \sum_{j=1}^p a(j)x'_{ij}, i = (1, 2, \dots, n) \quad (2)$$

Specifically, where n represents the number of provinces

Step 3: Optimize the Projection Index Function

$$\text{Max } Q(a) = S_Z \times D_Z \quad (3)$$

$$\text{s. t. } \sum_{j=1}^p a^2(j) = 1 \quad (4)$$

In this context, S_Z represents the standard deviation of the projection values. D_Z denotes the local density of the projection values. Equations (3) and (4) represent a complex nonlinear optimization problem. This paper employs accelerating genetic algorithm to solve it.

(2) Spatial Durbin Model

The construction industry is influenced not only by local economic and policy factors but also by external factors from neighboring regions (Li et al., 2024; Zhou et al., 2024). Spatial econometric models are essential for a comprehensive analysis of the impact of DIG on HDCl. The model is constructed as follows:

$$\text{HDCl}_{it} = \alpha_0 + \rho \sum_{j=1}^n w_{ij} \text{HDCl}_{it} + \alpha_1 \text{DIG}_{it} + \alpha_2 Z_{it} + \beta_1 \sum_{j=1}^n w_{ij} \text{DIG}_{it} + \beta_2 \sum_{j=1}^n w_{ij} Z_{it} + \mu_i + \varphi_t + \varepsilon_{it} \quad (5)$$

$$\varepsilon_{it} = \tau \sum_{j=1}^n w_{ij} \delta_{it} + \sigma_{it} \quad (6)$$

where i represents provinces, t represents years, Z_{it} represents the control variable. μ_i and φ_t are individual and time effects, respectively. α and β are the coefficients of the corresponding variables. w_{ij} represents the weight matrix. ε_{it} represents the random disturbance term.

(3) Mediating Effect Model

To test Hypotheses 2-3, we constructed a mediation effect model based on the testing approach proposed, as indicated in the following equation:

$$\text{HDCl}_{it} = a_0 + a_1 \text{DIG}_{it} + a_2 Z_{it} + \varepsilon_{it} \quad (7)$$

$$M_{it} = b_0 + b_1 \text{DIG}_{it} + b_2 Z_{it} + \varepsilon_{it} \quad (8)$$

$$\text{HDCl}_{it} = c_0 + c_1 \text{DIG}_{it} + c_2 M_{it} + c_3 Z_{it} + \varepsilon_{it} \quad (9)$$

Where a_1 stands for the aggregate utility of the role of DIG, and b_1 refers to the extent of the role of M (the mediator variable). If the coefficient c_1 is significant, it indicates the presence of both direct and indirect effects.

3.2 Variable declaration

(1) Dependent variable: high-quality development of the construction industry

Based on the new development principles of innovation, coordination, green development, openness, and sharing, a measurement system for HDCl has been established. The specific indicator system for HDCl is outlined in Table 1. Figure 2(a) and Figure 2(b) illustrate HDCl for each province in selected years.

In summary, HDCl shows a gradual increase in each province. However, the distribution of HDCl across different regions in China is uneven, with development levels generally decreasing from the southeastern coastal provinces to the northwestern inland areas. In the southeastern coastal provinces, the construction industry has benefited from a relatively stable industrial structure and early urbanization, giving these areas an inherent advantage in HDCl. The western regions face weaker development foundations, although specific regions, such as Sichuan and Chongqing, demonstrate considerable potential in the construction industry.

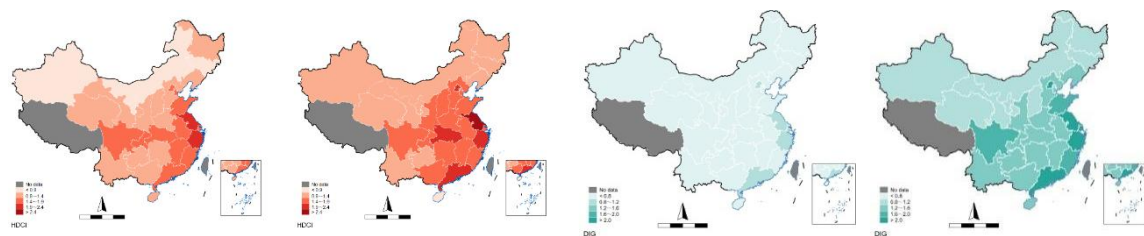


Figure 2 Spatial distribution of HDCl and DIG

(2) Key independent variable: digital economy

According to the connotation of DIG, this paper constructs an indicator system to measure DIG. The specific indicator system for DIG is detailed in Table 2. Figure 2(c) and Figure 2(d) illustrate DIG for each province in selected years.

The results indicate a significant upward trend in DIG across all provinces. Regionally, the highest levels of DIG development are predominantly concentrated in the eastern coastal regions, which is consistent with the location of the HDCl index.

(3) Mediating variables

Government transparency (GOV) is measured using the China Government Transparency Index Report (2012–2021), with missing data filled by interpolation. Marketization (MAR) is assessed using the regional marketization index as calculated by Fan et al.

(4) Control variables

Considering various factors influencing HDCl, this study selected economic development (PGDP), population density (POP), unemployment rate (UNE), and tax burden level (TAX) as control variables. These

variables are expressed by the logarithm of per capita GDP, population density, the unemployment rate, and the ratio of tax revenue to GDP.

3.3 Data source

Considering data accessibility, this study uses panel data from 30 provinces in China from 2012–2021. The primary data sources include the China Statistical Yearbook (2013–2022), the China Construction Industry Statistical Yearbook (2013–2022), and the Peking University Digital Inclusive Finance Index (2012–2021). Descriptive statistics of the variables are presented in Table 3.

Table 1 Indicator System for the Development Level of HDCI

Dimension	Indicators	Attribute	References
Innovation	Number of patents in the construction industry	+	Yang et al., 2024; Wang & Cheng, 2024
	Technical equipment rate	+	
	Power equipment utilization rate	+	
	Regional innovation capability	+	
Coordination	Proportion of total output value of state-owned enterprises	+	Zhang et al., 2023; Gao & Li, 2021; Wang et al., 2019
	Proportion of total output value of general contracting construction enterprises	+	
	Proportion of turnover for special and grade I general contracting	+	
	Proportion of turnover for one-level professional contracting	+	
Green	Cement consumption per 100 million yuan of output value	-	Gao & Li, 2021; Wang & Cheng, 2024
	Wood consumption per 100 million yuan of output value	-	
	Green coverage rate of built-up areas	+	
	Carbon emissions in the construction industry per 100 million yuan of output value	-	
Openness	Proportion of total output value of private enterprises	+	Zhang et al., 2023; Wang et al., 2019
	Proportion of total output value completed outside the province	+	
	Turnover of overseas contracting projects	+	
Sharing	Proportion of total output value in the construction industry	+	Wang & Cheng, 2024; Zhang et al., 2023;
	Average salary of urban employed personnel in the construction industry	+	
	Completed area of residential housing	+	
	Employment contribution	+	

Table 2 Indicator System for the Development Level of the Digital Economy

Dimension	Indicators	Attribute	References
Digital Infrastructure	Number of Internet access ports	+	Dai et al., 2023; Wang & Shao, 2023
	Proportion of Internet users to the resident population	+	
	Mobile phone penetration rate	+	
	Fiber optic cable route density	+	
Digital Industrialization	Number of domain names	+	Li et al., 2024; Lyu et al. 2024; Zhang et al. 2022
	Telecom service total volume	+	
	Number of employees in the information services industry	+	
	Revenue of the electronic information manufacturing industry	+	
Industrial digitalization	Software business revenue	+	Wu et al., 2023; Luo et al., 2022
	Number of websites per hundred enterprises	+	
	Number of computers used per one hundred employees in the enterprise	+	
	E-commerce sales volume	+	
Digital financial	Proportion of enterprises engaged in e-commerce transactions	+	Wu et al., 2023; Luo et al., 2022
	Equivalent full-time personnel for R&D staff in industrial enterprises above a certain scale	+	
	Digital financial coverage breadth	+	
	Depth of digital financial usage	+	
	Degree of digitization of digital finance	+	

4. Empirical results and discussion

4.1 Spatial spillover effect

(1) Spatial correlation test

The prerequisite for applying spatial econometric models is the presence of spatial dependence among variables. Spatial autocorrelation is commonly measured with Moran's I. The results are shown in Table 4. From 2012 to 2021, Moran's I for HDICI and DIG was significant, indicating that HDICI and DIG had significant spatial autocorrelation in China. Thus, applying the spatial Durbin model is appropriate.

Table 3. Descriptive statistics.

Var	Obs	Mean	SD	Min	Median	Max
HDCI	300	1.5432	0.427	0.7432	1.54	2.79
DIG	300	1.0687	0.522	0.1705	1.01	2.96
GOV	300	4.1927	0.172	3.6297	4.23	4.50
MAR	300	2.0669	0.252	1.2116	2.11	2.52
PGDP	300	9.3251	0.464	8.5984	9.20	10.78
POP	300	4.7514	7.079	0.0791	2.93	39.26
UEM	300	3.2173	0.636	1.2100	3.30	4.61
TAX	300	8.2060	2.934	4.4275	7.52	19.97

Table 4. Moran's I of HDICI and DIG from 2012 to 2021

Year	HDICI			DIG		
	Moran's	Z-statistic	p value	Moran's	Z-statistic	p value
2012	0.136	4.766	0.000	0.062	2.715	0.007
2013	0.121	4.389	0.000	0.052	2.455	0.014
2014	0.124	4.493	0.000	0.047	2.330	0.020
2015	0.126	4.543	0.000	0.035	2.000	0.045
2016	0.125	4.497	0.000	0.034	1.952	0.051
2017	0.127	4.547	0.000	0.035	1.983	0.047
2018	0.123	4.398	0.000	0.032	1.915	0.056
2019	0.122	4.329	0.000	0.036	2.036	0.042
2020	0.110	4.021	0.000	0.033	1.955	0.051
2021	0.109	3.979	0.000	0.034	1.959	0.050

(2) Regression results of the Spatial Durbin model

Taking into account the potential biases caused by individual differences and temporal effects, this study utilized the spatial Durbin model with two-way fixed effects to analyze spatial spillover effects.

The results indicate that the regression coefficient for DIG is significantly positive at the 1% level, confirming Hypothesis 1, as shown in Table 5 (2). The existing research also supports the positive impact of the digital economy on the high-quality development of industries (Ding et al., 2021; Li et al., 2024; Feng & Wang, 2024). However, the coefficient of $W \cdot DIG$ is negative and significant, indicating that when considering spatial spillover effects, DIG has a negative impact on HDICI in the adjacent region. This could be due to disparities in the conditions for applying digital technologies. These disparities can form siphoning effects (Scheerder et al., 2017). Regions with a more developed digital economy attract greater concentrations of production factors such as technology and talent, which may restrain HDICI in surrounding regions (Cheng et al., 2023).

4.2 Regional heterogeneity

There is significant regional variation in the impact of DIG on HDICI, as shown in Table 5 (3)–(6). DIG positively influences HDICI in both the eastern and central regions, where the construction industries benefit from higher market demand and more abundant resources. However, the effect in the eastern region is weaker compared to that of the central region. This may be due to digitalization efforts offering relatively limited support for core technological innovation, thus restricting the positive impact of DIG on HDICI in the eastern region. In contrast, the coefficient for the western region is negative, likely due to the relatively weak technological and industrial infrastructure (Hao et al., 2022).

4.3 Robustness tests

(1) Endogenous test

To address potential endogeneity between DIG and HDICI, this study employed instrumental variable estimation. First, we use a one-period lag of the digital economy as an instrumental variable, denoted as $L \cdot DIG$. At the same time, following Nunn and Qian (2014), we constructed interaction terms between the number of international internet users in the previous year and the number of telephones per hundred people in 1984 as further instrumental variables (IV). The results in Columns (1)–(2) of Table 6 show that DIG remains statistically significant and positive at the 1% level, confirming its beneficial impact on HDICI.

(2) Replace the weight matrix

For robustness checks, we replaced the spatial geographical weight matrix with the spatial adjacency weight matrix. The results in Columns (3) of Table 6, confirming that the model results are reliable.

Table5 Result of the benchmark regression

	(1) OLS HDCI	(2) SDM HDCI	(3) East HDCI	(4) Mid HDCI	(5) West HDCI	(6) East-North HDCI
DIG	0.2741*** (0.0444)	0.3148*** (0.0488)	0.5160*** (0.1036)	0.5240*** (0.1789)	-0.5025*** (0.0933)	-0.9759 (0.8097)
_cons	-1.6130*** (0.5785)		7.0877*** (2.4422)	5.3009*** (1.5055)	-1.7465 (1.0825)	5.4769* (2.8433)
W*DIG		-1.1665*** (0.3380)				
Spatial rho		-0.9789*** (0.2486)				
Variance sigma2_e		0.0036*** (0.0003)				
Controls	YES	YES	YES	YES	YES	YES
Year FE	NO	YES	YES	YES	YES	YES
Provincial FE	NO	YES	YES	YES	YES	YES
N	300	300	100	60	110	30

Standard errors in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 6 The robustness tests results.

	(1) DIG	(2) HDCI	(3) HDCI
L.DIG	0.9022*** (0.0124)		
IV	0.0000*** (0.0000)		
DIG		0.2595*** (0.0527)	0.3742*** (0.0550)
Kleibergen-Paap rk LM		260.921***	
Kleibergen-Paap rk Wald F		3779.003 [19.93]	
_cons	-0.0772 (0.1379)	-1.8563*** (0.6175)	
W*DIG			-0.4874*** (0.1102)
Controls	YES	YES	YES

4.4 Mechanism analysis

Based on the theoretical mechanisms, DIG may promote HDCI by enhancing government transparency and advancing marketization processes. To verify whether these mechanisms are valid, this study employed a mediation effect model to empirically test the transmission mechanisms. The result is shown in Table 7.

DIG significantly enhances the HDCI through government transparency, supporting Hypothesis 2. Enhancing transparency significantly reduces information search costs for enterprises, enhances overall industry transparency, diminishes irregular practices and corruption (Li et al., 2017). Moreover, in the early stages of digital transformation, governments provide substantial funding support or assistance for the digital R&D of construction enterprises, promoting technological innovation and enhancing R&D vitality. As transparency increases, government subsidies such as tax breaks, subsidized loans, and financial aid can be more accurately targeted toward enterprises.

By accelerating marketization, DIG can drive HDCI, supporting Hypothesis 3. One possible reason for this is that the application of digital technologies improves the reliability of production planning and effectively addresses the diverse needs of customers (Liang et al., 2024). Furthermore, there is a significant gap between R&D and the commercial application of technology in the construction industry (Zhang et al., 2021). In highly marketized environments, where technical standards and technology transfer systems are more developed, companies receive feedback on technology applications and market validation, reducing the risk of investing in new technologies. If market feedback is positive, companies can confidently proceed with large-scale implementation, further stimulating their innovative enthusiasm.

Table7 Results of the Mediating effect model.

	(1) GOV	(2) HDCI	(3) MAK	(4) HDCI
DIG	0.2224*** (0.0205)	0.1982*** (0.0520)	0.2028*** (0.0249)	0.1120* (0.0440)
GOV		0.3414** (0.1252)		
MAR				0.7996*** (0.0932)
_cons	5.0392*** (0.2665)	-3.3335*** (0.8520)	0.6154 (0.3241)	-2.1051*** (0.5213)
Controls	YES	YES	YES	YES

5. CONCLUSIONS AND IMPLICATIONS

In the context of the explosive growth of the digital economy, its technological advancements and innovative management models offer new pathways for the construction industry to achieve high-quality development. This study used panel data from 30 provinces in China from 2012 to 2021. A projection pursuit model, based on an accelerated genetic algorithm, is employed to assess the levels of DIG and HDCI. Then, the study thoroughly examined the impact and mechanisms through which DIG influences HDCI.

The conclusions are as follows: First, both DIG and HDCI show an upward trend, indicating the expansion of the digital economy and high-quality growth in the construction industry. Second, DIG has a significant, direct, and positive effect on HDCI, although its spatial effect is negative. In terms of impact mechanisms, DIG promotes HDCI by enhancing government transparency and accelerating marketization processes. Geographically, DIG significantly and positively impacts HDCI in the eastern and central regions, while exerting a notable negative effect in the western region.

Based on the main conclusions, this paper proposes the following policy recommendations. First, the government should prioritize the development of digital infrastructure, promote the adoption of digital technologies, and actively drive the digital transformation of traditional industries. Second, regions should implement targeted measures that align with their specific development conditions, fostering a synergistic and mutually reinforcing ecosystem to support coordinated regional growth. Finally, the government should vigorously promote the standardization and normalization of government transparency initiatives, while improving the market order for digital technology innovation in the construction industry.

This study offers a comprehensive framework and empirical analysis of the impact of DIG on HDCI, but has limitations that suggest directions for future research. First, constrained by data availability, the analysis was conducted at the provincial level; future studies could explore city-level data for finer insights. Second, while focusing on DIG's positive impacts, potential negative effects were overlooked, warranting more balanced future analyses.

ACKNOWLEDGMENTS

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: the National Natural Science Foundation of China [grant number 72301044]; the Chongqing Social Sciences Planning Project - Youth Fund [grant number 2022NDQN43]; the Fundamental Research Funds for Central Universities [grant number 2024CDJSKXYGK11].

REFERENCES

- Alizadehsalehi S., Hadavi A., Huang J. (2020). From BIM to Extended Reality in AEC Industry. *Automation in Construction*, 116(103254): 1-13.
- Ammar, M., Haleem, A., Javaid, M., Walia, R., & Bahl, S. (2021). Improving material quality management and manufacturing organizations system through Industry 4.0 technologies. *Materials Today: Proceedings*, 45, 5089-5096.
- B Babalola, A., & Harinarain, N. (2024). Policy barriers to sustainable construction practice in the Nigerian construction industry: an exploratory factor analysis. *Journal of Engineering, Design and Technology*, 22(1), 214-234.
- Bartelsman, E., Haltiwanger, J., & Scarpetta, S. (2013). Cross-country differences in productivity: The role of allocation and selection. *American economic review*, 103(1), 305-334.
- Cheng, W., Li, C., & Zhao, T. (2024). The stages of enterprise digital transformation and its impact on internal control: Evidence from China. *International Review of Financial Analysis*, 92, 103079.
- Cheng, Y., Zhang, Y., Wang, J., & Jiang, J. (2023). The impact of the urban digital economy on China's carbon

- intensity: Spatial spillover and mediating effect. *Resources, Conservation and Recycling*, 189, 106762.
- China Digital Economy Development Research Report. (2024), *CAICT*, Retrieved from website: http://www.caict.ac.cn/kxyj/qwfb/bps/202408/t20240827_491581.htm
- Cong, L. W., Xie, D., and Zhang, L. (2021). Knowledge accumulation, privacy, and growth in a data economy. *Management Science*, 67(10), 6480–6492.
- Dai, S., Su, M., Liu, Y., & Xu, Z. (2023). Digital economy, resource richness, external conflicts, and ecological footprint: Evidence from emerging countries. *Resources Policy*, 85, 103976.
- Del Río Castro, G., González Fernández, M. C., & Uruburu Colsa, Á. (2021). Unleashing the convergence amid digitalization and sustainability towards pursuing the Sustainable Development Goals (SDGs): A holistic review. *Journal of Cleaner Production*, 280, 122204.
- Ding, S., Hu, H., Xu, F., Chai, Z., and Wang, W. (2024). Blockchain-based security-minded information-sharing in precast construction supply chain management with scalability, efficiency and privacy improvements. *Automation in Construction*, 168, 105698.
- Ding, C., Liu, C., Zheng, C., & Li, F. (2021). Digital economy, technological innovation and high-quality economic development: Based on spatial effect and mediation effect. *Sustainability*, 14(1), 216.
- Djibo, B. O. S., Horsey, E. M., & Zhao, S. (2022). Government institutional support and eco-innovation: the moderating role of market performance in Benin's industrial sector. *Journal of Cleaner Production*, 378, 134598.
- Fan, G., Wang, X.L., Ma, G.R., 2011. Contribution of China's marketization process to economic growth. *Econ. Res*, 46 (09), 4–16.
- Feng, J., & Wang, Y. (2024). Does digital inclusive finance promote agricultural development? A test based on threshold and spillover effects. *Finance Research Letters*, 69, 106104.
- Gao, H., & Li, X. (2021). Research on the Evaluation of High-Quality Development of the Construction Industry Based on the Matter-Element Extension Model. *Construction Economics*, 42(11), 85-89.
- Golpîra, H. (2020). Optimal integration of the facility location problem into the multi-project multi-supplier multi-resource Construction Supply Chain network design under the vendor managed inventory strategy. *Expert systems with applications*, 139, 112841.
- Hao, X., Wang, X., Wu, H., & Hao, Y. (2023). Path to sustainable development: Does digital economy matter in manufacturing green total factor productivity?. *Sustainable Development*, 31(1), 360-378.
- Hong, Y., Liu, W., & Song, H. (2022). Spatial econometric analysis of effect of New economic momentum on China's high-quality development. *Research in International Business and Finance*, 61, 101621.
- Jiang, W., & Li, J. (2024). Digital Transformation and Its Effect on Resource Allocation Efficiency and Productivity in Chinese Corporations. *Technology in Society*, 102638.
- Khalesi, H., Balali, A., Valipour, A., Antucheviciene, J., Migilinskas, D., & Zigmund, V. (2020). Application of hybrid SWARA–BIM in reducing reworks of building construction projects from the perspective of time. *Sustainability*, 12(21), 8927.
- Lei, P., Li, X., & Yuan, M. (2024). The consequence of the digital economy on energy efficiency in Chinese provincial and regional contexts: Unleashing the potential. *Energy*, 311, 133371.
- Li, Z., Ouyang, X., Du, K., & Zhao, Y. (2017). Does government transparency contribute to improved eco-efficiency performance? An empirical study of 262 cities in China. *Energy Policy*, 110, 79-89.
- Li, L., Yi, Z., Jiang, F., Zhang, S., & Zhou, J. (2023). Exploring the mechanism of digital transformation empowering green innovation in construction enterprises. *Developments in the Built Environment*, 15, 100199.
- Li, C., Zhu, J., & Tao, C. (2024). How does the innovative factor allocation promote the high quality development of manufacturing industry?. *Journal of Innovation & Knowledge*, 9(3), 100511.
- Liang, S., & Tan, Q. (2024). Can the digital economy accelerate China's export technology upgrading? Based on the perspective of export technology complexity. *Technological Forecasting and Social Change*, 199, 123052.
- Liu, Y., Zhao, X., & Kong, F. (2023). The dynamic impact of digital economy on the green development of traditional manufacturing industry: Evidence from China. *Economic Analysis and Policy*, 80, 143-160.
- Luo, K., Liu, Y., Chen, P. F., & Zeng, M. (2022). Assessing the impact of digital economy on green development efficiency in the Yangtze River Economic Belt. *Energy Economics*, 112, 106127.
- Lyu, Y., Zhang, J., Wang, W., Li, Y., & Geng, Y. (2024). Toward low carbon development through digital economy: A new perspective of factor market distortion. *Technological Forecasting and Social Change*, 208, 123685.
- Ma, D., & Zhu, Q. (2022). Innovation in emerging economies: Research on the digital economy driving high-quality green development. *Journal of Business Research*, 145, 801-813.
- Mubarak, M. F., Tiwari, S., Petraite, M., Mubarik, M., & Raja Mohd Rasi, R. Z. (2021). How Industry 4.0 technologies and open innovation can improve green innovation performance?. *Management of Environmental Quality: An International Journal*, 32(5), 1007-1022.

- Nunn, N., & Qian, N. (2014). US food aid and civil conflict. *American economic review*, 104(6), 1630-1666.
- Nambisan, S., Lyytinen, K., Majchrzak, A., & Song, M. (2017). Digital innovation management. *MIS quarterly*, 41(1), 223-238.
- Ren, X., & Fang, L. (2023). The Integrated Development of Digital Economy and Real Economy: Driving Mechanisms, Constraining Factors, and Path Choices. *Academic Frontier*, 6, 108-111.
- Rodrigues, M., and Franco, M. (2021). Digital entrepreneurship in local government: Case study in Municipality of Fundão, Portugal. *Sustainable Cities and Society*, 73, 103115.
- Schiavi, B., Havard, V., Beddiar, K., & Baudry, D. (2022). BIM data flow architecture with AR/VR technologies: Use cases in architecture, engineering and construction. *Automation in Construction*, 134, 104054.
- Shahbaz, M., Wang, J., Dong, K., & Zhao, J. (2022). The impact of digital economy on energy transition across the globe: The mediating role of government governance. *Renewable and Sustainable Energy Reviews*, 166.
- Scheerder, A., Van Deursen, A., & Van Dijk, J. (2017). Determinants of Internet skills, uses and outcomes. A systematic review of the second-and third-level digital divide. *Telematics and informatics*, 34(8), 1607-1624.
- Trevisan, A.H., Lobo, A., Guzzo, D., Gomes, L.A.d.V., Mascarenhas, J. (2023). Barriers to employing digital technologies for a circular economy: a multi-level perspective. *J. Environ. Manage.* 332, 117437.
- Wan, Q., Ye, J., Zheng, L., Tan, Z., & Tang, S. (2023). The impact of government support and market competition on China's high-tech industry innovation efficiency as an emerging market. *Technological Forecasting and Social Change*, 192, 122585.
- Wang, D., & Cheng, X. (2024). Study on the path of high-quality development of the construction industry and its applicability. *Scientific Reports*, 14(1), 14727.
- Wang, H., & Guo, J. (2024). New way out of efficiency-equity dilemma: Digital technology empowerment for local government environmental governance. *Technological Forecasting and Social Change*, 200, 123184.
- Wang, X., Han, R., and Zhao, M. (2023). Evaluation and impact mechanism of high-quality development in China's coastal provinces. *International Journal of Environmental Research and Public Health*, 20 (2), 1336.
- Wang, W., Qi, Z., & Zhang, L. (2019). The Construction and Evaluation of the Measurement System for High-Quality Development of the Construction Industry in the New Era. *Construction Economics*, 40(12), 21-26.
- Wang, L., & Shao, J. (2023). Digital economy, entrepreneurship and energy efficiency. *Energy*, 269, 126801.
- Wang, J., Wang, B., Dong, K., & Dong, X. (2022). How does the digital economy improve high-quality energy development? The case of China. *Technological Forecasting and Social Change*, 184, 121960.
- Wang, K., Zhao, Y., Gangadhari, R. K., & Li, Z. (2021). Analyzing the adoption challenges of the internet of things (IoT) and artificial intelligence (AI) for smart cities in China. *Sustainability*, 13, 10983.
- Wen, H., Zhang, B., Li, S., Zhang, L., Gui, B., Liu, Z. (2024). Evaluating High-Quality Development in the Construction Industry via the Matter-Element Extension Method: A Case Study of 11 Cities in Zhejiang, China. *Buildings*, 14 (11), 3499.
- Wei, X. (2021). "The Boundary Choices of 'Two Hands' in China's High-Quality Economic Development: An Empirical Study Based on East Asian Transformation Experience. *Asia-Pacific Economy*, 6, 119-125.
- Wu, L., Wan, X., Jahanger, A., Li, M., Murshed, M., & Balsalobre-Lorente, D. (2023). Does the digital economy reduce air pollution in China? A perspective from industrial agglomeration. *Energy Reports*, 9, 3625-3641.
- Xue, Y., Tang, C., Wu, H., Liu, J., Hao, Y., 2022. The emerging driving force of energy consumption in China: does digital economy development matter? *Energy Policy*, 165, 112997.
- Yang, S., Shen, J., Deng, X., Lu, K., Liu, Z., & Cheng, B. (2024). Digital economy revolutionizing green total factor productivity in construction industry. *Technological Forecasting and Social Change*, 206, 123564.
- Zhang, A, Liu, R, & Ma, Y. (2023). Coupling coordination between the digital economy and high-quality development of the construction industry: An empirical analysis based on samples from 31 provincial administrative regions. *Science, Technology and Industry*, 23(02), 60-67.
- Zhang, J., Ouyang, Y., Ballesteros-Pérez, P., Li, H., Philbin, S. P., Li, Z., & Skitmore, M. (2021). Understanding the impact of environmental regulations on green technology innovation efficiency in the construction industry. *Sustainable Cities and Society*, 65, 102647.
- Zhou J., Xu W., Yan D. (2024). Breaking the resource curse: The impact of digital economy on the sustainable transformation of resource-based cities. *Sustainable Cities and Society*, 113: 105707.



Part 9

**AI for Safety and Health in
Building Environment**

TRAJECTORY FORECASTING FOR WORKER SAFETY IN CONSTRUCTION USING TRANSFORMER AND GRAPH ATTENTION NETWORKS

Mohammed Alduais¹, Xinming Li², and Qipei Mei³

1) MSc Student, Department of Civil and Environmental Engineering, University of Alberta, Edmonton, Alberta, Canada. Email: alduais@ualberta.ca

2) Ph.D., Assoc. Prof., Department of Mechanical Engineering, University of Alberta, Edmonton, Alberta, Canada. Email: xinming.li@ualberta.ca

3) Ph.D., Assist. Prof., Department of Civil and Environmental Engineering, University of Alberta, Edmonton, Alberta, Canada. Email: qipei.mei@ualberta.ca

Abstract: With the rise of residential housing demand worldwide, offsite construction emerges as a possible option to speed up construction while improving the safety of workers. However, offsite construction sites are normally a dynamic environment in which workers collaborate with various machinery and large moving objects, resulting in additional safety concerns. Accurate prediction of future trajectories is an important step in building a collision alarm system that can be utilized to mitigate such safety risks. Traditional methods, such as Kalman filters (KF) and Markov processes, rely heavily on past trajectories and hand-crafted features, which fail to account for the dynamic nature of construction sites. With the rising interest in data-driven approaches, several studies have explored different methods of trajectory prediction. Long Short-Term Memory (LSTM) network is one of the major methods used for forecasting future trajectories by leveraging both past individual and contextual information. However, one of the main limitations of LSTM is error accumulation, which limits the model from providing accurate results. Inspired by the success of the transformer model in natural language processing, this paper proposes the use of transformer encoder-decoder architecture with graph attention networks (GATs) to predict worker trajectories on construction sites. The temporal interactions of the workers are captured by the transformer model, while GAT captures the spatial relationships of the workers, which allow the model to build more comprehensive view of the workers behavior. The model is able to take 8 frames, covering 3.2 seconds, and predict the next 12 frames, covering 4.8 seconds, with an average displacement error (ADE) of 1.25 m and a final displacement error (FDE) of 2.3 m. The proposed model improves performance compared to traditional methods such as LSTM.

Keywords: Offsite Construction, Safety in Construction, Worker Trajectory Prediction, LSTM, GAT, Transformer

1. INTRODUCTION

In the coming few years, the construction industry will be one of the main drivers of the global economy. Its growth is anticipated to increase by an average of 4.4% between 2020 and 2025, exceeding the growth rates of the manufacturing and services industry (Robinson et al., 2021). With the rise of new real estate construction, the modular construction market in Europe and the United States will exceed a market value of \$130 billion by 2030 (Bertram et al., 2019). Modular construction is shifting operations from an on-site to an off-site fabrication facility. The fabricated building components are then transported to the site for installation (Thai et al., 2020). One of the main reasons for the rise of modular construction is its elevated safety standards, as emphasized by Jin et al. (2020). Regardless, fabrication facilities are dynamic working environments where unsafe scenarios may occur. These events result from congested working environments and the utilization of construction equipment, such as robotic machines (Teizer & Cheng, 2015). There are several reasons for such hazardous situations: workers' lack of positional awareness when working with construction machinery, which can be due to its sudden movement and rotation, obscured sightlines, and ambient noise (Teizer & Cheng, 2015). Monitoring and predicting the future trajectories of workers is essential for mitigating the risk of potential injuries arising from such unsafe situations. This approach yields vital information that can be employed to create a contact collision alarm system. Several studies have proposed methods for trajectory prediction in the construction industry. For example, Zhu et al. (2016) presented a method designed to predict the trajectory of workers and mobile equipment using a Kalman filter. In this method, the trajectories were obtained using multiple video cameras. However, this method relies on past trajectories to predict future trajectories, neglecting other important information, such as the influence of other entities. The behavior of workers is influenced by their interaction with others and the task they are working on; this is due to the dynamic working environment in a construction site, where multiple entities coexist (Cai et al., 2020). Another study proposed to model the worker's trajectory as a Markov process, and the prediction of future trajectories depends on past historical records. However, this method relies on hand-crafting the features, which is hard to implement for broad range scenarios.

Data-driven methods have emerged as viable techniques in trajectory prediction due to rapid advances in computational power and deep learning techniques. For instance, Cai et al. (2020) proposed a method for trajectory prediction using Long-Short-Term Memory (LSTM). This method takes advantage of both the individual

trajectories, represented by the past trajectories, and the contextual information, represented by the movement of the target neighbors, information about the working group, and the final destination. However, error accumulation is one of the drawbacks of using LSTM, since the model uses the output from the last step to predict the following step, affecting its effectiveness. In 2020, Giuliani et al. proposed a transformer inspired trajectory prediction model. Transformer networks showed a promising result compared to other methods. This paper expands on Giuliani et al. model by integrating graph attention networks and proposes a complete framework from detection to prediction.

2. RELATED WORK

With the advances in the field of autonomous vehicles, trajectory prediction emerges as one of the fundamental challenges that need to be solved, especially in applications such as analyzing pedestrian behavior (Cai et al., 2020). Bayesian models, probabilistic planning, and data driven methods are the three main approaches used in the field of trajectory prediction. Firstly, Bayesian models, Lerner et al. (2007) are considered one of the pioneers who introduced the use of the Extended Kalman filter algorithm for trajectory prediction. This algorithm alternates between a prediction step, where current internal states are computed based on previous states, and an update step, where the current states are refined using observations. In addition, Antonini et al. (2006) proposed a Discrete Choice framework, and Treuille et al. introduced continuum dynamics. One of the disadvantages that hinder the use of Bayesian models in trajectory prediction is the need for handcrafting the internal states, which can be time consuming and very difficult to implement in complex situations (Yang et al., 2023) (Huang et al., 2019). Secondly, probabilistic planning is where the problem we have is modeled using a Markov decision process. This method aims to find the optimal solution by maximizing a reward function and ultimately reaching the goal (Ma et al., 2017). However, A key challenge in implementing probabilistic planning methods is accurately defining a reward function, specifically when dealing with real word trajectory prediction tasks. Thirdly, data-driven approaches show great promise for handling complicated trajectory prediction issues. For instance, Yi et al. (2016) proposed a Convolutional Neural Network (CNN) technique for predicting pedestrian behavior in an area with high density. On the other hand, LSTM networks can detect spatiotemporal information in sequential data. Several research publications proposed using LSTM networks to handle trajectory prediction difficulties. For example, Alahi et al. (2016) added a social pooling layer to an LSTM network to allow LSTM cells to share information internally. In addition, Huang et al. (2019) proposed combining graph attention and LSTM networks to capture the spatial and temporal interactions at each time step. Furthermore, an environment aware LSTM model was proposed by Yang et al. (2023). The proposed model improved the vanilla LSTM model by introducing worker to worker and environment to worker interactions to the model input. Giuliani et al. (2020) introduced a new model that draws inspiration from the renowned transformer networks commonly employed in Natural Language Processing (NLP), as described in the paper "Attention Is All You Need" (Vaswani et al., 2017). The suggested model represents each person independently, with no human-human or scene interaction terms. As a result, this paper aims to extend Giuliani et al.'s (2020) research by introducing graph attention networks to capture spatial interactions.

3. METHOD

During construction activities, construction workers rely on their environmental awareness, such as accounting for walking paths, other construction equipment, and other construction workers, to help mitigate any risk of collision. To integrate such vital information, we propose incorporating transformer networks to capture the temporal interactions and graph attention networks to capture the spatial relationships between construction workers. This becomes particularly essential given that the original transformer model lacked the incorporation of contextual information beyond historical trajectories. Figure 1 shows the proposed methodology. In order to automate the trajectory prediction, YOLOv10 model proposed by Wang et al. (2024) was used to for object detection and to keep track of the workers present in the scene DeepSort proposed by Wojke et al. (2017) was incorporated to the overall framework. Once worker detection and tracking are completed, further data preprocessing is implemented to prepare the data for the training phase. In the model training, the model is initially trained using ETH dataset to initialize the model parameters and using transfer learning the model is re-trained using the new construction dataset. In the following sections, each phase is presented in detail.

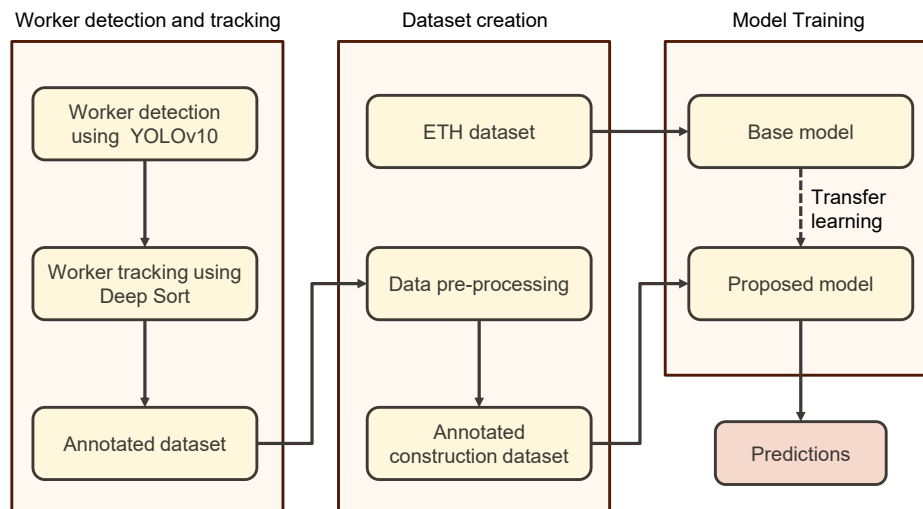


Figure 1. Overview of the proposed methodology

3.1 Worker detection and tracking

In order to automate the processes of detecting and tracking workers and panels, YOLOv10 model was trained using real world construction dataset. In addition to using YOLOv10, DeepSort algorithm was utilized to add another parameter to the dataset, which is tracking ID. This will help in preparing the data to match the input format.

(1) Object detection using YOLOv10

Before starting the object detection phase, a pre-trained model is needed to be used during inference. Therefore, a real-world construction dataset is used to train the YOLOv10 model to detect workers. The scenes were taken from a surveillance camera inside an off-site construction facility in Edmonton, Alberta. The construction dataset contains two scenes where workers were working on different building components such as floors and panels. The construction dataset contains one class: (1) Workers. For consistency with the literature, the scene was annotated at 2.5 fps meaning each frame will be annotated every 0.4 seconds. As Kima et al. (2019) research found that increasing the observation period does not increase the accuracy of the prediction in a construction setting. In the first scene, we annotated 1951 frames with 4 workers. In the second scene, we annotated 212 frames with 4 workers. The annotation was done using Roboflow website. With the help of augmentation functions within their website, data augmentation was implemented to enhance the model performance. The total images used after applying data augmentation is 9388 images.

(2) Object tracking using DeepSort

Tracking workers and moving obstacles is important to help improve the predictive performance of the proposed model. Therefore, DeepSort algorithm was utilized with a trained YOLOv10 model to track the IDs of workers. The extra parameter gained from this step is used later to aggregate the movements based on the IDs. For instance, let's assume worker number 1 appears for 2 minutes. Therefore, in the data pre-processing phase, the worker trajectory can be grouped into a single sequence and then prepare it for training.

3.2 Dataset creation

In this paper, two datasets were used for training and fine-tuning the proposed model. The first dataset is described in the following section will be used to train a base model in order to initialize the model wights. Then, a new construction dataset is created to be used to fine tune the base model using transfer learning.

(1) ETH dataset

ETH name was derived from ETH Zurich University (Pellegrini et al., 2009), this dataset contains two scenes in a bird's eye view from RGB cameras that is usually used for surveillance purposes: (1) Eth; and (2) Hotel. Eth scene was taken from the top of the ETH main building and the annotation was done at 2.5 fps, which means that one frame is annotated every 0.4 seconds. This part comprises of a total of 365 different pedestrian trajectories. For Hotel, the scene was taken from the 4th floor of a hotel in Bahanhofstr in Zurich and the video was taken at a 25 fps, but the annotation was done at 2.5 fps for consistency. The original annotation format is listed below:

[frame number, pedestrian_ID, pos_x, pos_z, pos_y, v_x, v_z, v_y]

However, in our case the format is changed to a new format presented below:

[frame number, pedestrian_ID, pos_x, pos_y]

(2) Construction dataset

The second dataset used in this paper is a real-world construction dataset. This dataset is the output from worker detection and tracking phase after pre-processing with a total of 1900 frames. The data pre-processing comprises of taking the rectangle coordinates of the detected worker and convert it into a single point, in our case the center of the rectangle. After that the pixel coordinates are converted to real world coordinates. Resulting into [pos_x, pos_y]. For Frame_number can be obtained from the YOLOv10 results and pedestrian_ID can be obtained from DeepSort results. Therefore, the final format will follow the same format as ETH dataset. A sample frame from the construction dataset is shown in Figure 2.



Figure 2. A sample frame from the construction dataset

3.3 Model Training

In following sections, we will describe the proposed model architecture, base model training, and the proposed model training.

(1) Proposed model architecture

As described earlier, we propose to incorporate graph attention networks (GAT) (Veličković et al., 2017), to the transformer encoder-decoder architecture to help decode the spatial information of the workers see Figure 4. GAT is used on graph-based data, which uses self-attention to calculate the importance of each neighbor on the target node. GAT have two main components: (1) Nodes, and (2) Edges. Where nodes can represent individual points in the data, worker in our case, and the edges represent the relationship between the target node and its neighbors. The transformer architecture consists of three main parts: (1) Positional encoding, (2) Encoder, and (2) Decoder.

Time is an important factor that should be considered in sequential problems. Unlike most RNNs that incorporate time, transformer models do not possess the ability to incorporate time as a factor. Therefore, the authors of the transformer model proposed the use of positional encoding. This positional encoding has the same dimensions as the input/output embedding. So, it can be added to the input/output embedding. In their work, they used sine and cosine functions to encode some information about the token position. The positional encoding function is described below:

$$PE_{(pos,2i)} = \sin\left(\frac{pos}{10000^{\frac{2i}{d}}}\right) \quad (1)$$

$$PE_{(pos,2i+1)} = \cos\left(\frac{pos}{10000^{\frac{2i+1}{d}}}\right) \quad (2)$$

Where pos is the position of the token in the sequence, i is the dimension index, and d is the dimension of the model.

The main objective of the Encoder is to encode the source inputs, so that it can later be used by the Decoder. The encoder consists of N identical layers working in parallel, normally $N = 6$. There are two sub-layers for each of the main layers: (1) Multi-head Attention, and (2) Fully connected Feed-Forward Network (FFN). The authors employed another connection called residual connection, which will connect the two sub-layers. After that a normalization layer is used, which takes the output from the previous sub-layer and add the residual connection than apply layer normalization.

The main objective of the decoder is taking the information from the encoder and GAT and generate the target sequence. The structure of the decoder is similar to that of the encoder with the difference it employs two attention modules. The first attention module will take the encoded information from the encoder and the second module will take the output embedding by applying masking. The attention mechanism used in both the encoder and the decoder is shown in Eq. 3.

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) \quad (3)$$

where Q , K , and V are entry sequence matrices named “query”, “keys”, and “values” respectively. And d_k represents the dimensions vector corresponding to the “key” matrix.

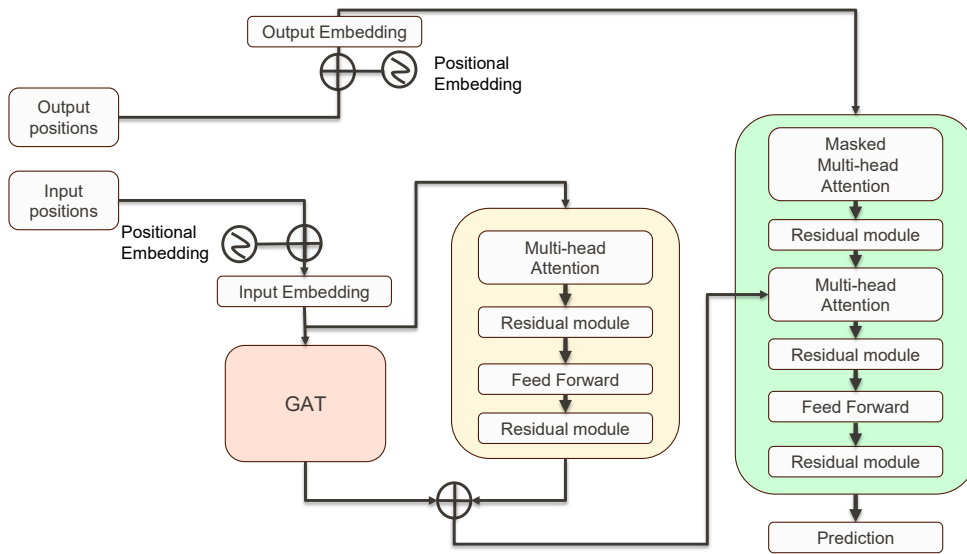


Figure 3. Overview of proposed model architecture

(3) Implementation details evaluation metrics

The model was trained using PyTorch on a desktop computer equipped with an RTX 3060 GPU, an AMD Ryzen 9 5900 12-Core Processor 3.00 GHz, and 32.0 GB RAM. The network was trained with the Adam optimizer, employing a learning rate of 0.001, a batch size of 64, and a total of 600 epochs.

Initially, the model was trained on the ETH dataset, then transfer learning was applied, and the model was retrained using a real construction dataset. The dataset was split as follows: 70% for training, 20% for validation, and 10% for testing the model performance. The model will take eight frames as input and predict twelve future frames. We conducted tests using actual construction operations data to assess the trajectory prediction accuracy of the trained models. Our evaluation metrics include both average displacement error (ADE) and final displacement error (FDE) (Pellegrini et al., 2009). ADE is calculated as the mean square error across all predicted positions of the trajectory and the corresponding ground-truth trajectory, represented by the following formula:

$$ADE = \frac{\sum_{i=1}^N \sum_{t=T_{obs}+1}^{T_{pred}} \|(\hat{x}_i^t, \hat{y}_i^t) - (x_i^t, y_i^t)\|}{N \times (T_{pred} - T_{obs} - 1)} \quad (4)$$

In the given formula, N represents the number of workers, and $(\hat{x}_i^t, \hat{y}_i^t)$ and (x_i^t, y_i^t) denote the predicted coordinates of worker i at time instant t and the corresponding ground-truth coordinates of worker i at time instant t , respectively. $\| \cdot \|$ is the Euclidean distance. FDE is determined as the distance between the predicted final destination and the corresponding ground truth destination, represented by the following formula:

$$FDE = \frac{\sum_{i=1}^N \left\| \left(\hat{x}_i^{T_{pred}}, \hat{y}_i^{T_{pred}} \right) - \left(x_i^{T_{pred}}, y_i^{T_{pred}} \right) \right\|}{N} \quad (5)$$

Where N represents the number of workers, and $\left(\hat{x}_i^{T_{pred}}, \hat{y}_i^{T_{pred}} \right)$ and $\left(x_i^{T_{pred}}, y_i^{T_{pred}} \right)$ denote the predicted coordinates of worker i at time instant t and the corresponding ground-truth coordinates of worker i at time instant t , respectively. $\| \cdot \|$ is the Euclidean distance. The units of ADE and FDE is in meters, given that both the ETH dataset and the construction datasets use real-world coordinates.

3. RESULTS AND DISCUSSION

In the upcoming section, I will present the implementation details, outline the evaluation metrics employed, and present the results of the testing phase.

3.3 Results

In this section, the results obtained from the proposed model is discussed. First, the results from training YOLOv10 are presented. Second, the results obtained from the proposed model is compared with the literature to identify the improvements.

(1) YOLOv10 results

The training setup was as follows: a total of 9,288 images were used for training, validation, and testing, with a split of 70%, 20%, and 10% for training, validation, and testing, respectively. The training was conducted on a PC equipped with an NVIDIA RTX 3070 GPU with 12 GB of memory, an AMD Ryzen 9 5900 12-Core Processor at 3.00 GHz, and 32 GB of RAM. The model was trained for 400 epochs with a batch size of 64 and an image size of 640.

The model achieved the following results: a precision of 99.39%, a recall of 99.4%, a mean Average Precision (mAP) of 93.7%, and a mean Average Precision at 50 (mAP50) of 99.4%. The results indicated effectiveness of the model, as it achieved high values of precision, recall, mAP, and mAP50. Figure 4 shows three training loss for three parameters: (1) Box loss, (2) Class loss, and (3) DFL loss. Where Box loss (bounding box regression loss), is the difference between the predicted bounding boxes and the actual. For Class loss, is the loss associated with classification of objects within the predicted bounding box. These two help the model in accurately learn how to predict an approximate location of the bounding box and the class associated with it. Lastly, DFL (Distribution Focal Loss), is a loss function that is used to enhance the bounding box regression localization accuracy. Also, Figure 5 shows the variation of mAP and mAP50 across the training epochs.

The YOLOv10 model results demonstrated its ability to accurately detect both classes with high accuracy. However, performance challenges arise when applying the model to a different working environment. This can be attributed to variations in lighting, different equipment, and changes in camera positioning. These factors result in an under-performing model that cannot accurately detect the presence of workers.

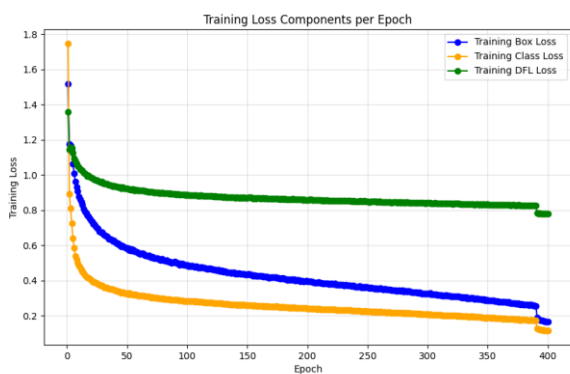


Figure 4. Training loss curves for YOLOv10

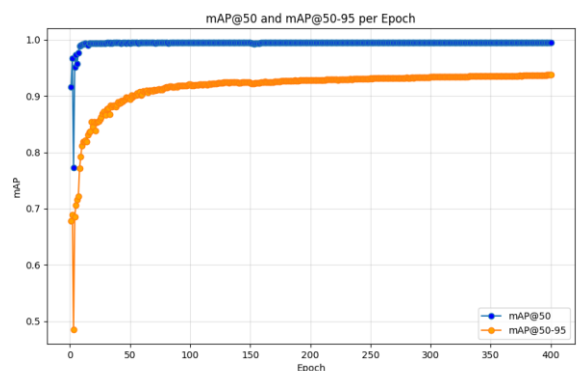


Figure 5. mP50 and mAP curves for YOLOv10

(2) Proposed model results

To measure the performance of the proposed model, four models are used as a base line. The training for all base models was conducted in accordance with a set of parameters to reduce any discrepancies. The

observation and the prediction windows were kept the same. As changing this parameter, the overall performance will dramatically change. As discussed in previous sections the observation window will be 8 frames and the model predicts the next 12 frames. Similarly, the batch size and the epochs used are 4 and 600, respectively. The results are summarized in Table 1, which shows that the proposed model outperformed all the baseline models with an improvement percentage range from approximately 27.92% to 31.51% for MAD and an improvement percentage range from approximately 26.57% to 29.84% for FAD.

Table 1. Results comparison

Model Name	MAD	FAD
LSTM	1.753	3.332
SGAN	1.767	3.343
EA-Distance	1.825	3.349
EA-Direction	1.734	3.201
Proposed model	1.25	2.35

Figure 6 presents a comparison between a single prediction made by the proposed model and the corresponding ground-truth values. The figure illustrates the actual trajectories of three workers alongside their predicted trajectories. The results indicate that the model successfully captures the overall movement pattern of the workers, even when irregular movements occur, as seen with worker 3. However, in some cases, the predicted trajectories' lengths do not align with the ground truth. A possible explanation for this discrepancy lies in the structure of the dataset itself. Specifically, the workers' movements are significantly different from those of pedestrians. Workers typically have prolonged stops while completing tasks before moving to the next location. As a result, their trajectories should reflect a balance between long and short movements.

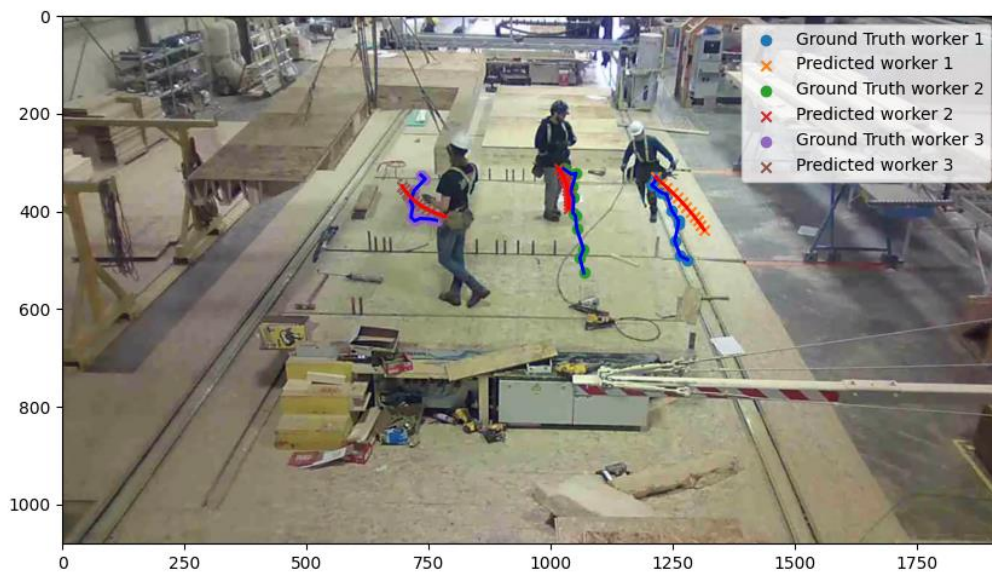


Figure 6. Predicted vs ground truth trajectories of the proposed model

5. CONCLUSIONS

In conclusion, this study has presented an innovative approach to predicting worker trajectories on dynamic construction sites by integrating transformer encoder-decoder architecture with graph attention networks (GATs). The proposed model effectively captures both the temporal and spatial interactions of workers, enabling more accurate trajectory predictions. By leveraging both transformers for temporal analysis and GATs for spatial context, our model outperforms traditional methods, such as LSTM, particularly in mitigating error accumulation and addressing the complex, interactive environments characteristic of construction sites. Testing with real-world data validated the model's superior performance, demonstrating lower average displacement error (ADE) and final displacement error (FDE) than baseline methods.

This trajectory forecasting framework holds substantial potential for enhancing worker safety by forming the basis for collision warning systems in construction environments. Future work could explore adapting

the model to various construction environments and expanding it to incorporate additional dynamic elements, such as machinery or varying site conditions, to further enhance robustness.

ACKNOWLEDGMENTS

This study was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) through Alliance Grant with Alberta Innovates (ALLRP 561120 - 20). The authors would also like to acknowledge the support from their industry partner in Canada.

REFERENCES

- Alahi, A., Goel, K., Ramanathan, V., Robicquet, A., Fei-Fei, L., & Savarese, S. (2016). Social LSTM: Human Trajectory Prediction in Crowded Spaces. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 961–971. <https://doi.org/10.1109/CVPR.2016.110>
- Antonini, G., Bierlaire, M., & Weber, M. (2006). Discrete choice models of pedestrian walking behavior. *Transportation Research Part B: Methodological*, 40(8), 667–687. <https://doi.org/10.1016/j.trb.2005.09.006>
- Bertram, N., Fuchs, S., Mischke, J., Palter, R., Strube, G., Woetzel, J., & McKinsey. (2019). Modular construction: From projects to products. *Capital Projects & Infrastructure*. Retrieved from <https://www.mckinsey.com/~media/mckinsey/business%20functions/operations/our%20insights/modular%20construction%20from%20projects%20to%20products%20new/modular-construction-from-projects-to-products-full-report-new.pdf>
- Cai, J., Zhang, Y., Yang, L., Cai, H., & Li, S. (2020). A context-augmented deep learning approach for worker trajectory prediction on unstructured and dynamic construction sites. *Advanced Engineering Informatics*, 46, 101173. <https://doi.org/10.1016/j.aei.2020.101173>
- Giuliani, F., Hasan, I., Cristani, M., & Galasso, F. (2021). Transformer Networks for Trajectory Forecasting. *2020 25th International Conference on Pattern Recognition (ICPR)*, 10335-10342. <https://doi.org/10.1109/ICPR48806.2021.9412190>
- Huang, Y., Bi, H., Li, Z., Mao, T., & Wang, Z. (2019). STGAT: Modeling Spatial-Temporal Interactions for Human Trajectory Prediction. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 6271–6280. <https://doi.org/10.1109/ICCV.2019.00637>
- Jin, R., Hong, J., & Zuo, J. (2020). Environmental performance of off-site constructed facilities: A critical review. *Energy and buildings*, 207, 109567. <https://doi.org/10.1016/j.enbuild.2019.109567>
- Kim, D., Liu, M., Lee, S., & Kamat, V. R. (2019). Trajectory prediction of mobile construction resources toward pro-active struck-by hazard detection. In M. Al-Hussein (Ed.), *Proceedings of the 36th International Symposium on Automation and Robotics in Construction (ISARC)* (pp. 982–988). Banff, Canada: International Association for Automation and Robotics in Construction (IAARC). <https://doi.org/10.22260/ISARC2019/0131>
- Lerner, A., Chrysanthou, Y., & Lischinski, D. (2007). Crowds by Example. *Computer Graphics Forum*, 26(3), 655–664. <https://doi.org/10.1111/j.1467-8659.2007.01089.x>
- Ma, W.-C., Huang, D.-A., Lee, N., & Kitani, K. M. (2017). Forecasting Interactive Dynamics of Pedestrians with Fictitious Play. *arXiv*. <http://arxiv.org/abs/1604.01431>
- Pellegrini, S., Ess, A., Schindler, K., & van Gool, L. (2009). You'll never walk alone: Modeling social behavior for multi-target tracking. *2009 IEEE 12th International Conference on Computer Vision*, 261-268. <https://doi.org/10.1109/ICCV.2009.5459260>
- Robinson, G., Leonard, J., Whittington, T., Oxford Economics, Marsh, Guy Carpenter, & Turner, J. (2021). Future of construction. *Oxford Economics*. Retrieved from <https://www.oxfordeconomics.com/wp-content/uploads/2023/08/Future-of-Construction-Full-Report.pdf>
- Teizer, J., & Cheng, T. (2015). Proximity hazard indicator for workers-on-foot near miss interactions with construction equipment and geo-referenced hazard areas. *Automation in Construction*, 60, 58–73. <https://doi.org/10.1016/j.autcon.2015.09.003>
- Thai, H.-T., Ngo, T., & Uy, B. (2020). A review on modular construction for high-rise buildings. *Structures*, 28, 1265–1290. <https://doi.org/10.1016/j.istruc.2020.09.070>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. *arXiv*. <https://doi.org/10.48550/arxiv.1706.03762>
- Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2017). Graph attention networks. *arXiv*. <https://doi.org/10.48550/arxiv.1710.10903>
- Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., & Ding, G. (2024). YOLOV10: Real-Time End-to-End Object Detection. *arXiv*. <https://doi.org/10.48550/arxiv.2405.14458>
- Wojke, N., Bewley, A., & Paulus, D. (2017). Simple online and realtime tracking with a deep association metric. *2017 IEEE International Conference on Image Processing (ICIP)*, 3645-3649. <https://doi.org/10.1109/ICIP.2017.8296962>

- Yang, Q., Mei, Q., Fan, C., Ma, M., & Li, X. (2023). Environment-Aware Worker Trajectory Prediction Using Surveillance Camera in Modular Construction Facilities. *Buildings*, 13(6), 1502. <https://doi.org/10.3390/buildings13061502>
- Yi, S., Li, H., & Wang, X. (2016). Pedestrian Behavior Understanding and Prediction with Deep Neural Networks. In *Computer Vision – ECCV 2016* (Vol. 9905, pp. 263–279). Springer International Publishing. https://doi.org/10.1007/978-3-319-46448-0_16
- Zhu, Z., Park, M.-W., Koch, C., Soltani, M., Hammad, A., & Davari, K. (2016). Predicting movements of onsite workers and mobile equipment for enhancing construction site safety. *Automation in Construction*, 68, 95–101. <https://doi.org/10.1016/j.autcon.2016.04.009>

A SELF-ATTENTION FUSION-BASED BI-LSTM FRAMEWORK FOR OCCUPANT-CENTRIC PREDICTION OF INDOOR ENVIRONMENTAL FACTORS

Min Jae Lee¹ and Ruichuan Zhang, Ph.D.²

1) Ph.D. Student, Myers-Lawson School of Construction, Virginia Polytechnic Institute and State University, Blacksburg, VA. Email: minjae0624@vt.edu

2) Assistant Professor, Myers-Lawson School of Construction, Virginia Polytechnic Institute and State University, Blacksburg, VA. Email: ruichuanz@vt.edu

Abstract: The integration of occupant data into the management of indoor environment factors is gaining increasing attention for creating intelligent and inclusive built environments. Traditional approaches have mostly relied on static models, often failing to account for the ever-changing nature of occupant behavior and environmental factors across time and dimensions. Recent advancements in deep learning, especially deep sequential models capable of capturing both local and global dependencies between time steps, provide an opportunity to overcome these challenges. To address these challenges, the authors propose an LSTM-based model framework that utilizes self-attention-based multimodal fusion, real-time occupant data, indoor environmental quality (IEQ) data, outdoor air quality data, and meteorological data to predict future IEQ conditions, preferred IEQ conditions, and classify current IEQ conditions based on collected occupant feedback. To develop and test the proposed framework, four key steps were followed: (1) collecting IEQ data through smart sensors, (2) collecting perceived occupant feedback, (3) collecting meteorological and Outdoor Air Quality (OAQ) data, and (4) developing an attention-fusion-based Bi-Directional LSTM(Bi-LSTM) model. The proposed framework was tested at the Virginia Tech Blacksburg campus, showing promising results.

Keywords: Deep Learning, Indoor Environment Quality, Attention-based Multimodal Fusion, and LSTM.

1. INTRODUCTION

Indoor Environmental Quality (IEQ) includes various aspects, including indoor air quality, lighting, thermal conditions, and acoustics, which are crucial in how occupants feel in indoor spaces (USGBC 2014). Furthermore, in today's modern society, humans have largely become an indoor species, spending up to 90% of their time indoors (EPA 2023). Many studies have found that well-maintained IEQ conditions result in improved occupant well-being, while sub-optimal IEQ conditions result in sick building syndromes and other health risks (Samet and Spengler 2003, Shan et al. 2018, Cincinelli and Martellini 2017, Jones 1999). Furthermore, buildings are becoming more dynamic and complex, especially with climate changes, which calls for integrating holistic and multiple factors into IEQ modeling and assessment (Smarra et al. 2018, Du et al. 2021, Liu et al. 2022).

Previous research in IEQ modeling and assessment for indoor environment management, while comprehensive, has been limited. Tien et al. (2022) stated that existing efforts for IEQ modeling and assessment often rely on traditional statistical models, which do not leverage the capabilities of deep learning to analyze complex interactions among multiple IEQ factors. Furthermore, Aldakheel (2023) reviewed the application of Artificial Intelligence (AI) in building management systems, noting that while there is a growing interest in using deep learning for energy efficiency and comfort, many studies have yet to explore its full potential in real-time IEQ monitoring and management. Second, previous studies have approached IEQ modeling and assessment using isolated data streams, neglecting the holistic integration of indoor and outdoor environmental factors. Lolli et al. (2022) stated that many existing models overlook the integration of indoor and outdoor data, which can result in an incomplete understanding of the indoor environment. Similarly, Appau (2023) conducted a post-occupancy evaluation of university student housing, highlighting that many studies focus solely on indoor conditions without considering the influence of external environmental factors, such as outdoor air quality and weather conditions. Furthermore, Sun et al. (2019) stated that the development of indoor air quality monitoring systems that often rely on isolated data streams can compromise the quality and reliability of the data collected.

To address these gaps, this paper proposes a novel framework that utilizes attention-based multimodal fusion with Bi-LSTM model architecture that can predict future IEQ conditions and classify current IEQ conditions from multidimensional IEQ data, meteorological data, outdoor air quality (OAQ) data, and occupant feedback. The proposed framework for IEQ prediction and classification consists of four main components: (1) collecting IEQ data through smart sensors, (2) collecting perceived occupant feedback, (3) collecting meteorological and OAQ data, and (4) developing an attention-fusion-based Bi-LSTM model to predict future IEQ conditions as well as classify current IEQ conditions based on collected occupant feedback.

2. BACKGROUND

2.1 Importance of IEQ

As humans spend more time in indoor spaces, reaching up to 90% of their time (EPA 2023), Indoor Environmental Quality (IEQ) conditions are increasingly recognized as critical factors impacting occupant health and productivity within indoor environments. Recent systematic reviews highlighted the need for comprehensive and standardized IAQ data and measurement methods and emphasized that understanding of IAQ is still limited compared to outdoor air quality (Vardoulakis et al. 2020). Additionally, studies have shown that high concentrations of indoor air pollutants, such as CO₂, can lead to symptoms associated with Sick Building Syndrome (SBS), underscoring the importance of effective IEQ assessment in office environments (Shamsudin 2023). Zahaba et al. (2022) also showed that human exposure to indoor air pollutants can be significantly higher than that of outdoor pollutants, which raises concerns about the health impacts of poor IAQ. Furthermore, Deng et al. (2023) demonstrated that IEQ significantly affects office productivity, particularly in the context of the COVID-19 pandemic, highlighting the need for improved IAQ management strategies (Deng et al., 2023). A study has also found a relationship between occupant mental health and the indoor environment, further highlighting the importance of well-maintained IEQ conditions (Beemer et al. 2019).

As the importance of well-maintained IEQ conditions increases, accurately modeling and assessing IEQ conditions are also becoming important. Tagliabue et al. (2021) predicted future IAQ following a data-driven approach with an IoT sensor network. Similarly, Lee and Zhang (2024) predicted future IEQ conditions at educational facilities utilizing multimodal occupant feedback along with multidimensional IEQ data. Kapoor et al. (2022) predicted future CO₂ concentration levels for office rooms using a machine learning-based approach. Fritz et al. (2022) evaluated machine learning models to classify occupants' perceptions of their indoor environment. IEQ models can be integrated into downstream applications such as smart building systems to enable occupant-centric control, where environmental settings are adjusted based on the predictions from IEQ models. Despite the advancements in predictive modeling and machine learning applications for IEQ, there remains a gap in fully understanding the complex relationship between IEQ and its effects on occupants along with meteorological and OAQ conditions.

2.2 Deep Learning in Indoor Environments

The application of deep learning-based models in indoor environments has gained attraction, particularly for tasks such as indoor localization and environmental monitoring. Deep learning, which is a subset of machine learning, holds the improved ability to compute and classify large and complex datasets. Deep learning, when combined with multimodal data, can result in improved model performance (Meng et al. 2020). Recent studies have demonstrated the effectiveness of deep learning models in enhancing accuracy. For instance, Chenari et al. (2017) developed a CO₂-based demand-controlled ventilation strategy using deep learning algorithms, which not only improves IAQ but also optimizes energy consumption. Lee and Zhang (2024) developed an IEQ prediction model, based on Convolutional Neural Network (CNN), achieving high performance. Similarly, Rizk et al. (2019) showcased the use of deep learning models, specifically Convolutional Neural Networks (CNNs), for indoor localization based on Received Signal Strength (RSS) and Channel State Information (CSI), achieving notable improvements in localization accuracy. However, challenges remain regarding the generalization of these models across different indoor environments and the variability of data inputs, highlighting a need for further research to enhance model robustness and applicability (Crețescu et al. 2019).

2.3 Multimodal Data Fusion

Multimodal data, collected from different sources and sensors, encompasses information that can show unique insights into each type of data as well as the overall system of interest (Lahat et al. 2016). Gao et al. (2020) surveyed deep learning for multimodal data fusion, emphasizing the importance of effectively combining different modalities to improve event understanding, especially when one modality is incomplete. Multimodal data fusion is increasingly recognized as a vital approach for improving Indoor Environmental Quality (IEQ) assessments and modeling systems. Out of different algorithms for fusing multimodal data, attention-based fusion is gaining traction due to its ability to selectively focus on the most relevant features across modalities, thereby enhancing the overall model performance and resilience to missing data. Lin et al. (2023) fused multi-sensor data based on a self-attention mechanism. Similarly, Chan-To-Hing and Veeravalli (2024) proposed a cross-attention-based data fusion approach for Masked Autoencoders in remote sensing. Zhao et al. (2022) proposed an attention-based multimodal fusion model for human activity recognition. Liu et al. (2023) proposed a Transformer-based fusion model with modality-specific tokens to achieve effective cross-modal interaction. While multimodal data fusion is increasingly being applied across various domains, its application to IEQ or indoor environment modeling remains limited. Despite its potential to

enhance the accuracy and comprehensiveness of indoor environment assessments, the use of attention-based fusion techniques in this field is still in its early stages, with only a few studies exploring these capabilities for effective IEQ prediction and control.

3. PROPOSED FRAMEWORK

The proposed framework for IEQ prediction and classification consists of four main components (Figure 1): (1) collecting IEQ and IAQ condition data through smart sensors, (2) collecting occupant feedback through a developed HBI UI, (3) collecting meteorological and OAQ data through an API, and (4) developing an self-attention based Bi-LSTM model to predict future IEQ conditions as well as classify current IEQ conditions based on collected occupant feedback.

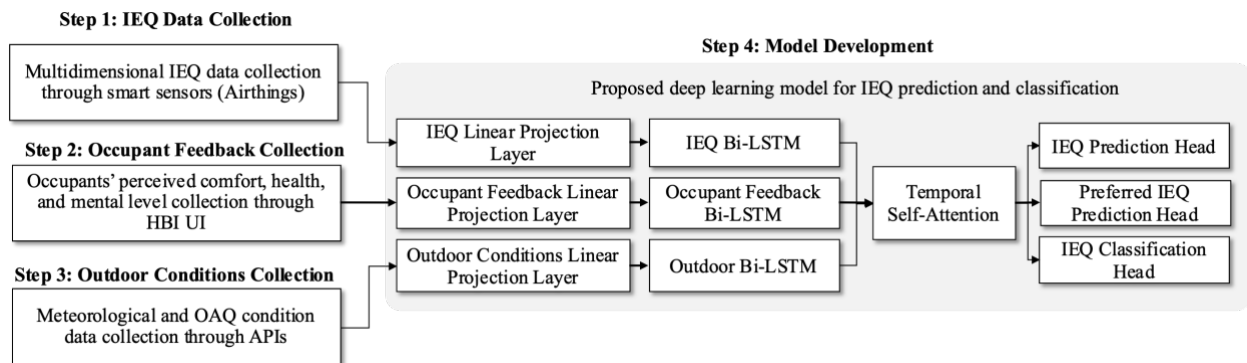


Figure 1. Proposed Framework

3.1 IEQ Data Collection

The first step of the proposed framework involves collecting real-time IEQ condition data using smart sensors, specifically the Airthings View Plus device. This sensor collects variable IEQ and IAQ condition data, including temperature, humidity, CO₂ levels, VOC concentration levels, indoor air pressure, PM₁, PM_{2.5}, and Radon levels. All these variables are crucial for assessing and modeling indoor environmental conditions as well as occupant comfort. For example, high levels of CO₂ can indicate insufficient ventilation, leading to potential cognitive impairment, while VOCs and particulate matter are linked to respiratory issues and general discomfort. By capturing these diverse parameters continuously in real-time, the framework aims to provide a comprehensive dataset that reflects the dynamic nature of indoor environments.

3.2 Occupant Feedback Collection

The second step of the proposed framework includes collecting occupant feedback using a Human-Building Interaction (HBI) User Interface (UI) developed by the authors. This UI is designed to be accessible and intuitive, allowing occupants to indicate their perceived comfort, health, and mental level on a 5-point Likert scale. This subjective feedback serves as a valuable complement to the objective sensor data collected in the first step, as it helps bridge the gap between measured environmental conditions and their perceived impact on the occupants. By capturing individual perceptions, the framework can better account for variability in comfort preferences among different users, which may be influenced by factors such as age, health, or personal sensitivity to environmental changes. The collection of this feedback also allows for the identification of temporal patterns in occupant comfort, which can be analyzed alongside the objective IEQ data to create a more adaptive indoor environment that responds to the occupants' needs in real-time.

3.3 Outdoor Conditions Data Collection

The third step of the proposed framework involves collecting meteorological data and OAQ data through an API, specifically the National Oceanic and Atmospheric Administration (NOAA) API, and the OAQ data being collected from a local weather station based on the geographical location, providing additional context and perspective for the indoor environment. These data include outdoor temperature (°F), outdoor humidity (%), wind speed (mph), precipitation (in), cloud cover (%), PM_{2.5} (µg/m³), PM₁₀ (µg/m³), CO₂ (ppb), O₃ (ppb), and CO (ppm), which can all collectively influence IEQ and IAQ conditions as well as how occupant feel in indoor environments. Integrating outdoor condition data allows the model to consider external factors that can potentially impact the indoor

environment, allowing predictions and classifications to be more comprehensive and holistic, and account for variations in weather and OAQ conditions.

3.4 Attention-Fusion based Bi-LSTM Model Development

The proposed framework's final step involves developing a Bi-directional LSTM (Bi-LSTM) with a temporal self-attention model for IEQ predictions and classifications. The proposed model is a deep learning-based architecture designed to handle temporal and multimodal data, incorporating Bi-LSTM layers and interpretable temporal attention mechanisms. This architecture allows the model to capture both local temporal patterns through the LSTM layers and global dependencies across time steps using the self-attention mechanism. The proposed model also includes three final heads: (1) predicting future IEQ conditions without considering occupant preferences, (2) predicting preferred IEQ conditions based on collected occupant feedback (i.e., comfort and health levels) using a weighted loss function, and (3) classifying IEQ conditions based on occupant feedback. This design allows the model architecture to simultaneously focus on current conditions, adjust for occupant preferences, and provide a classification of occupant well-being. The interpretable temporal self-attention mechanism enables the identification of influential features from the collected data that affect the predictions. The proposed model architecture consists of three key components: (1) data processing layers, (2) attention-fusion layers, and (3) prediction and classification heads.

- (1) **Data Processing Layers:** The model begins with linear projection layers that project each data modality (IEQ, occupant feedback, and outdoor conditions) into a common hidden dimension. This ensures consistency across all input data types and prepares each modality for further processing. The input sequences are then fed into separate Bi-LSTM layers for each modality, following the model fusion approach. The Bi-LSTM layers allow the model to capture temporal dependencies within each data stream, leveraging the bidirectional processing to incorporate both past and future data.
- (2) **Attention-Fusion Layers:** The outputs of the Bi-LSTM layers are then processed by temporal self-attention layers. These attention layers are applied separately to each modality, allowing the model to focus on important time steps within the data sequences. The self-attention mechanism helps identify which time steps in the sequence contribute the most to the prediction and classification tasks, capturing both short-term variations and long-term trends. The attention outputs from all three modalities (IEQ, outdoor, feedback) are then fused, combining the learned features into a unified representation. This fusion process involves stacking the attention outputs from each modality and averaging them, followed by a fully connected layer to reduce the dimensionality and integrate the information. Additionally, the model incorporates learnable weights for each modality, which allows the model to prioritize certain data streams based on their relevance to the task at hand.
- (3) **Prediction and Classification Heads:** The fused representation is passed through separate fully connected layers for both regression and classification tasks. The proposed model includes three output heads: IEQ Prediction Head: Predicts future IEQ conditions without considering occupant preferences, using a regression approach. Preferred IEQ Prediction Head: Predicts conditions that are more favorable to the occupants based on collected feedback, incorporating a weighted loss function to give higher importance to preferred conditions. Classification Heads for Occupant Feedback: Outputs predictions for occupant comfort and health levels, treating these as classification tasks.

4. EXPERIMENTS AND RESULTS

4.1 Data Collection

To test the proposed framework, a set of experiments was conducted from an academic building on the Virginia Tech Blacksburg campus. The selected building represents a typical academic environment with varied occupant activities ranging from self-studying to listening to lectures. IEQ and IAQ data were recorded every 2 minutes, resulting in a dataset capable of capturing subtle changes in environmental conditions. Occupant feedback was collected through a Human-Building Interaction (HBI) UI, and additional feedback data were augmented using an AI-based algorithm to increase the volume and diversity of the dataset for this experiment. Meteorological and Outdoor Air Quality (OAQ) data were collected from an API and augmented using a similar AI-based approach to provide more comprehensive coverage. After augmentation, the total number of data points for each data type was 23,636.

4.2 Data Preprocessing

Several preprocessing steps were followed to ensure that the data was in the correct format and suitable for effective and accurate analysis as well as prediction and classification by the proposed model. (1) *Loading and Merging Datasets*. The IEQ, outdoor, and occupant feedback datasets were loaded and merged based on a timestamp, double-checking that all modalities were aligned, resulting in a cohesive dataset for training the model. Furthermore, the recorded timestamp was converted to a consistent datetime format, and the merged dataset was sorted chronologically to maintain temporal dependencies. (2) *Data Normalization*. All collected data, including IEQ/IAQ data, occupant feedback data, and meteorological and OAQ data, were normalized using the MinMaxScaler to bring all features to a common scale, allowing more effective learning by the proposed model. (3) *Imputation for Feedback*. Missing feedback values were handled by first forward-filling to extend previously recorded feedback until new feedback was recorded. This ensured the continuity of feedback data, which is crucial for time series training. In cases where initial feedback was missing, a neutral value of 3 was used to fill the gaps. (4) *Sequence Generation and Data Splitting*. Sequences of data were generated with a length of 15 timesteps to capture temporal patterns effectively. The target for regression (future IEQ values) was set to 1 timestep ahead of the input sequence. The dataset was split into training, validation, and test sets in a 60-20-20 ratio to ensure proper evaluation of the model.

4.3 Evaluation Metrics

To evaluate the proposed model, five key metrics were used: Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), R^2 , and accuracy. MSE, as shown in Equation (1), measures the average squared difference between the predicted and actual values, indicating how well the model predicts future IEQ conditions, with a lower MSE suggesting better performance. MAE, as shown in Equation (2), measures the average magnitude of the errors without considering their direction, providing a straightforward interpretation of prediction accuracy. Compared to MSE, MAE is less sensitive to outliers, with lower MAE indicating better performance. RMSE, shown in Equation (3), is the square root of MSE and provides an interpretable error metric in the same units as the target variable. R^2 , as shown in Equation (4), represents the proportion of the variance in the dependent variable that is predictable from the independent variables, with a value closer to 1 indicating a better fit.

$$MSE = \frac{1}{n} \sum_{i=1}^n (T_i - \hat{T}_i)^2 \quad (1)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |T_i - \hat{T}_i| \quad (2)$$

$$RMSE = \sqrt{MSE} \quad (3)$$

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (4)$$

Accuracy, as shown in Equation (5), is used for classification tasks to assess the proportion of correct predictions made by the model out of all predictions.

$$Accuracy = \frac{TP + TN}{\text{Total number of samples}} \quad (5)$$

4.5 Ablation Analysis

The proposed model was tested in Python 3 using an NVIDIA A100 GPU. Comparative analyses were conducted to evaluate the impact of model architecture and data fusion strategies on the prediction and classification of indoor environmental quality (IEQ). The analyses included two primary model configurations: (1) the proposed Bi-LSTM with Self-Attention, and (2) the standard LSTM with Self-Attention. These configurations were chosen to assess the ability to capture complex temporal dependencies and deliver accurate predictions. Additionally, the evaluation considered two different data fusion strategies: (1) Attention-Based Fusion, which utilized multi-head attention mechanisms to integrate IEQ, outdoor, and occupant feedback data, and (2) GAT-Based Fusion, which employed Graph Attention Networks to explicitly model relationships between data modalities.

The evaluation compared two key aspects: model architecture and fusion strategies. In terms of model architecture (Table 1), the proposed Bi-LSTM with Self-Attention demonstrated significantly better performance than the standard LSTM with Self-Attention. The Bi-LSTM achieved a 38.7% lower MSE, indicating its superior ability

to capture complex temporal dependencies in the data, which is crucial for accurate indoor environmental quality (IEQ) predictions. Both models maintained perfect classification accuracy for comfort and health levels, but the proposed Bi-LSTM consistently provided more precise regression results, reflecting its improved predictive capabilities.

Table 1. Performance Difference between Different Model Architectures

Model	MSE	% Difference in MSE	MAE	RMSE	R ²	Comfort Accuracy	Health Accuracy
Bi- LSTM with Self- Attention	0.1515	-	0.2239	0.3889	0.8093	1.0000	1.0000
LSTM with Self- Attention	0.2469	+38.7%	0.3569	0.4968	0.6424	1.0000	1.0000

For fusion strategies (Table 2), Attention-based Fusion outperformed GAT-Based Fusion by 19.3% in terms of MSE. The use of multi-head attention allowed the model to effectively integrate IEQ, outdoor conditions, and occupant feedback data, resulting in more accurate predictions. While GAT Fusion was less effective than Attention-Based Fusion, it still offered meaningful benefits over other methods by modeling the relationships between data modalities explicitly. However, the superior performance of Attention-Based Fusion highlighted its capability to better capture and utilize the diverse information available across different data sources. Despite the differences in MSE, both fusion strategies showed strong classification results, with Attention-Based Fusion maintaining a slight advantage in terms of overall predictive accuracy.

Table 2. Fusion Method Comparison

Fusion Strategy	MSE	% Difference in MSE)	MAE	RMSE	R ²	Comfort Accuracy	Health Accuracy
Attention- Based Fusion	0.1515	-	0.2239	0.3889	0.8093	1.0000	1.0000
GAT- Based Fusion	0.1878	+19.3%	0.2908	0.4330	0.7483	0.2381	1.0000

The results showed that the proposed Bi-LSTM with Self-Attention and Attention-Based Fusion consistently delivered superior performance, achieving lower prediction errors and more accurate integration of data compared to the other models and fusion strategies. This confirms the effectiveness of the proposed approach in capturing temporal dependencies and utilizing diverse information for predicting and classifying IEQ conditions.

4.6 Predicted vs. Actual Values Plot

Figure 2 compares actual and predicted values for four indoor environmental quality (IEQ) parameters: RADON, CO₂, PM_{2.5}, and Humidity. The plots show that the predicted values closely follow the general trend of the actual data, indicating the model's ability to track fluctuations in the IEQ parameters. However, some deviations are observed, suggesting areas where prediction accuracy could be further refined. Overall, the model demonstrates its capacity to capture key patterns in the data across different IEQ variables.

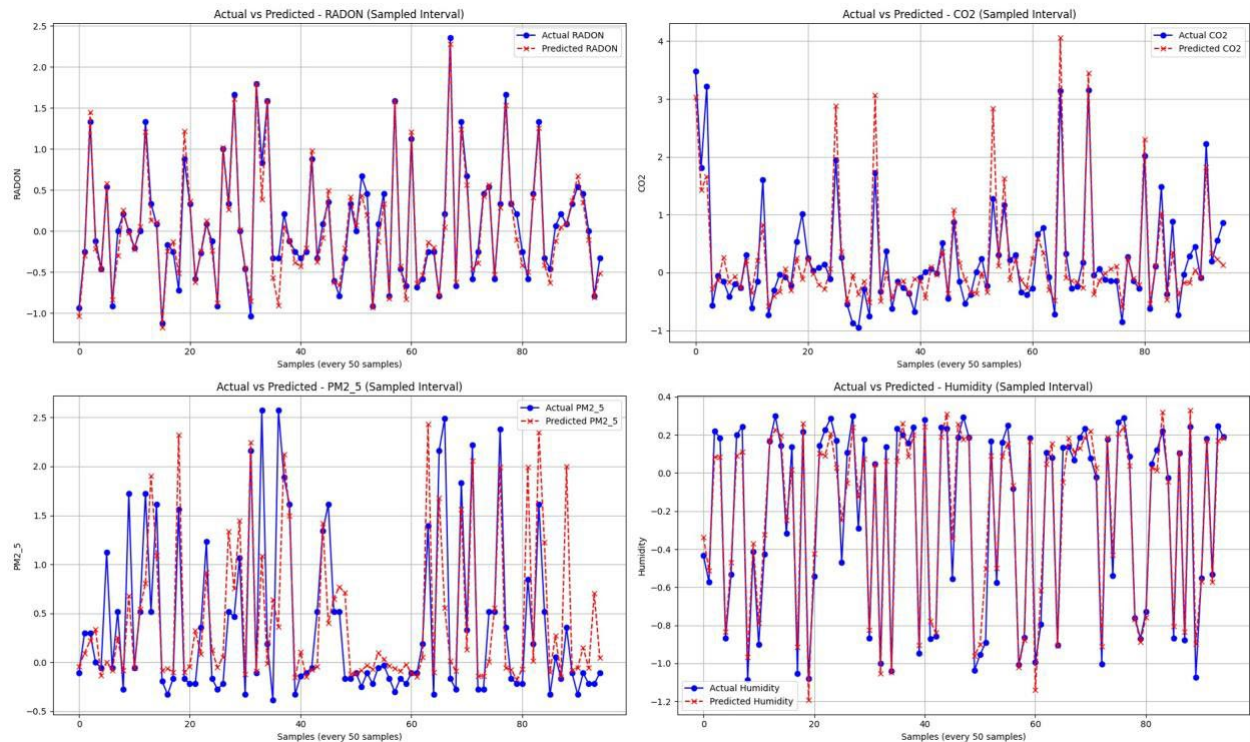


Figure 2. Predicted vs. Actual Values

5. CONCLUSION

In conclusion, this paper presents a novel Bi-Directional LSTM model with self-attention that integrates IEQ data, outdoor environmental data, and occupant feedback through attention-based fusion to predict and classify indoor environmental quality conditions. The proposed framework leverages the advantages of deep learning and multimodal data fusion to provide a comprehensive assessment of indoor environments. Results from the comparative analyses demonstrate that the proposed Bi-LSTM model with attention-based data fusion consistently outperforms the baseline models, highlighting the benefits of using advanced model architectures and fusion techniques. This research contributes to the growing body of work on IEQ assessment, offering valuable insights for improving occupant comfort, health, and overall indoor environment quality.

While the proposed framework's results are promising, several areas for future work remain. Additional data modalities, such as occupancy patterns, activity data, and human vital signs, could be incorporated to further enhance the model's capabilities. This will allow the model to capture the dynamic interactions between environmental conditions and occupant behavior more effectively, creating adaptive and personalized indoor environments. Another avenue for future research is to evaluate the scalability of the proposed framework across different building types, including residential, commercial, and mixed-use facilities, to assess its generalizability. Furthermore, integrating Model Predictive Control (MPC) could be explored to enable real-time, proactive adjustments of indoor environmental conditions based on the model's predictions. By employing MPC, the system would provide an intelligent, closed-loop control mechanism to maintain optimal IEQ conditions continuously.

REFERENCES

- Aldakheel, J. (2023). Indoor environmental quality evaluation of smart/artificial intelligence techniques in buildings – a review. *E3s Web of Conferences*, 396, 01101. <https://doi.org/10.1051/e3sconf/202339601101>
- Appau, W. (2023). Indoor environmental quality and energy use intensity: an empirical post-occupancy evaluation test of on-campus university student housing in ghana. *Property Management*, 42(3), 333-352. <https://doi.org/10.1108/pm-07-2023-0058>

- Beemer, C. J., Stearns-Yoder, K. A., Schuldt, S. J., Kinney, K. A., Lowry, C. A., Postolache, T. T., Brenner, L. A., & Hoisington, A. J. (2019). A brief review on the mental health for select elements of the built environment. *Indoor and Built Environment*, 1420326X1988965. <https://doi.org/10.1177/1420326x19889653>
- Chan-To-Hing, H., & Veeravalli, B. (2024). Fus-MAE: A cross-attention-based data fusion approach for Masked Autoencoders in remote sensing. *ArXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2401.02764>
- Chenari, B., Lamas, F. B., Gaspar, A. R., & Silva, M. G. d. (2017). Development of a new co2-based demand-controlled ventilation strategy using energyplus. <https://doi.org/10.31224/osf.io/h6mdn>
- Cincinelli, A., & Martellini, T. (2017). Indoor Air Quality and Health. *International Journal of Environmental Research and Public Health*, 14(11), 1286. <https://doi.org/10.3390/ijerph14111286>
- Cipollone, V., Morresi, N., Serroni, S., Casaccia, S., Giovanardi, M., Pracucci, A., Arnone, D., & Revel, G. M. (2024). AI-Based Methodology for Thermal Comfort Measurement: Application of a Simplified Comfort Model on a Real-Life Case Study. *2024 IEEE International Workshop on Metrology for Living Environment (MetroLivEnv)*, 11, 286–291. <https://doi.org/10.1109/metrolivenv60384.2024.10615461>
- Crețescu, I., Isopescu, D. N., Lutić, D., & Soreanu, G. (2019). Indoor air pollutants and the future perspectives for living space design. *Indoor Environment and Health*. <https://doi.org/10.5772/intechopen.87309>
- Deng, Z., Dong, B., Guo, X., & Zhang, A. (2023). A pilot study on the combined multi-domain impact of indoor air quality and noise on office productivity. *E3s Web of Conferences*, 396, 01047. <https://doi.org/10.1051/e3sconf/202339601047>
- Du, Y., Zandi, H., Kotevska, O., Kurte, K., Munk, J., Amasyali, K., McKee, E., & Li, F. (2021). Intelligent multi-zone residential HVAC control strategy based on deep reinforcement learning. *Applied Energy*, 281, 116117. <https://doi.org/10.1016/j.apenergy.2020.116117>
- EPA. 2023. "Indoor air quality." Accessed October 12, 2024. <https://www.epa.gov/report-environment/indoor-air-quality>
- Fang, Z., Crimier, N., Scanu, L., Midelet, A., Alyafi, A., & Delinchant, B. (2021). Multi-zone indoor temperature prediction with LSTM-based sequence to sequence model. *Energy and Buildings*, 245, 111053. <https://doi.org/10.1016/j.enbuild.2021.111053>
- Fritz, H., Tang, M., Kinney, K., & Nagy, Z. (2022). Evaluating machine learning models to classify occupants' perceptions of their indoor environment and sleep quality from indoor air quality. *Journal of the Air & Waste Management Association*. <https://doi.org/10.1080/10962247.2022.2105439>
- Gao, J., Li, P., Chen, Z., & Zhang, J. (2020). A survey on deep learning for multimodal data fusion. *Neural Computation*, 32(5), 829-864. https://doi.org/10.1162/neco_a_01273
- Jones, A. P. (1999). Indoor air quality and health. *Atmospheric Environment*, 33(28), 4535–4564. [https://doi.org/10.1016/s1352-2310\(99\)00272-1](https://doi.org/10.1016/s1352-2310(99)00272-1)
- Kapoor, N. R., Kumar, A., Kumar, A., Kumar, A., Mohammed, M. A., Kumar, K., Kadry, S., & Lim, S. (2022). Machine Learning-Based CO2 Prediction for Office Room: A Pilot Study. *Wireless Communications and Mobile Computing*, 2022, 1–16. <https://doi.org/10.1155/2022/9404807>
- Lahat, D., Adali, T., & Jutten, C. (2015). Multimodal Data Fusion: An Overview of Methods, Challenges, and Prospects. *Proceedings of the IEEE*, 103(9), 1449–1477. <https://doi.org/10.1109/jproc.2015.2460697>
- Lahat, D., Adali, T., & Jutten, C. (2015). Multimodal Data Fusion: An Overview of Methods, Challenges, and Prospects. *Proceedings of the IEEE*, 103(9), 1449–1477. <https://doi.org/10.1109/jproc.2015.2460697>
- Lee, M. J., & Zhang, R. (2024). Human-Centric Artificial Intelligence of Things-Based Indoor Environment Quality Modeling Framework for Supporting Student Well-Being in Educational Facilities. *Journal of Computing in Civil Engineering*, 38(2). <https://doi.org/10.1061/jccee5.cpeng-5632>
- Lin, X., Chao, S., Yan, D., Guo, L., Liu, Y., & Li, L. (2023). Multi-Sensor Data Fusion Method Based on Self-Attention Mechanism. *Applied Sciences*, 13(21), 11992. <https://doi.org/10.3390/app132111992>
- Liu, J., Capurro, D., Nguyen, A., & Verspoor, K. (2023). Attention-based multimodal fusion with contrast for robust clinical prediction in the face of missing modalities. *Journal of Biomedical Informatics*, 145, 104466–104466. <https://doi.org/10.1016/j.jbi.2023.104466>
- Liu, X., Ren, M., Yang, Z., Yan, G., Guo, Y., Cheng, L., & Wu, C. (2022). A multi-step predictive deep reinforcement learning algorithm for HVAC control systems in smart buildings. *Energy*, 124857. <https://doi.org/10.1016/j.energy.2022.124857>
- Lolli, F., Marinello, S., Coruzzolo, A., & Butturi, M. (2022). Post-occupancy evaluation's (poe) applications for improving indoor environment quality (ieq). *Toxics*, 10(10), 626. <https://doi.org/10.3390/toxics10100626>
- Meng, T., Jing, X., Yan, Z., & Pedrycz, W. (2020). A survey on machine learning for data fusion. *Information Fusion*, 57, 115–129. <https://doi.org/10.1016/j.inffus.2019.12.001>

- Mitkov, R., Petrova-Antonova, D., & Hristov, P. O. (2023). Predictive Modeling of Indoor Environmental Parameters for Assessing Comfort Conditions in a Kindergarten Setting. *Toxics*, 11(8), 709–709. <https://doi.org/10.3390/toxics11080709>
- Rizk, H., Torki, M., & Youssef, M. (2019). Cellindeep: robust and accurate cellular-based indoor localization via deep learning. *Ieee Sensors Journal*, 19(6), 2305–2312. <https://doi.org/10.1109/jsen.2018.2885958>
- Samet, J. M., & Spengler, J. D. (2003). Indoor Environments and Health: Moving Into the 21st Century. *American Journal of Public Health*, 93(9), 1489–1493. <https://doi.org/10.2105/ajph.93.9.1489>
- Shamsudin, S. (2023). Indoor air quality assessment in the office of the transformer manufacturing factory in selangor, malaysia. *Malaysian Journal of Medicine and Health Sciences*, 19(5), 24–31. <https://doi.org/10.47836/mjmhs.19.5.5>
- Shan, X., Melina, A. N., & Yang, E.-H. (2018). Impact of indoor environmental quality on students' wellbeing and performance in educational building through life cycle costing perspective. *Journal of Cleaner Production*, 204, 298–309. <https://doi.org/10.1016/j.jclepro.2018.09.002>
- Smarra, F., Jain, A., de Rubeis, T., Ambrosini, D., D'Innocenzo, A., & Mangharam, R. (2018). Data-driven model predictive control using random forests for building energy optimization and climate control. *Applied Energy*, 226, 1252–1272. <https://doi.org/10.1016/j.apenergy.2018.02.126>
- Sun, S., Zheng, X., Villalba-Díez, J., & Ordieres-Meré, J. (2019). Indoor air-quality data-monitoring system: long-term monitoring benefits. *Sensors*, 19(19), 4157. <https://doi.org/10.3390/s19194157>
- Tagliabue, L. C., Re Cecconi, F., Rinaldi, S., & Ciribini, A. L. C. (2021). Data driven indoor air quality prediction in educational facilities based on IoT network. *Energy and Buildings*, 236, 110782. <https://doi.org/10.1016/j.enbuild.2021.110782>
- Tien, P., Wei, S., Darkwa, J., Wood, C., & Calautit, J. (2022). Machine learning and deep learning methods for enhancing building energy efficiency and indoor environmental quality – a review. *Energy and Ai*, 10, 100198. <https://doi.org/10.1016/j.egyai.2022.100198>
- USGBC. (2014). Green Building 101: What is indoor environmental quality? | U.S. Green Building Council. [Usgbc.org. https://www.usgbc.org/articles/green-building-101-what-indoor-environmental-quality](https://www.usgbc.org/articles/green-building-101-what-indoor-environmental-quality)
- Vardoulakis, S., Giagloglou, E., Steinle, S., Davis, A., Sleenwenhoek, A., Galea, K., ... & Crawford, J. (2020). Indoor exposure to selected air pollutants in the home environment: a systematic review. *International Journal of Environmental Research and Public Health*, 17(23), 8972. <https://doi.org/10.3390/ijerph17238972>
- Xing, T., Sun, K., & Zhao, Q. (2023). MITP-Net: A deep-learning framework for short-term indoor temperature predictions in multi-zone buildings. 239, 110388–110388. <https://doi.org/10.1016/j.buildenv.2023.110388>
- Zahaba, M., Tamsi, N., Engliman, S., Kamaruddin, A., Hassan, N., & Ariffin, A. (2022). Indoor air quality (iaq) in a naval ship after refit program: a time variation analysis. *Iop Conference Series Earth and Environmental Science*, 1013(1), 012004. <https://doi.org/10.1088/1755-1315/1013/1/012004>
- Zhang, X., & Li, P. (2023). Transfer Learning in the Transformer Model for Thermal Comfort Prediction: A Case of Limited Data. *Energies*, 16(20), 7137–7137. <https://doi.org/10.3390/en16207137>
- Zhao, Y., Guo, S., Chen, Z., Shen, Q., Meng, Z., & Xu, H. (2022). Marfusion: An Attention-Based Multimodal Fusion Model for Human Activity Recognition in Real-World Scenarios. *Applied Sciences*, 12(11), 5408–5408. <https://doi.org/10.3390/app12115408>
- Zhu, Y., Al-Ahmed, S. A., Shakir, M. Z., & Olszewska, J. I. (2022). LSTM-Based IoT-Enabled CO2 Steady-State Forecasting for Indoor Air Quality Monitoring. *Electronics*, 12(1), 107. <https://doi.org/10.3390/electronics12010107>

INTEGRATED FRAMEWORK FOR COMPLIANCE CHECKING AND PERFORMANCE EVALUATION IN BUILDING DESIGN

Jin Han¹, Zhe Zheng², Zhen Xu³, Xin-Zheng Lu⁴, and Jia-Rui Lin^{5,*}

1) Ph.D., Department of Civil Engineering, Tsinghua University, Beijing, China. Email: han-j23@mails.tsinghua.edu.cn

2) Ph.D., Department of Civil Engineering, Tsinghua University, Beijing, China. Email: zhengz97@qq.com

3) Ph.D., Prof., School of Civil and Resource Engineering, University of Science and Technology Beijing, Beijing, China. Email: xuzhen@ustb.edu.cn

4) Ph.D., Prof., Department of Civil Engineering, Tsinghua University, Beijing, China. Email: luxz@tsinghua.edu.cn

5) Ph.D., Assoc. Prof., Department of Civil Engineering, Tsinghua University, Beijing, China. Email: lin611@tsinghua.edu.cn

Abstract: Compliance checking and performance evaluation are crucial components of the iterative design and review process. However, traditional manual reviews and refined simulations are often subjective, time-consuming, and require a large number of input parameters. Therefore, automated rule checking (ARC) and performance simulation based on surrogate models have gained increasing attention. Despite this, most existing research focuses on one aspect or the other, without effectively integrating both, leading to gaps in reliability and efficiency when applied in practice. Therefore, this study proposes an integrated framework that combines automated compliance checking and efficient performance evaluation based on surrogate models, enabling rapid design review iterations. The framework comprises three interconnected modules: the NLP-AutoChecking module for rule checking based on IFC (Industry Foundation Classes) and semantic alignment; the DiffEvac module for evacuation performance simulation based on diffusion models; and a BIM design software module that connects both through a unified data interaction approach to modify design. Specifically, the BIM design can be converted into IFC format for rule checking to identify non-compliant elements and specify the violated regulations. Designers can then modify the design within the BIM software accordingly. It is important to note that multiple solutions may meet regulatory requirements, but not all are scientifically or practically optimal, as some may compromise safety or increase costs. Therefore, the modified BIM designs are exported as floor plans, cleaned and annotated, and then fed into the surrogate model for performance simulation, which evaluates and selects the optimal solution from the available options. This iterative cycle of compliance checking, simulation, and design modification continues until the design meets all regulatory standards and achieves optimal performance. Case studies demonstrate that this framework enables quick iterations and adjustments throughout both the design and review stages, significantly improving design quality and offering strong potential for widespread practical adoption.

Keywords: Building design, Automated rule checking (ARC), Evacuation performance simulation, Efficient simulation

1. INTRODUCTION

The design process is crucial for ensuring the quality, safety, and functionality of buildings, often requiring multiple iterations of review and modifications to achieve optimal building performance. During these reviews, compliance checking and performance evaluation are essential components. Compliance checking ensures the design meets regulatory requirements, while performance evaluation assesses the building's functionality under various scenarios, typically through simulation, to optimize the design. Traditional manual reviews, however, are often subjective, time-consuming, and costly, prompting extensive research into ARC and performance simulation based on surrogate models.

To enhance the automation of the review process, work on ARC can be traced back to the 1960s, when Fenves developed a tabular decision system for structural design review (Fenves, 1966). Building on this, Rasdorf and Lakmazaheri, employed predicate logic to formalize codes and regulations, enhancing the SASE (Standards Analysis, Synthesis, and Expression) system's ability to process standards (Rasdorf & Lakmazaheri, 1990). These early ARC studies primarily focused on 2D drawing reviews (Eastman et al., 2009). Since the 2000s, advancements in BIM models and IFC have significantly improved the capacity to describe design information, driving further ARC development (Ismail et al., 2017). For instance, Tan developed an automated code compliance checking system for building envelope design based on BIM and tabular decision systems (Tan et al., 2010). However, these studies often depend on manually created and maintained rule libraries, which pose challenges in flexibility and maintainability (Eastman et al., 2009; Nawari, 2019). Therefore, many scholars have proposed semiautomated and automated methods for interpreting the regulation text into a computer-processable format. One of the typical semi-automated rule interpretation methods is RASE (Requirement, Applicabilities, Selection, Exceptions) (Hjelseth & Nisbet, 2011), which helps experts analyze the semantic structure of regulatory requirements and uses document annotation techniques to mark different components of the regulations. Based on this, Beach developed a parser for the RASE method using ontology technology, reducing the manual effort required for semantic alignment and code generation in semiautomated methods (Beach et al., 2015). However, these methods still function at a coarse-grained level and rely heavily on manual input, leading to a relatively low level of automation.

For this reason, natural language processing (NLP), widely used for analyzing and interpreting human language in text form (Fuchs, 2021), has been applied to automate rule interpretation from regulatory documents. Zhang and El-Gohary (Zhang & El-Gohary 2016) introduced a fully automated tagging and conversion method, developing a rule-based BIM extraction tool to align BIM objects with regulatory instances, enabling automated rule interpretation and checking. Zheng (Zheng et al., 2022) developed a large-scale domain corpus and pre-trained domain-specific language models for the AEC domain to assist in automated rule interpretation. Zhou (Zhou et al., 2022) introduced an innovative automatic rule extraction method using a deep learning model combined with context-free grammars (CFGs), which can interpret regulatory rules into pseudocode formats without manual intervention, increasing generality, accuracy, and interpretability. While these studies advanced automated rule interpretation, challenges remain: (1) High costs and low generalizability of regular expression-based methods (Tomassetti G, 2017); (2) Semantic misalignment between regulatory texts and BIM models due to differing terminologies (Zhou & El-Gohary, 2021); (3) Current methods rely on simple logic clauses, which are limited by missing implicit information (Zhang et al., 2018). To address these issues, we proposed NLP-AutoChecking in our previous research, a novel automated rule interpretation and design review method that automatically interprets the domain regulatory texts expressed in natural language into computer-processable code based on a pipeline of NLP techniques (Zheng et al., 2022; Zheng et al., 2023).

Design regulations traditionally provide clear parameters and requirements to ensure building safety. However, due to variations in structure, space usage, and layout, designing strictly according to regulatory requirements may not always result in the most scientifically sound or cost-effective solutions, potentially leading to suboptimal safety levels or increased costs (Meacham & Custer, 1995). As a predictive and analytical tool, evacuation simulation facilitates the understanding of complex evacuation behaviors, enhances building design, and refines emergency plans. Consequently, it has garnered significant attention from both academic researchers and industry practitioners. Evacuation simulations can be broadly classified into macro-level and micro-level models, based on whether they emphasize overall crowd dynamics or individual interactions. Early models, such as fluid mechanics, Cellular Automata (CA), and social force models (Klote & Hadjisophocleous, 2008; Helbing et al., 2000; Fu et al., 2015; Gao et al., 2022), provided a basic framework for simulating macroscopic pedestrian movements. However, these models often rely on generalized crowd behaviors, leading to oversimplifications and a lack of accuracy in capturing the finer details of evacuation scenarios. Recent advances in computational capabilities and data accessibility have shifted the focus towards micro-level simulations, encouraging the development of more detailed and nuanced models, such as agent-based, virtual reality (VR), and multi-level simulations (Cotfas et al., 2022; Wang et al., 2023; Alac et al., 2023; Lim et al., 2023). These newer approaches simulate each evacuee as an independent agent, capturing individual behaviors and interactions (Senanayake et al., 2024). While these models offer higher precision in representing evacuation processes, they also require extensive input data and complex modeling, making them time-consuming and labor-intensive. As a result, researchers have turned to more efficient evacuation simulation methods, leveraging artificial intelligence (AI) technologies capable of learning from existing data and knowledge as the foundation for these advancements. For instance, Nourkojouri (Nourkojouri et al., 2023) explored the use of image generation algorithms for rapid evacuation simulations and evaluations, focusing primarily on Conditional GANs (Generative Adversarial Networks) in their initial experiments. However, Conditional GANs have limitations in generalization and image generation detail compared to more advanced models like diffusion models (Gu et al., 2024). Furthermore, their study showed poor performance on irregular floor plans, highlighting the need to investigate alternative image generation models that could provide better quality, stability, and support for faster evaluations of various building layouts. To address these issues, we proposed DiffEvac in our previous research, a novel method to learn building evacuation patterns based on diffusion model, for efficient evacuation simulation and enhanced safety design (Han et al., 2024).

Although significant progress has been made in ARC and performance simulation, existing research often focuses on one aspect without effectively integrating both. This leads to discrepancies in data interfaces and processing logic, requiring designers to spend considerable time on manual adjustments and re-entry, which severely impacts efficiency. Therefore, this study proposes an integrated framework that combines automated compliance checking and efficient evacuation performance simulation based on surrogate models, ensuring that the final design not only meets all regulatory requirements but also achieves optimal functionality and safety.

2. FRAMEWORK

The proposed framework comprises three interconnected modules: the NLP-AutoChecking module for rule checking based on IFC and semantic alignment (Section 2.1); the DiffEvac module for evacuation performance simulation based on diffusion models (Section 2.2); and a BIM design software module that connects both to modify design (Section 2.3). Specifically, we propose a unified data interaction approach that makes various design software compatible with the NLP-AutoChecking and DiffEvac modules, allowing for rapid export of IFC formats for automated rule checking, as well as quick export, cleaning, annotation, and feature decoupling of floor plans for evacuation simulation. The relationship between the three modules is shown in Figure 1.

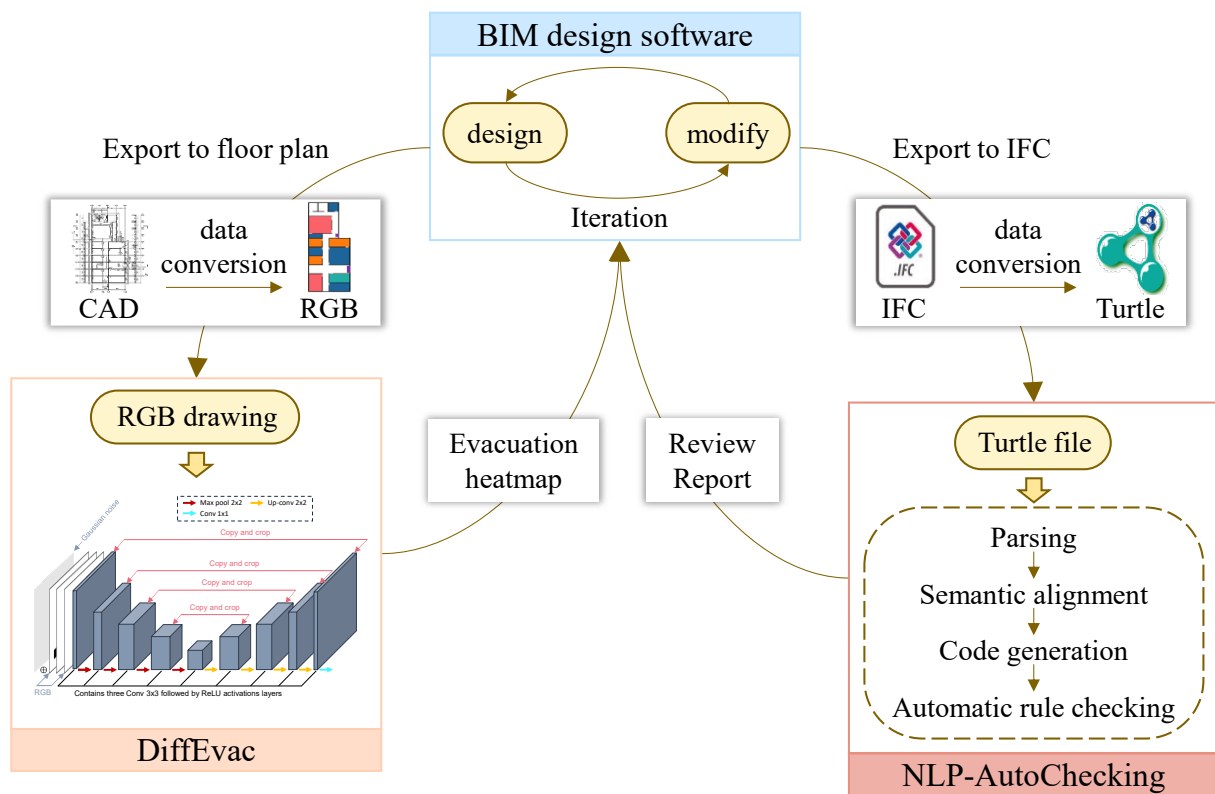


Figure 1. The relationship between the three modules

2.1 NLP-AutoChecking module for automated rule checking

This module is based on our previously proposed NLP-AutoChecking, a novel automated rule interpretation and design review method that automatically interprets the domain regulatory texts expressed in natural language into computer-processable code based on a pipeline of NLP techniques (Zheng et al., 2022; Zheng et al., 2023). To enhance the scalability of this module and enable rapid automated rule checking across various types of BIM software, we propose a unified data conversion method.

The IFC format is widely used for collaboration and model storage in BIM projects, with many BIM software programs supporting it (IFCwiki, 2018). However, since IFC uses the EXPRESS schema, it cannot be directly processed by the ontology-based reasoning engine in NLP-AutoChecking, requiring additional conversion. To address this, we developed a custom IFC-to-ontology mapping method to automatically convert BIM data stored in IFC format into Turtle (Terse RDF Triple Language) format. This process involves three key steps: concept mapping, IFC parsing, and generating the Turtle file.

Concept mapping aims to align IFC concepts and attributes with fire protection of building ontology (FPBO) classes, object properties, and data properties through the mapping table established in this study, as shown in Table 1. IFC object concepts are mapped to corresponding FPBO classes; for example, "IfcBuildingStory" maps to "BuildingStory." Similarly, IFC attributes are linked to FPBO data properties, such as "fire_resistance_rating" being mapped to "hasFireResistanceLimits_hour." The relationships between instances in FPBO are determined by three key IFC relationships: decomposition, contains, and BoundedBy. First, the domain and range objects of these relationships are identified by parsing the IFC file, then their corresponding FPBO classes are determined. A HashMap is used to define the object properties between two objects, using two keys: one for the domain object's class and the other for the range object's class.

Table 1. Part of the ontology mapping dictionary

IFC Schema	FPBO concepts
IfcBuildingStorey	BuildingStory
IfcBuilding	BuildingRegion
IfcSpace	BuildingSpace
IfcColumn	Column
IfcBeam	Beam
IfcSlab	Floor
IfcWallStandardCase	Wall

IfcWall	Wall
IfcWindow	Window
IfcDoor	Doors
IfcStair	Stairs
GlobalId	hasGlobalId
fire_resistance_rating (user defined attribute)	hasFireResistanceLimits hour

Several tools are available for parsing IFC formats, including IfcOpenShell (Python), IfcPlusPlus (C++), and the xBIM toolkit (.NET). Among these, IfcOpenShell, which is based on OpenCascade technology and operates in Python, is the most suitable for this study. Therefore, IfcOpenShell was selected to parse the IFC files (IfcOpenShell, 2023). During the parsing process, relevant objects and attributes are stored in memory and subsequently mapped to FPBO concepts using the concept mapping method described earlier.

Finally, in the Turtle file generation step, the mapped data stored in memory are formatted according to the Turtle syntax, a widely used format for ontology data exchange, and then outputted as a Turtle file.

With this approach, architects can efficiently export BIM models into a format compatible with the semantic parsing, semantic alignment, conflict resolution, and code generation processes in NLP-AutoChecking for automated rule checking. Specifically, a Turtle file is used as the input, and a file that documents non-compliant elements and specifies violated regulations is produced as the output, supporting design modifications. Additionally, the generated .ttl file allows for easy editing and further adjustments.

2.2 DiffEvac module for efficient evacuation performance simulation

This module is based on our previously proposed DiffEvac, a novel method to learn building evacuation patterns based on diffusion model, for efficient evacuation simulation and enhanced safety design (Han et al., 2024).









To ensure designs from various BIM software can be efficiently exported into a format suitable for evacuation performance simulation, this study introduces a unified data conversion method that includes four key steps: floor plan export, data cleaning, data annotation, and feature decoupling.

For floor plan export, each floor of the BIM model is individually exported as a floor plan in an editable CAD format. This process extracts the 2D layout of each level from the 3D BIM model, ensuring that the necessary spatial details are preserved for evacuation performance simulation.

Subsequently, irrelevant elements such as annotations, symbols, text, and redundant lines are removed, leaving only the essential structural features like walls, windows, and door openings (without indicating door swing directions). The drawings are then converted into RGB images, with each image adjusted to maximize the use of available space.

In addition to area capacity and exit locations, which can be reflected through pixel positions in building floor plans, crowd density is also a crucial parameter in evacuation simulations, requiring manual annotation. Therefore, based on the “Standard for Design of Office Building (JGJ/T67-2019)” (The Ministry of Housing and Urban-Rural Development of the People's Republic of China, 2019) and the “Code for Design of Library Buildings (JGJ 38-2015)” (The Ministry of Housing and Urban-Rural Development of the People's Republic of China, 2015), occupant densities for rooms of different functions were determined, as detailed in Table 2. To differentiate room occupancy density and function, we applied distinct colors, embedding crowd density information directly into the floor plans to ensure the model captures all key parameters. Afterward, the floor plans were annotated according to the color schemes outlined in Table 2, producing the functional layout drawings for each room.

Table 2 Occupant densities for rooms of different functions

Function of room	Density (m2/person)	Color
Ordinary office	6	
Meeting room (with table)	2	
Meeting room (no table)	1	
Exhibition Hall	1.43	
Other region	9	
Corridor & Restroom	0	
Exit (stairs)	0	
Exit (door)	0	

This method allows us to export various BIM models into a format compatible with the DiffEvac approach, using an RGB drawing as input. The surrogate models in DiffEvac can then be applied for rapid evacuation performance simulation, generating evacuation heatmaps to effectively guide design modifications.

2.3 BIM design software module for design modification

Using the unified data conversion method described in Sections 2.1 and 2.2, we can achieve rapid automated rule checking and evacuation performance simulation during the design and review process for various BIM models.

The process begins by converting the initial BIM design into IFC (Industry Foundation Classes) format, which is then input into the rule checking module to identify non-compliant elements and specify the violated regulations. Designers can use this information to modify the design within the BIM model. It is important to note that multiple modification solutions may meet regulatory requirements. However, due to differences in building structure, space usage, and other factors, designing strictly according to standardized code parameters may not always result in the most scientifically sound or cost-effective solution, potentially leading to suboptimal safety levels or increased costs. Therefore, we enable the modified BIM designs to be quickly exported as floor plans, cleaned, annotated, and then fed into the surrogate model for performance simulation. This step evaluates and selects the optimal solution from among the available options. This iterative cycle of compliance checking, simulation, and design modification continues until the design meets all regulatory standards and achieves optimal performance, as described in Figure 1.

3. CASE STUDY

This section demonstrates the practical application of the proposed integrated framework in the design and review process, using a case study of fire protection checking for a three-story office building. The workflow for optimizing design plans during the design and review process using the proposed framework is shown in Figure 2.

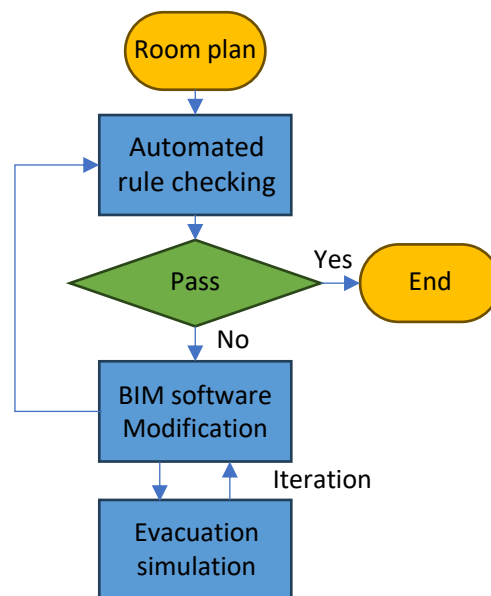


Figure 2. The workflow for optimizing design plans using the proposed framework

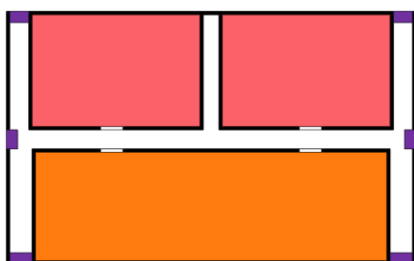
This building has a first-level fire resistance rating, spanning approximately 2500 m², and features a high-occupancy conference hall. In this case, Revit is used as the BIM software, and the regulation checked by the NLP-AutoChecking module is: "For crowded places such as conference halls and multi-purpose halls in buildings, there should be no less than two evacuation doors in a hall or room, and the building area should not exceed 400m²." (GB 50016-2014). The corresponding model is shown in Figure 3.



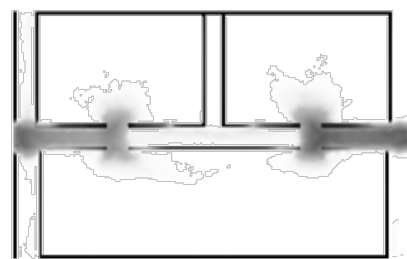
Figure 3. Three-dimensional model of the three-story office building

After the architect completes the preliminary design, the BIM model can be exported as a Turtle file following the process described in [Section 2.1](#), and then input into the previously developed NLP-AutoChecking method for rapid automatic rule checking. This process identifies that the number of evacuation doors in two rooms was less than the required minimum of two, as stipulated by the code. Since the model did not pass the review, adjustments are needed for the non-compliant rooms. The original layout is shown in [Figure 4\(a\)](#), where the two red areas each have only one emergency exit, failing to meet the code requirements. Therefore, the number of emergency exits in these areas must be increased. However, further research is needed to determine where in the room to add a new emergency exit to enhance the evacuation performance. [Figure 4](#) presents three modified room layouts. Option 1, shown in [Figure 4\(c\)](#), adds both evacuation exits to the front. Option 2, shown in [Figure 4\(e\)](#), places one evacuation exit at the front and the other at the side. Option 3, shown in [Figure 4\(g\)](#), places both evacuation exits at the side.

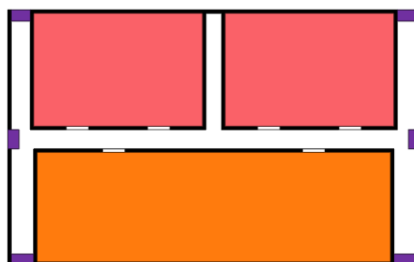
It can be seen that all three proposed modifications meet the regulatory requirements. To select the option with the best evacuation performance, we used the data processing method described in [Section 2.2](#) to convert these BIM models into three-channel input with decoupled features. These were then fed into the DiffEvac method proposed in our previous research, generating corresponding evacuation heatmaps to evaluate the performance of each plan, as shown in [Figures 4\(b\), 4\(d\), 4\(f\), and 4\(h\)](#). [Figure 4\(b\)](#) illustrates that the original room layout leads to inadequate use of exits located at the top and bottom of the building, leading to severe congestion in the central corridor, which significantly impairs evacuation efficiency. [Figure 4\(d\)](#) shows that the design option with additional evacuation exits at the front of the two rooms ([Figure 4\(c\)](#)) also fails to fully utilize the exits at the top and bottom of the building, similarly causing large-scale congestion in the central corridor. Even though this design meets code requirements, the additional exits do not substantially enhance evacuation efficiency. When an evacuation exit is added to the side of the room, as shown in [Figure 4\(f\)](#), some people evacuate through the upper exit, significantly alleviating the crowding in the central corridor. When both evacuation passages in the rooms with red backgrounds are located at the side ([Figure 4\(g\)](#)), [Figure 4\(h\)](#) shows that the level of crowding is lower compared to the results in [Figure 4\(f\)](#), indicating that this layout is more effective for facilitating evacuation.



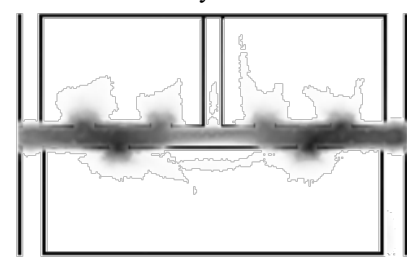
(a) Original room layout



(b) Evacuation heatmaps corresponding to the left layout



(c) Option 1 (both evacuation exits are placed at the front)



(d) Evacuation heatmaps corresponding to the left layout

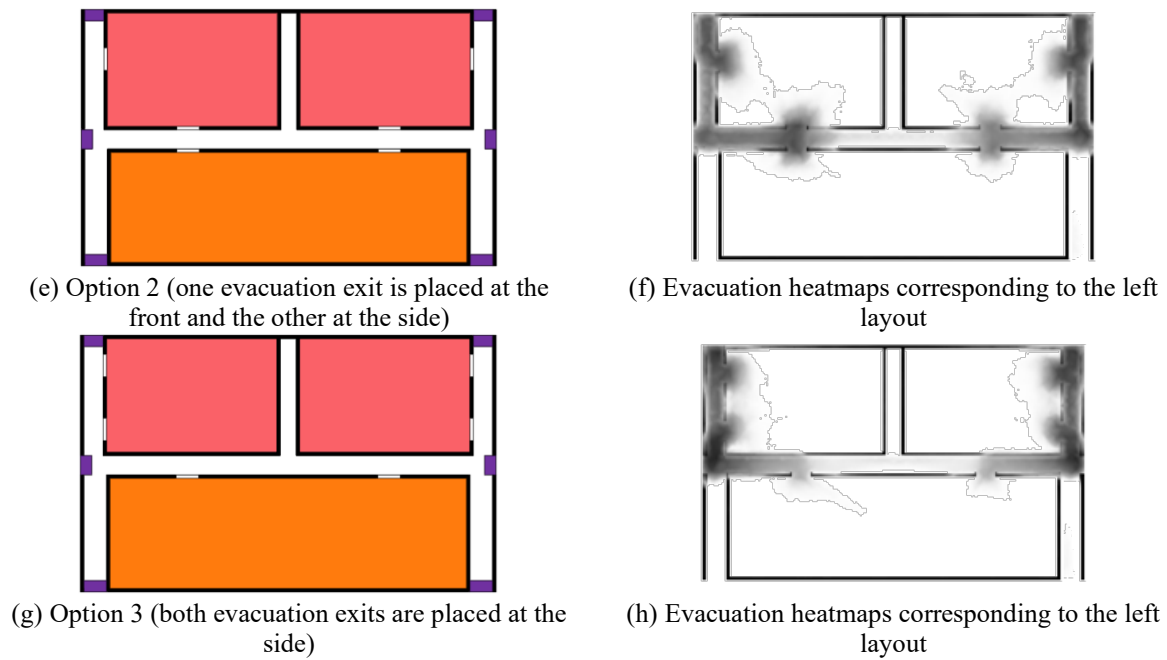


Figure 4. Room layouts and their corresponding evacuation heatmaps

To further analyze the evacuation effectiveness of different design options, this section utilized Pathfinder to develop the refined models and perform simulation analyses, with the evacuation times recorded in Figure 5. It is observed that while placing both evacuation exits at the front meets regulatory requirements, it does not result in a substantial reduction in evacuation time. In contrast, relocating the evacuation exits to the sides of the rooms alleviates crowd congestion and results in a considerable decrease in evacuation time.

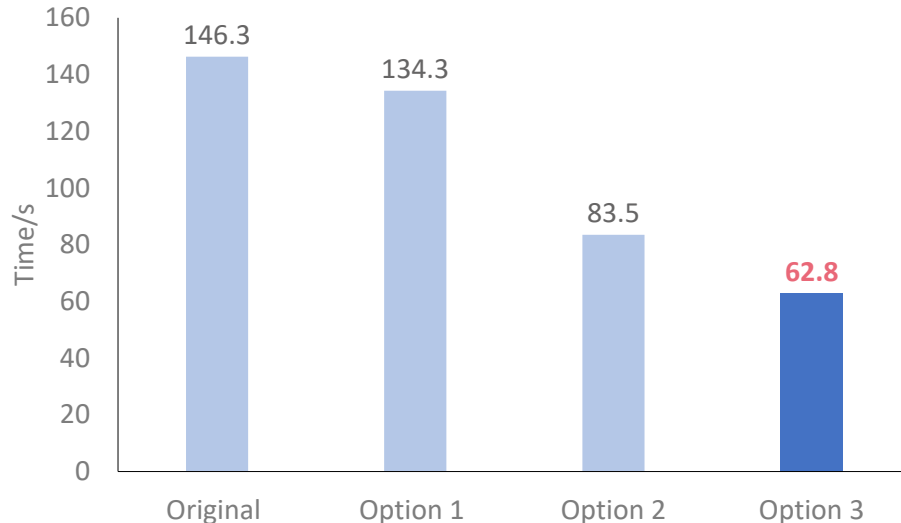


Figure 5. Evacuation time required for different layout plans

This case study highlights the practical value of the proposed integrated framework, for revising design plans following automatic rule checking and efficient evacuation performance simulation, demonstrating its effectiveness in aiding users with design optimization and validation. While this example focuses on a single regulation, real-world applications often involve multiple violations, and modifications may conflict with other regulations. Therefore, the process should be iterative to ensure the design complies with all regulatory standards and achieves optimal performance.

4. CONCLUSIONS

This study proposes an integrated framework that combines automated compliance checking and efficient

performance evaluation based on surrogate models, enabling rapid design review iterations. Specifically, we propose a unified data interaction approach that makes various design software compatible with the NLP-AutoChecking and DiffEvac modules, allowing for rapid export of IFC formats for automated rule checking, as well as quick export, cleaning, annotation, and feature decoupling of floor plans for evacuation simulation.

Case study demonstrate that this framework enables quick iterations and adjustments throughout both the design and review stages, significantly improving design quality and offering strong potential for widespread practical adoption.

It is important to note that the data processing framework proposed in this study is not limited to evacuation performance evaluation but can be extended to a wider range of applications. For example, by modifying the surrogate models, this framework can be expanded to assess other building performance criteria. Moreover, the data processing flow can also be adapted to support text-guided design modifications. Specifically, engineers' textual design requirements would be processed into a language understandable by computers, while existing design drawings would be transformed into feature tensors through image processing. To enable this, adjustments would need to be made to the deep learning models within the framework, ensuring that they are capable of effectively handling both textual input and image-based design features, thus facilitating automatic design modifications driven by textual descriptions.

ACKNOWLEDGMENTS

The authors are grateful for the financial support received from the National Natural Science Foundation of China (Nos. 52238011, 52378306).

REFERENCES

- Alac, R., Hammad, A. W., Hadigheh, A., & Opdyke, A. (2023). Optimising egress location in school buildings using mathematical modelling and agent-based simulation. *Safety science*, 167, 106265.
- Beach, T. H., Rezgui, Y., Li, H., & Kasim, T. (2015). A rule-based semantic approach for automated regulatory compliance in the construction sector. *Expert Systems with Applications*, 42(12), 5219-5231.
- Cotfas, L. A., Delcea, C., Iancu, L. D., Ioanăș, C., & Ponsiglione, C. (2022). Large event halls evacuation using an agent-based modeling approach. *IEEE Access*, 10, 49359-49384.
- Eastman, C., Lee, J. M., Jeong, Y. S., & Lee, J. K. (2009). Automatic rule-based checking of building designs. *Automation in construction*, 18(8), 1011-1033.
- Fenves, S. J. (1966). Tabular decision logic for structural design. *Journal of the Structural Division*, 92(6), 473-490.
- Fuchs, S. (2021). Natural language processing for building code interpretation: systematic literature review report. *August*, 21, 2023.
- Fu, Z., Zhou, X., Zhu, K., Chen, Y., Zhuang, Y., Hu, Y., ... & Li, J. (2015). A floor field cellular automaton for crowd evacuation considering different walking abilities. *Physica A: Statistical Mechanics and its Applications*, 420, 294-303.
- Gao, D. L., Lee, E. W. M., & Lee, Y. Y. (2022). Integration of cumulative prospect theory in cellular automata model for building evacuation. *International Journal of Disaster Risk Reduction*, 74, 102904.
- Gu, Y., Huang, Y., Liao, W., & Lu, X. (2024). Intelligent design of shear wall layout based on diffusion models. *Computer-Aided Civil and Infrastructure Engineering*.
- Han, J., Lu, X., Gu, Y., Liao, W., Cai, Q., & Xue, H. (2024). Optimized data representation and understanding method for the intelligent design of shear wall structures. *Engineering Structures*, 315, 118500.
- Han, J., Zheng, Z., Gu, Y., Lu, X. Z., & Lin, J. R. (2024). Learning and simulating building evacuation patterns for enhanced safety design using generative models. *Computers in Industry*. (under review)
- Helbing, D., Farkas, I., & Vicsek, T. (2000). Simulating dynamical features of escape panic. *Nature*, 407(6803), 487-490.
- Hjelseth, E., & Nisbet, N. (2011, October). Capturing normative constraints by use of the semantic mark-up RASE methodology. In *Proceedings of CIB W78-W102 conference* (pp. 1-10).
- IfcOpenShell. (2023). *IfcOpenShell 0.7.0 documentation*. Retrieved from blenderbim website: <https://blenderbim.org/docs-python/>.
- IFCwiki. (2018). *Open source projects supporting IFC*. Retrieved from ifcwiki website: http://www.ifcwiki.org/index.php/Open_Source.
- Ismail, A. S., Ali, K. N., & Iahad, N. A. (2017, July). A review on BIM-based automated code compliance checking system. In *2017 international conference on research and innovation in information systems (icriis)* (pp. 1-6). IEEE.
- Klote, John H.D.Sc, P.E., & Hadjisophocleous, George,PhD., P.Eng. (2008). An Overview of Evacuation Analysis with Application to Smoke Control Systems. *ASHRAE Transactions*, 114, 143-150.
- Lim, H., Lee, H., & Hwang, J. H. (2023). Multi-Agent Simulation on Staff Evacuation Behavior in Elderly

- Nursing Home Fire Emergencies. *Buildings*, 13(2), 400.
- Meacham, B. J., & Custer, R. L. (1995). Performance-based fire safety engineering: an introduction of basic concepts. *Journal of Fire Protection Engineering*, 7(2), 35-53.
- Nawari, N. O. (2019). A generalized adaptive framework (GAF) for automating code compliance checking. *Buildings*, 9(4), 86.
- Nourkojouri, H., Nikkhah Dehnavi, A., Bahadori, S., & Tahsildoost, M. (2023). Early design stage evaluation of architectural factors in fire emergency evacuation of the buildings using Pix2Pix and explainable XGBoost model. *Journal of Building Performance Simulation*, 16(4), 415-433.
- Rasdorf, W. J., & Lakmazaheri, S. (1990). Logic-based approach for modeling organization of design standards. *Journal of computing in civil engineering*, 4(2), 102-123.
- Senanayake, G. P., Kieu, M., Zou, Y., & Dirks, K. (2024). Agent-based simulation for pedestrian evacuation: A systematic literature review. *International Journal of Disaster Risk Reduction*, 104705.
- Tan, X., Hammad, A., & Fazio, P. (2010). Automated code compliance checking for building envelope design. *Journal of Computing in Civil Engineering*, 24(2), 203-211.
- The Ministry of Public Security of the People's Republic of China (2014). Code for fire protection design of buildings (GB 50016-2014). (in Chinese)
- The Ministry of Housing and Urban-Rural Development of the People's Republic of China (2015). Code for design of library buildings (JGJ 38-2015). (in Chinese)
- The Ministry of Housing and Urban-Rural Development of the People's Republic of China (2019). Standard for design of office building (JGJ/T67-2019). (in Chinese)
- Tomassetti, G. (2017). *The ANTLR mega tutorial*. Retrieved from tomassetti website: <https://tomassetti.me/antlr-mega-tutorial/>
- Wang, Z., Mao, Z., Li, Y., Yu, L., & Zou, L. (2023). VR-based fire evacuation in underground rail station considering staff's behaviors: model, system development and experiment. *Virtual Reality*, 27(2), 1145-1155.
- Zhang, C., Beetz, J., & de Vries, B. (2018). BimSPARQL: Domain-specific functional SPARQL extensions for querying RDF building data. *Semantic Web*, 9(6), 829-855.
- Zhang, J., & El-Gohary, N. M. (2016). Semantic NLP-based information extraction from construction regulatory documents for automated compliance checking. *Journal of Computing in Civil Engineering*, 30(2), 04015014.
- Zheng, Z., Chen, K. Y., Cao, X. Y., Lu, X. Z., & Lin, J. R. (2023). Llm-funcmapper: Function identification for interpreting complex clauses in building codes via llm. *arXiv preprint arXiv:2308.08728*.
- Zheng, Z., Lu, X. Z., Chen, K. Y., Zhou, Y. C., & Lin, J. R. (2022). Pretrained domain-specific language model for natural language processing tasks in the AEC domain. *Computers in Industry*, 142, 103733.
- Zheng, Z., Zhou, Y. C., Lu, X. Z., & Lin, J. R. (2022). Knowledge-informed semantic alignment and rule interpretation for automated compliance checking. *Automation in Construction*, 142, 104524.
- Zhou, P., & El-Gohary, N. (2021). Semantic information alignment of BIMs to computer-interpretable regulations using ontologies and deep learning. *Advanced Engineering Informatics*, 48, 101239.
- Zhou, Y. C., Zheng, Z., Lin, J. R., & Lu, X. Z. (2022). Integrating NLP and context-free grammar for complex rule interpretation towards automated compliance checking. *Computers in Industry*, 142, 103746.

VISUAL-DRIVEN INSPECTION FOR COLLECTING THE STATUS OF FIRE SAFETY EQUIPMENT STATUS IN BUILDING SPACES

Fangzhou Lin¹, Zhengyi Chen², Xiao Zhang³, Boyu Wang⁴, Jack C.P. Cheng⁵

1) Ph.D. Student, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: flinan@connect.ust.hk

2) Ph.D., Postdoctoral Fellow, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: zchenfq@connect.ust.hk

3) Ph.D. Candidate, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: xiao.zhang@connect.ust.hk

4) Ph.D., Postdoctoral Fellow, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: bwangbb@connect.ust.hk

5) Ph.D., Prof., Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: cejcheng@ust.hk

Abstract: Installed fire safety equipment within buildings plays a crucial role in ensuring the safety of personnel and minimizing losses. Nevertheless, if not maintained appropriately, these devices may fail to function optimally in emergency situations. As building sizes continue to grow, traditional manual inspection methods encounter significant challenges, including a heavy workload and complex information recording tasks. To tackle these issues, advanced emergency equipment detection frameworks and improvement plans have been put forward. This framework is specifically designed to overcome the problem of remote inspection being unable to accurately locate objects by establishing spatial relationships among devices, cameras, and trajectories. Firstly, the improved detection algorithm is utilized to detect objects of interest. Subsequently, these objects are located through a tracking algorithm and Visual Simultaneous Localization and Mapping (vSLAM). The on-site experimental results clearly show that the framework can effectively solve various types of equipment detection problems in a wide range of complex scenarios and holds great promise for replacing manual labor.

Keywords: Fire Safety Equipment, Visual Positioning, Object Tracking, Equipment Detection

1. INTRODUCTION

Emergency equipment, designed for crisis response, safeguards individuals during unforeseen events, reducing harm. For instance, fire extinguishers combat fires, exit signs guide evacuations, and alarms prompt action (Xin & Huang, 2013). Diverse buildings require tailored emergency setups, like fire extinguishers, lighting, and communication devices (Ivanov & Chow, 2022). Hospitals, for instance, need medical gear and backup power (Tang, Fitzgerald, Hou, & Wu, 2014). The state of emergency equipment is crucial for overall safety. Inadequate gear or outdated regulations in buildings impair emergency responses, emphasizing the need for well-maintained equipment (Dong, You, & Hu, 2014; Walters & Hastings, 1998). Regular checks are vital to ensure equipment integrity and functionality. Assessment of emergency equipment status occurs during installation, ensuring compliance with regulations, and during daily maintenance to prevent lapses in upkeep. Checks during installation guarantee regulatory adherence and adequate emergency preparedness, aiding construction progress tracking (Cao, Kamaruzzaman, & Aziz, 2022). In daily operations, maintenance gaps can lead to issues like inaccessible fire extinguishers or obstructed hose reels, impeding swift fire control (Xu, Chan, Leong, & Borondo, 2023). Routine inspections by authorities and community managers are essential to uphold equipment functionality (Guan, Fang, & Wang, 2018; Tse, 2002). Therefore, regular monitoring and maintenance of emergency equipment are vital for operational readiness, ensuring prompt and effective responses during crises.

During construction, delivery, daily operation, and maintenance phases, emergency equipment status checks rely on manual inspection at the current stage. However, manual checks are inefficient and costly. According to Hong Kong Fire Department guidelines (Lo, 1998), emergency equipment should be inspected annually by a registered contractor. As buildings grow larger, the number of fire safety facilities increases, straining manpower and time, potentially compromising inspection quality. This results in subpar maintenance practices (Kobes, Helsloot, de Vries, & Post, 2010). Intelligent inspection technologies (Spencer, Hoskere, & Narazaki, 2019), particularly computer vision, offer a solution by enhancing automation and data collection, potentially replacing manual labor. This technology enables automatic data recording and analysis, generating detailed reports for maintenance planning and equipment management. Ultimately, it streamlines inspections, boosts equipment reliability, and enhances maintenance efficiency.

Currently, emergency equipment management research primarily focuses on platform construction. For instance, Wang et al. (S.-H. Wang, Wang, Wang, & Shih, 2015) introduced a maintenance module for emergency equipment based on BIM. This module uses BIM's data storage to help maintenance personnel access fire safety equipment data swiftly. However, this data is manually verified, leading to concerns about accuracy. Vijayalakshmi et al. (Vijayalakshmi & Muruganand, 2017) applied IoT to monitor emergency facilities, enhancing

management in two stages: improving firefighting product quality and employing Radio Frequency Identification (RFID) for equipment tracking and deficiency identification. Li et al. (Li, Becerik-Gerber, Krishnamachari, & Soibelman, 2014) proposed an intelligent emergency response framework using metaheuristic algorithms for optimal solutions, integrating emergency equipment positions for decision-making. Existing research relies heavily on manual methods, leading to outdated information on equipment status, a major challenge in equipment management. On-site emergency facility management is still emerging, facing issues due to long life cycles and high costs. Neglected concerns include facility integrity, configuration, and hazards. Damaged or outdated firefighting facilities due to neglect pose risks during emergencies. Some buildings may lack adequate or properly placed fire protection equipment, impacting safety standards. Hidden dangers like obstructed fire extinguishers pose risks if not addressed promptly. The gap between emergency facility management platforms and actual equipment status highlights the need for intelligent frameworks for equipment status detection to enhance emergency response efficiency and reliability. Addressing these issues is crucial for ensuring the effectiveness and safety of emergency responses.

While computer vision technologies have found application in civil engineering inspections, they currently face limitations in effectively assessing the status of emergency equipment. Key challenges include the scattered placement of emergency devices throughout buildings, leading to difficulties in spatially locating targets during long-distance inspections. Additionally, dynamic perspectives in video detection hinder the accurate identification and counting of objects. Moreover, the varying sizes of emergency equipment, coupled with image quality issues in videos featuring different-sized targets, necessitate detectors with enhanced performance to address these complexities.

2. METHOD

This study presents an efficient automated inspection framework that utilizes cutting-edge computer vision and deep learning technologies for accurate positioning and identification of emergency equipment. The framework not only quickly detects the status of equipment but also compiles statistics on the equipment, facilitating enhanced management and maintenance tasks.

The proposed visual-based method, depicted in Figure 1, comprises three main components: an advanced object detection network, object ID allocation and tracking, and 3D mapping with detailed device information. Initially, the framework captures both RGB and depth images using a depth camera, which are used to create an inspection map by matching 2D and 3D features. The RGB images are processed through an enhanced detection network to identify emergency equipment, while depth images are converted into a point cloud using camera parameters to establish spatial relationships. Features within each equipment's bounding box are analyzed to assign unique IDs, enabling effective tracking. These devices are then mapped onto the inspection trajectory to generate a comprehensive device distribution map. This method facilitates precise detection and tracking of emergency equipment, essential for monitoring progress, managing emergencies, and performing maintenance tasks.

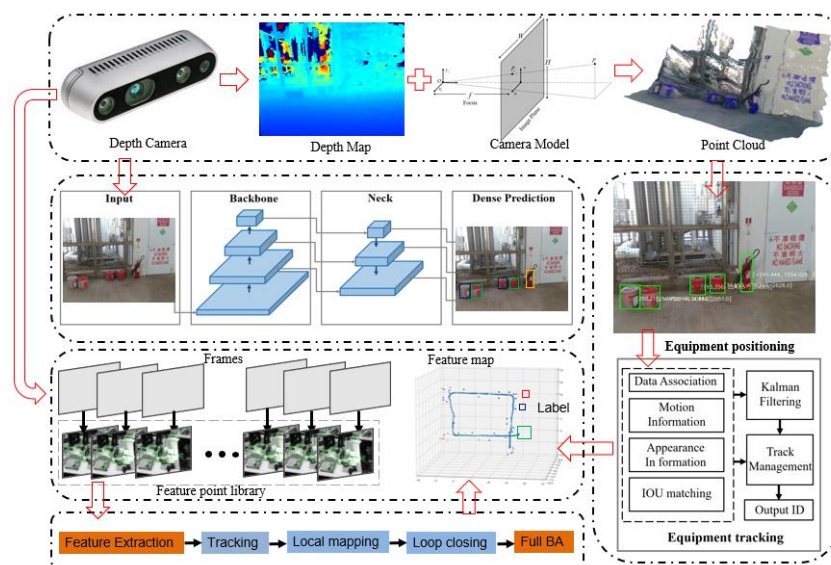


Figure 1. Fire safety equipment inspection framework

2.1 Improved Fire Safety Equipment Detection Network

Deep learning-based object detection algorithms like YOLOv5 are prevalent in industrial inspection but face challenges in detecting fire safety equipment due to issues like varying object sizes and image blurring from camera motion. To enhance performance, we propose an improved detection network that integrates C2F modules for advanced feature extraction, CPCA modules to accentuate and merge features, and DyHead to boost feature perception capabilities. This combination significantly improves detection accuracy and stability for device tracking in dynamic environments.

In the proposed network, the C2F module (Reis, Kupec, Hong, & Daoudi, 2023) is introduced instead of the C3 module. It combines three interlaced layers with 1×1 kernel and includes the DarkNet bottleneck. This module adds more skip connections, eliminates branch convolution operations, and adds splitting operations. This enriches feature information while reducing computational complexity. The C2F module optimizes the feature extraction network's information flow and ensures lightness with gradient diversion connections. It also draws on the ELAN module's ideas to optimize the network structure for easier training. By using feature vector diversion and multi-level nested convolution, it learns multi-scale features, expands the receptive field range, and improves the network's trainability and object detection performance. The attention mechanism highlights key features, removes background interference, and fuses features effectively. The proposed network uses the CPCA attention mechanism (Huang, Chen, Zou, Lu, & Chen, 2023) to address the network's limitations in handling different-scale, -shape, and -direction information. The CPCA attention mechanism consists of two sub-modules: the channel attention module and the spatial attention module. Compared to traditional ones, CPCA has innovative designs for better capturing and enhancing key information. The proposed network adds the CPCA attention mechanism before outputting three-sized feature maps on the backbone to enhance neck section feature fusion. The model reassigns weights to different-resolution feature maps, enhancing useful features and suppressing irrelevant ones. This makes the model focus on potential fire safety equipment areas. Adding CPCA before outputting feature maps in the backbone network can globally select and weight feature map channels and spatial positions. Thus, the model can interpret the entire image's contextual information, distinguish objects and backgrounds, improving detection accuracy. The backbone network's downsampling causes object information loss, especially for small objects. Video factors also degrade object quality. To enhance head module's feature perception, DyHead (Dai et al., 2021) with multiple self-attention mechanisms is adopted. After inputting the feature map into DyHead, it becomes a 3D tensor. Features are input to scale, spatial, and task perception attention modules. DyHead unifies these perceptions, enabling the network to focus on useful info. Its cascade attention mechanism allows handling multiple tasks. Three-scale feature maps are input to a unified branch and processed by DyHead.

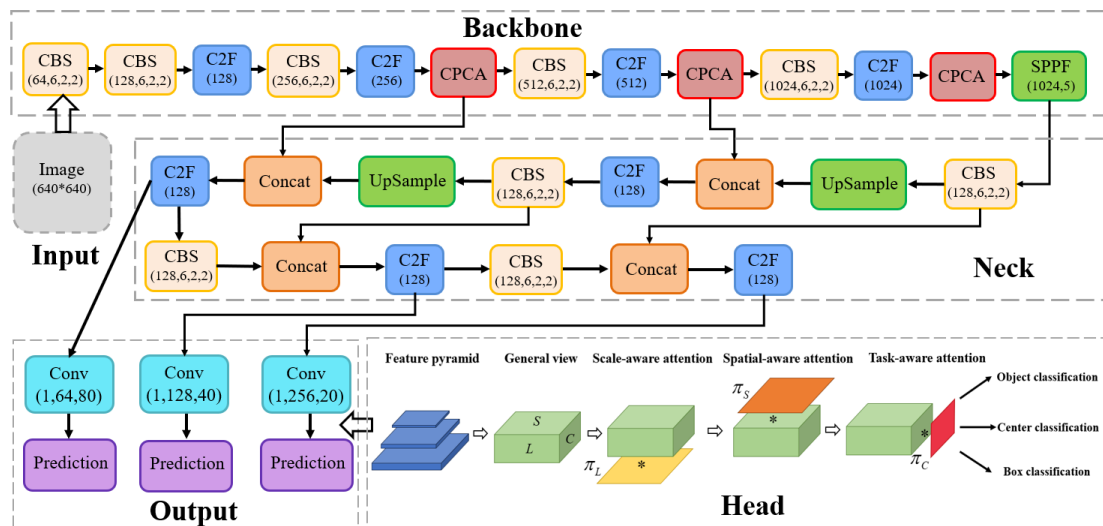


Figure 2. Enhanced architecture of the detection network

Object detection is a supervised learning task. The loss function measures the difference between predicted results and ground truth labels to guide model training. Minimizing the loss function helps object detection models learn more accurate bounding boxes and classification predictions, improving detection accuracy. For localization loss, IoU is commonly used to measure the overlap between predicted and actual bounding boxes. However, when IoU is used as a loss function, there are two issues: inability to distinguish bounding box shape differences and the problem of non-overlapping objects. For aspect ratio bounding boxes, they may have the same overlap but can't accurately surround objects. IoU can't distinguish them, resulting in the same value and inability to make reasonable weight adjustments. When predicting small objects, there may be no overlap between the ground truth and prediction box, with an IoU of 0 and a gradient of 0, making the model unable to optimize effectively.

The introduction of GIoU (Rezatofighi et al., 2019) helps alleviate the gradient problem caused by

variations in the shapes of bounding boxes. The concept of minimum closure region C is added on the basis of IoU. The minimum closure region is a minimum box C that can completely enclose the two bounding boxes A and B. By incorporating the area of the minimum closure region into the calculation, GIoU can better consider the position and shape differences between bounding boxes. In addition, to address the issue of small objects, the Normalized Wasserstein Distance (NWD) (Liu, Johns, & Davison, 2019; J. Wang, Xu, Yang, & Yu, 2021) is used instead of the degree of overlap to evaluate the prediction accuracy. For most objects, their bounding boxes are usually not strictly rectangular and may contain some background pixels. In these bounding boxes, the object is often concentrated at the center of the bounding box, while the background pixels are distributed at the boundaries. To better describe the weights of different pixels in the bounding box, the bounding box is regarded as a two-dimensional Gaussian distribution. In this model, the importance of pixels from the center of the bounding box to the boundary gradually decreases to distinguish the importance of the object and the background.

The loss of the proposed $Loss_{GN}$ function can be expressed as Eq.(1). Among them, λ_G and λ_N are dynamic parameters based on the size of gradients for different loss weights. A faster decrease in loss indicates a quicker learning speed. Nevertheless, for losses with higher learning speeds, the weight values should be controlled to strike a balance among various losses (He, Zhu, Wang, Savvides, & Zhang, 2019; Liu et al., 2019; Rezatofighi et al., 2019; J. Wang et al., 2021). The dynamic learning weight is defined as Eq. (2), which represents the loss ratio at time $t-1$ and time $t-2$. Next, Equation (3) is employed to normalize the learning speed through exponential processing to obtain the weights for each type of loss. Here, B is the equilibrium coefficient that serves to adjust the degree of difference in different loss weights. As the value of B rises, the disparity in weights is reduced, leading to a more even distribution of weights. A larger B implies a greater emphasis on balancing the weights among different components or tasks.

$$Loss_{GN} = \lambda_G Loss_G + \lambda_N Loss_N \quad (1)$$

$$\lambda_k(t) = \frac{\exp(w_k(t-1)/B)}{\exp(w_G(t-1)/B) + \exp(w_N(t-1)/B)} \quad (2)$$

$$w_k(t-1) = \frac{Loss_k(t-1)}{Loss_k(t-2)}, (k = N, G) \quad (3)$$

2.2 Equipment Spatial Positioning Based on vSLAM and Improved Visual Tracking

The DeepSort algorithm (Lu et al., 2021) is used to distinguish different fire safety equipment by tracking various objects in the video and assigning unique IDs to each object. The detector provides DeepSort with information such as location, size, and appearance features. DeepSort uses the kalman filtering algorithm to predict the position and state of objects in the current frame based on the parameters detected in the previous frame. It also uses deep neural networks to extract object RGB features. Cosine distance measures the similarity of appearance features, while mahalanobis distance calculates the similarity of motion features to form a cost matrix. Hungarian algorithm evaluation cost matrix. If the matching distance is less than the predefined threshold, it is considered that the IDs are the same, indicating a successful match. For mismatched trajectories, perform secondary matching based on IoU. If successful, use the kalman filter to update the object state. Determine whether to remove based on the trajectory that does not match the lifespan assessment. Unlike previous tracking tasks that used mobile cameras as inspection sensors, this results in more intense object motion in the image. The IoU matching strategy based on bounding box position similarity is prone to errors or omissions. Additional object depth information can improve tracking stability. Considering the uniform variation of distance between objects and cameras, we introduce a depth change rate based on depth cameras to optimize the data association evaluation method. When the matching distance is near the threshold, depth information can be used as an auxiliary judgment to improve the robustness of tracking accuracy.

Once the 2D bounding box and ID details of every fire safety equipment are obtained, the next step is to get the 3D position information of the equipment with respect to the camera. Since a depth camera is used, the depth value d of each pixel in the image is known. The camera model allows us to combine the camera's internal parameters and depth values to obtain the 3D coordinates of any pixel point in the camera coordinate system. Given that the framework only needs an approximate position of the fire safety device, which is sufficient for subsequent maintenance or information updates on the management platform. Hence, we take the center of the 2D bounding box as input to acquire the 3D coordinates of that point, representing the 3D positional information of the equipment. The projection process is presented. For the single frame, only the spatial transformation relationship (R_{wc}, T_{wc}) from the camera coordinate system $O_c - xyz$ to the 3D coordinate system $O_w - xyz$ needs to be obtained, and the target point can be projected into world coordinates by rigid transformation.

Fire safety equipment in buildings is scattered in distribution. As a result, cameras need to obtain the status of all such equipment from multiple frames. Hence, it is necessary to establish a unified world coordinate system and convert the fire safety equipment detected from each perspective into this coordinate system. In the case of multiple frames, when dealing with multiple frames, it is crucial to determine the spatial transformation

relationship $(R_{k:k+1}, T_{k:k+1})$ between adjacent frames, which indicates the camera's spatial pose at each moment. In order to obtain the set of spatial poses of the camera at different timestamps, the proposed framework computes the correspondence between feature points in adjacent images to determine the camera's pose during motion, which includes transformation and rotation. Suppose the rigid spatial variation matrix between the two frames of the camera at time k and time $k+1$ is $T_{k:k+1} \in \mathbb{R}^{4 \times 4}$, it can be expressed follows:

$$T_{k:k+1} = \begin{bmatrix} R_{k:k+1} & t_{k:k+1} \\ 0 & 1 \end{bmatrix} \quad (4)$$

where, $R_{k:k+1} \in \mathbb{R}^{3 \times 3}$ is the rotation matrix and $t_{k:k+1} \in \mathbb{R}^{3 \times 1}$ is the translation matrix. As can be seen in Figure 5(c), along the inspection path of the camera. By setting $T_{1:k} = \{T_{1:0}, T_{2:1}, \dots, T_{k:k-1}\}$ as the set of cameras pose transformations, it can be achieved through spatial coordinate transformation. It is assumed that at a certain time step k , the spatial coordinate of any point under the camera's perspective can be converted into a unified world coordinate system by using Equation (5), as shown below.

$$\begin{bmatrix} x_{pw} \\ y_{pw} \\ z_{pw} \\ 1 \end{bmatrix} = T_{1:0} \cdot T_{2:1} \cdots T_{k-1:k-2} \cdot T_{k:k-1} \begin{bmatrix} x_p^k \\ y_p^k \\ z_p^k \\ 1 \end{bmatrix} \quad (5)$$

To attain higher accuracy and robustness, our framework employs one of the most advanced visual SLAM algorithms, ORB-SLAM3 (Campos, Elvira, Rodriguez, M. Montiel, & D. Tardos, 2021) to obtain $(R_{k:k+1}, T_{k:k+1})$. Moreover, for the purpose of obtaining a real-scaled camera trajectory, the RGB and depth maps gathered by the depth camera are employed as inputs to the SLAM system. Within this system, first, all 3D coordinates related to the same device ID are retrieved. Then, after setting a threshold to remove outliers, the average of the remaining points is regarded as the 3D position of the equipment. Eventually, by projecting the detected devices onto the SLAM map, a map containing 3D device location information can be constructed, as depicted in Figure 3.

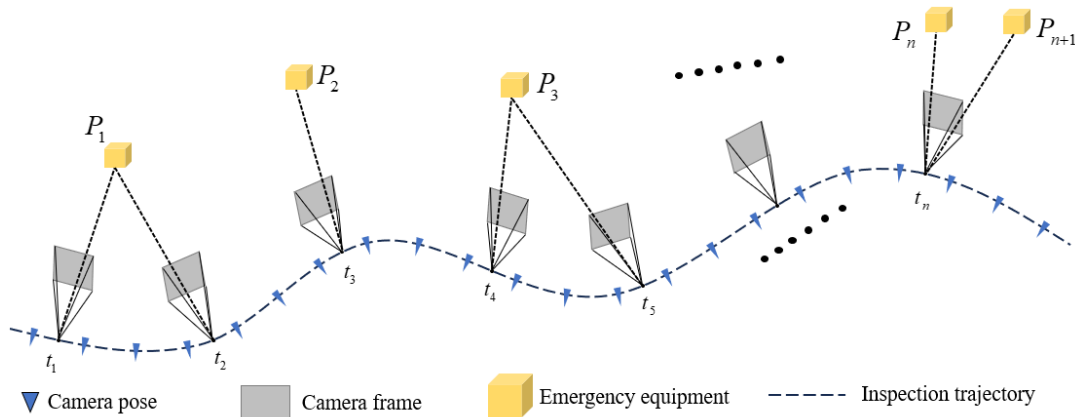


Figure 3. Creation of an equipment information map

3. RESULTS

3.1 Dataset and training

The existing public datasets are insufficient for meeting the training and testing requirements of this study as they lack consideration of an adequate amount of emergency equipment. The study identified eight crucial emergency equipment types: alarm bell, alarm button, emergency shower, escape sign, fire extinguisher, sand bucket, hose reel, and warning light. To address this gap, a dataset featuring 2793 instances of diverse emergency equipment was meticulously curated. Images, captured using smartphones or D435i depth cameras under varying lighting conditions, were intentionally marred by occlusion, lighting inconsistencies, and motion blur to simulate real-world challenges. Leveraging the LabelImg annotation tool, bounding boxes were meticulously marked around each item, followed by data augmentation using Mosaic to enhance dataset diversity. Figure 4 presents several annotated and enhanced images. The dataset was then partitioned randomly into training, validation, and testing subsets, each comprising 70%, 15%, and 15% of the total images, respectively.



Figure 4. The dataset after data augmentation

In Figure 5, information regarding bounding boxes in the dataset is depicted. Figure 5(a) illustrates the object quantities for each type within the dataset, primarily influenced by the frequency of emergency equipment use and their placements in buildings. Meanwhile, Figure 5(b) showcases the diverse sizes and aspect ratios of object bounding boxes. The distribution of the center points of the bounding boxes is depicted in Figure 5(c), highlighting their concentration in the central region of the image data. Finally, Figure 5(d) displays a scatter plot representing the width and height of the bounding boxes, with the darkest hues clustered in the bottom left corner, indicating a prevalence of small targets within the dataset.

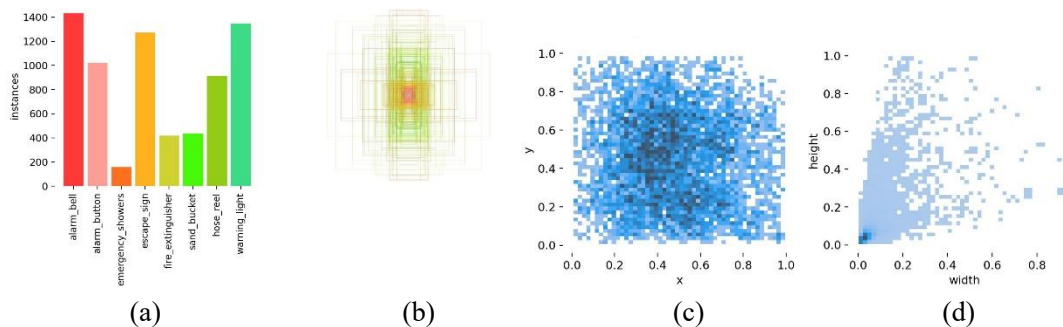


Figure 5. Information about the manually labeling of objects in dataset.

3.1 Experimental validation

To assess the effectiveness of the proposed method and establish a direct link between equipment status and the corresponding devices, four distinct corridor scenarios were selected for on-site experiments (refer to Figure 6). Corridors, typically housing a high density of fire safety equipment, were chosen due to their significance in this context. Scenario ① features a straight corridor with consistent artificial lighting, offering a uniform background that facilitates the easy identification of fire safety equipment in a structured environment. In contrast, scenario ②, also a straight corridor, transitions from an outdoor section with strong natural light to an indoor area with relatively weaker illumination, potentially causing fire safety equipment to be obscured by debris. Finally, scenario ③ presents an indoor Z-shaped corridor with even artificial and natural lighting. Despite this, the corridor's spaciousness results in some emergency maps appearing as small objects in the images.

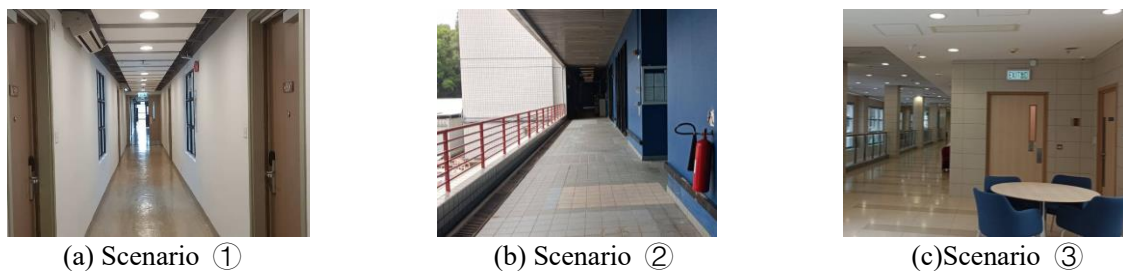
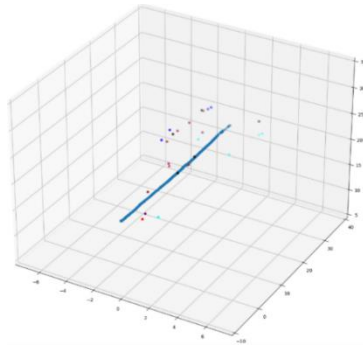


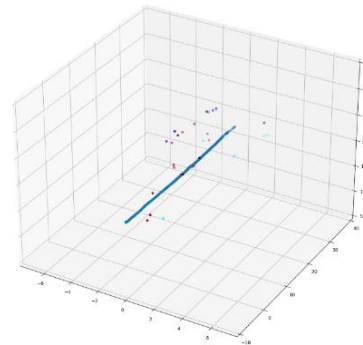
Figure 6. Four on-site scenarios

Figures 7 to 9 show the distribution of camera inspection paths and fire safety equipment, comparing the results of the original method and the improved method. The statistical results are shown in Tables 1-3,

respectively. Both the original method and the improved method can effectively cope with this scenario ①, counting the correct number, thanks to the strong contrast backgrounds. In scenario ②, due to the issue of obstruction and debris, the original method is unable to detect one of the fire extinguishers, while poor lighting caused an alarm bell to be missed. The improved method can detect these two ignored devices. In scenario ③, the improved method can also provide similar improvements, especially for smaller emergency devices such as alarm bells and warning lights.



(a) Inspection map of the original method

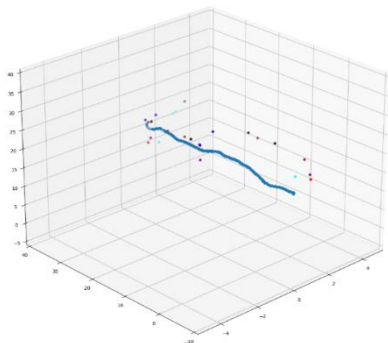


(b) Inspection map of the improved method

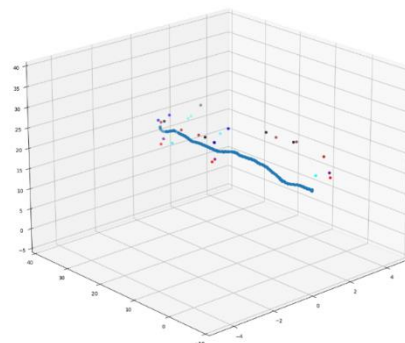
Figure 7 Inspection map for scenario①

Table 1. Scenario ① fire safety equipment inspection statistics

Scenario ①	Ground truth	Original method	Improved method
Alarm_bell	3	3	3
Alarm_button	2	2	2
Emergency_showers	0	0	0
Escape_sign	7	7	7
Fire_extinguisher	0	0	0
Sand_bucket	0	0	0
Hose_reel	2	2	2
Warning_light	0	0	0



(a) Inspection map of the original method

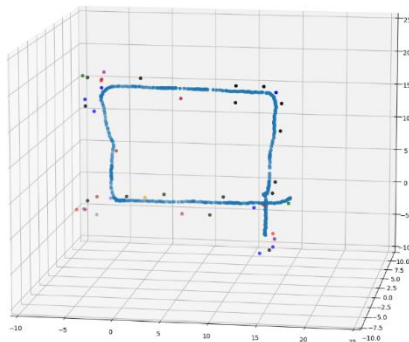


(b) Inspection map of the improved method

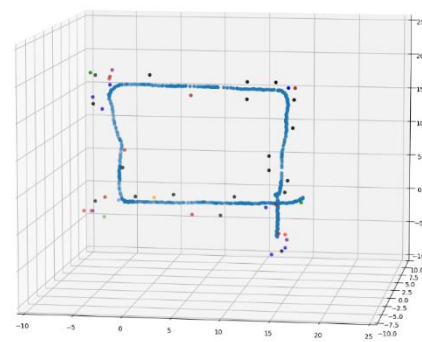
Figure 8 Inspection map for scenario②

Table 2. Scenario ② fire safety equipment inspection statistics

Scenario ②	Ground truth	Original method	Improved method
Alarm_bell	<u>7</u>	<u>6</u>	<u>7</u>
Alarm_button	3	2	3
Emergency_showers	0	0	0
Escape_sign	4	4	4
Fire_extinguisher	<u>5</u>	<u>4</u>	<u>5</u>
Sand_bucket	0	0	0
Hose_reel	3	3	3
Warning_light	5	5	5



(a) Inspection map of the original method



(b) Inspection map of the improved method

Figure 9 Inspection map for scenario ③

Table 3. Scenario ③ fire safety equipment inspection statistics

Scenario ④	Ground truth	Original method	Improved method
Alarm_bell	<u>9</u>	<u>7</u>	<u>9</u>
Alarm_button	3	3	3
Emergency_showers	3	3	3
Escape_sign	8	8	8
Fire_extinguisher	0	0	0
Sand_bucket	1	1	1
Hose_reel	3	3	3
Warning_light	<u>19</u>	<u>16</u>	<u>19</u>

4. DISCUSSION

Buildings often face the issue of fire safety equipment malfunctioning due to insufficient maintenance (Dong et al., 2014). As building scales increase, the workload for regular inspections grows significantly, necessitating substantial manpower and time. To address this, it is crucial to invest in automated inspection and maintenance systems for fire safety equipment to ensure they remain in optimal working condition, providing reliable safety for the building and its occupants. We have developed an advanced detection framework that leverages vSLAM to construct a three-dimensional relationship among equipment, cameras, and trajectories, effectively pinpointing the spatial positions of dispersed devices. DeepSort is utilized to track objects in video feeds, assigning unique IDs and enabling accurate object counting and identity recognition. Additionally, feature enhancement strategies and improvements to the loss function have been integrated to boost the detection network's performance, minimizing missed detections and false positives. The results demonstrate that this method adeptly handles various inspection scenarios, significantly improving accuracy and reliability of the system.

5. CONCLUSIONS

Compared to the original method, the improved method significantly enhances the robustness and accuracy of emergency equipment detection, offering more reliable support for emergency management and response. Looking ahead, developing an integrated management platform for fire safety equipment could be a

priority. This platform would integrate with existing detection frameworks to provide comprehensive oversight and operational support for equipment. It could include features like intelligent scheduling and advanced path planning algorithms to streamline inspection processes, thus reducing inspection times and distances traveled. Furthermore, the integration of mobile terminals and apps could transform inspection management, enabling robots to remotely handle and execute tasks. This would allow inspection personnel to seamlessly receive task allocations, access optimized path plans, retrieve device-specific information, and easily record and submit inspection results. Such a holistic approach promotes real-time data sharing, boosts work efficiency and precision, and ultimately improves operational workflows in fire safety equipment management.

REFERENCES

- Campos, C., Elvira, R., Rodriguez, J. J. G., M. Montiel, J. M., & D. Tardos, J. (2021). ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial, and Multimap SLAM. *IEEE Transactions on Robotics*, 37(6), 1874-1890. doi:10.1109/tro.2021.3075644
- Cao, Y., Kamaruzzaman, S. N., & Aziz, N. M. (2022). Building Information Modeling (BIM) Capabilities in the Operation and Maintenance Phase of Green Buildings: A Systematic Review. *Buildings*, 12(6). doi:10.3390/buildings12060830
- Dai, X., Chen, Y., Xiao, B., Chen, D., Liu, M., Yuan, L., & Zhang, L. (2021). *Dynamic head: Unifying object detection heads with attentions*. Paper presented at the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.
- Dong, Q., You, F., & Hu, S.-q. (2014). Investigation of Fire Protection Status for Nanjing Representative Historical Buildings and Future Management Measures. *Procedia Engineering*, 71, 377-384. doi:10.1016/j.proeng.2014.04.054
- Guan, Y.-x., Fang, Z., & Wang, T.-r. J. P. e. (2018). Fire risk assessment and daily maintenance management of cultural relic buildings based on ZigBee technology. *Procedia Engineering*, 211, 192-198. doi:10.1016/j.proeng.2017.12.004
- He, Y., Zhu, C., Wang, J., Savvides, M., & Zhang, X. (2019). *Bounding box regression with uncertainty for accurate object detection*. Paper presented at the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.
- Huang, H., Chen, Z., Zou, Y., Lu, M., & Chen, C. J. a. p. a. (2023). Channel prior convolutional attention for medical image segmentation. *Computers in Biology and Medicine*. doi:10.1016/j.combiomed.2024.108784
- Ivanov, M. L., & Chow, W.-K. (2022). Fire safety in modern indoor and built environment. *Indoor and Built Environment*, 32(1), 3-8. doi:10.1177/1420326x221134765
- Kobes, M., Helsloot, I., de Vries, B., & Post, J. G. (2010). Building safety and human behaviour in fire: A literature review. *Fire Safety Journal*, 45(1), 1-11. doi:10.1016/j.firesaf.2009.08.005
- Li, N., Becerik-Gerber, B., Krishnamachari, B., & Soibelman, L. J. A. i. C. (2014). A BIM centered indoor localization algorithm to support building fire emergency response operations. *Automation in Construction*, 42, 78-89. doi:10.1016/j.autcon.2014.02.019
- Liu, S., Johns, E., & Davison, A. J. (2019). *End-to-end multi-task learning with attention*. Paper presented at the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.
- Lo, S. M. J. S. S. (1998). A building safety inspection system for fire safety issues in existing buildings. *Structural Survey*, 16(4), 209-217. doi:10.1108/02630809810243220
- Lu, T., Tervola, S., Lü, X., Kibert, C. J., Zhang, Q., Li, T., & Yao, Z. (2021). A novel methodology for the path alignment of visual SLAM in indoor construction inspection. *Automation in Construction*, 127. doi:10.1016/j.autcon.2021.103723
- Pan, Y., Braun, A., Brilakis, I., & Borrmann, A. (2022). Enriching geometric digital twins of buildings with small objects by fusing laser scanning and AI-based image recognition. *Automation in Construction*, 140. doi:10.1016/j.autcon.2022.104375
- Reis, D., Kupec, J., Hong, J., & Daoudi, A. J. a. p. a. (2023). Real-Time Flying Object Detection with YOLOv8. *arXiv preprint, arXiv:2305.09972*. doi:10.48550/arXiv.2305.09972
- Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., & Savarese, S. (2019). *Generalized intersection over union: A metric and a loss for bounding box regression*. Paper presented at the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.
- Spencer, B. F., Hoskere, V., & Narazaki, Y. (2019). Advances in Computer Vision-Based Civil Infrastructure Inspection and Monitoring. *Engineering*, 5(2), 199-222. doi:10.1016/j.eng.2018.11.030
- Tang, R., Fitzgerald, G., Hou, X. Y., & Wu, Y. P. (2014). Building an evaluation instrument for China's hospital emergency preparedness: a systematic review of preparedness instruments. *Disaster Med Public Health Prep*, 8(1), 101-109. doi:10.1017/dmp.2014.10
- Tse, P. W. J. J. o. Q. i. M. E. (2002). Maintenance practices in Hong Kong and the use of the intelligent scheduler.

- Journal of Quality in Maintenance Engineering*, 8(4), 369-380. doi:10.1108/13552510210448540
- Vijayalakshmi, S., & Muruganand, S. J. I. R. J. E. T. (2017). Internet of Things technology for fire monitoring system. 4(6), 2140-2147.
- Walters, M., & Hastings, E. J. F. (1998). Fire safety legislation in Hong Kong. *Facilities*, 16(9/10), 246-253. doi:10.1108/02632779810229066
- Wang, D., Gao, L., Zheng, J., Xi, J., & Zhong, J. (2025). Automated recognition and rebar dimensional assessment of prefabricated bridge components from low-cost 3D laser scanner. *Measurement*, 242. doi:10.1016/j.measurement.2024.115765
- Wang, J., Xu, C., Yang, W., & Yu, L. J. a. p. a. (2021). A normalized Gaussian Wasserstein distance for tiny object detection. *arXiv preprint, arXiv:2110.13389*. doi:10.48550/arXiv.2110.13389
- Wang, S.-H., Wang, W.-C., Wang, K.-C., & Shih, S.-Y. (2015). Applying building information modeling to support fire safety management. *Automation in Construction*, 59, 158-167. doi:10.1016/j.autcon.2015.02.001
- Xin, J., & Huang, C. (2013). Fire risk analysis of residential buildings based on scenario clusters and its application in fire risk management. *Fire Safety Journal*, 62, 72-78. doi:10.1016/j.firesaf.2013.09.022
- Xu, W., Chan, S. C., Leong, W. Y., & Borondo, F. (2023). Effectiveness Study of Artificial Intelligent Facility System in Maintaining Building Fire Safety (Case Study: Typical Public Building Cases of Fire-Fighting Facilities Management in China). *Discrete Dynamics in Nature and Society*, 2023, 1-21. doi:10.1155/2023/2592322

Vision LLM-Driven Operational Hazard Recognition for Building Fire Safety Compliance Checking

Dayou Chen¹, Long Chen², Yiheng Zeng³, Craig Hancock⁴, Russell Lock⁵ and Simon Sølvsten⁶

1) Ph.D. Candidate, School of Architecture, Building and Civil Engineering, Loughborough University, United Kingdom. Email: d.chen@lboro.ac.uk

2) Ph.D., Assoc. Prof., Department of Architecture and Civil Engineering, City University of Hong Kong, Hong Kong, SAR. Email: longchen@cityu.edu.hk

3) Department of Computer Science, University College London, United Kingdom. Email: leo.zeng.22@ucl.ac.uk

4) Ph.D., Reader, Architecture, Building and Civil Engineering, Loughborough University, United Kingdom. Email: c.m.hancock@lboro.ac.uk

5) Ph.D., Reader, Department of Computer Science, Loughborough University, United Kingdom. Email: r.lock@lboro.ac.uk

6) Ph.D., Assoc. Prof., European Center for Risk & Resilience Studies, University of Southern Denmark, Denmark. Email: simos@sam.sdu.dk

Abstract: Building fire incidents pose significant risks to human lives and property, making fire safety compliance a critical aspect of building management. Traditional compliance checks are largely manual, relying on expert inspectors to assess and report on fire safety standards. While prior research has explored Automated Compliance Checking (ACC) during the design phase, limited attention has been given to the operational phase, where dynamic risks necessitate continuous monitoring. This study proposes a novel approach that leverages vision Large Language Models (vLLMs) to automate fire safety compliance monitoring in the operational phase. The developed method frames hazard recognition as a Visual Question Answering (VQA) task, enabling the model to analyze visual data and respond to textual queries regarding potential fire hazards. The system employs a Vision Transformer (ViT) for visual encoding and a multimodal fusion process, allowing the vLLM to generate contextually relevant descriptions of observed hazards, along with regulatory references including Occupational Safety and Health Administration (OSHA) standards. Evaluation results demonstrate significant improvements in hazard recognition over a generic vLLM baseline, with an average BLEU score of 0.1355 compared to 0.0410 and higher ROUGE scores reflecting superior precision and coherence. The model's ability to automatically generate structured hazard description reports has practical implications for assisting expert-driven inspections, offering a comprehensive and effective solution for long-term fire safety management. This study thus advances ACC research by providing a comprehensive, automated method for continuous fire safety compliance in operational building environments.

Keywords: Fire Safety Compliance, Automated Compliance Checking (ACC), Vision Large Language Models (vLLM), Visual Question Answering (VQA), Computer Vision, Operational Phase Monitoring

1. INTRODUCTION

Building fire incidents pose significant threats to human lives and properties, making fire safety compliance an essential aspect of building management. Regulatory frameworks are in place to ensure a minimum standard of safety and performance for built assets, yet these checks predominantly depend on manual inspections. Although prior research has focused on automated compliance checking (ACC) during the design phase, there is a notable gap in automating compliance checking throughout the operational phase of buildings. This gap is especially concerning as fire risks require continuous monitoring to ensure safety measures remain effective beyond initial design compliance.

The operational phase presents unique challenges for fire safety compliance, where risks are dynamic and necessitate ongoing assessment. Fire safety measures, such as the installation of fire doors and maintenance of unobstructed escape routes, are critical during the design and construction phase but may be compromised over time. For instance, fire doors could be propped open, or escape paths could be blocked by stored items, undermining their intended purpose. Thus, continuous monitoring is essential to maintain the integrity of these safety measures, effectively mitigating risks that may emerge during a building's operational lifespan.

Table 1. Summary of Automated Code Compliance Checking Methods

Literature	Applicable Phases	Technological Approach
Malsane et al. (2015)	Design Phase	Developing object model for automated compliance checking
Jiang et al. (2022)	Design Phase	Using ontology mapping and rule-based reasoning for ACC
Fitkau and Hartmann (2024)	Design Phase	Ontology-based knowledge formalization for ACC in fire safety
Zhang and El-	Design Phase	Using NLP and logic reasoning for

Gohary (2017)		fully automated code checking
Zhou et al. (2022)	Design Phase	Integrating NLP and CFG for rule interpretation in ACC
Bloch et al. (2023)	Design Phase	Using graph neural networks for ACC
Chen et al. (2024)	Design Phase	Using computer vision and deep generative models for ACC
Bosché (2010)	Construction Phase	Automated recognition of CAD objects in laser scans for compliance control
Cheng et al. (2022)	Construction Phase	Computer vision and deep learning for safety compliance monitoring
Ding et al. (2022)	Construction Phase	Using visual question answering and deep learning for safety compliance checking
Beach et al. (2024)	Operational Phase	Feasibility study on moving ACC to operational phase

Table 1 summarizes the key contributions of previous research in ACC, highlighting the applicable phases and technological approaches employed. Historically, the literature on ACC has emphasized Building Information Modeling (BIM) as a primary tool for compliance verification during the design phase. This emphasis has led to substantial advancements in areas such as structural safety, fire risk assessment, and water distribution systems. These studies typically rely on BIM data combined with rule-based reasoning, aligning design models with regulatory standards. For example, Martins and Monteiro (2013) developed the "LicA" application for automating water distribution network checks using IFC-based BIM models, while Malsane et al. (2015) introduced an object model for fire safety compliance. Zhang et al. (2013) further extended BIM's application to safety rule checking, applying algorithms to prevent fall hazards in construction planning. These design-phase solutions illustrate BIM's capacity to enhance safety and regulatory compliance early in a building's lifecycle.

Recent advances in Natural Language Processing (NLP) and Machine Learning (ML) have further enhanced ACC automation by reducing reliance on manual interpretation of regulatory texts. Zhang and El-Gohary (2017) and Zhou et al. (2022) developed NLP frameworks to extract regulatory requirements directly from text, converting complex legal language into computable rules. These innovations are crucial for addressing one of the most labor-intensive components of ACC, facilitating a more scalable approach to compliance checking. Additionally, Bloch et al. (2023) applied Graph Neural Networks (GNN) to ACC, specifically targeting accessibility requirements in residential designs. By bypassing traditional hard-coded rules, GNNs enable a more flexible and scalable approach, demonstrating the potential of ML techniques in overcoming the limitations of conventional rule-based systems.

While recent studies have started to address the gap in lifecycle-wide compliance with an emphasis on extending ACC into the operational phase, significant work remains to be done. Beach et al. (2024) advocated for automated data capture and analysis to facilitate compliance during building operations. However, fully automated compliance checking and monitoring specifically tailored for the operational phase of buildings is yet to be developed.

The use of computer vision in ACC has become an emerging focus, particularly for compliance monitoring and hazard recognition. Cheng et al. (2022) developed a deep learning model for classifying Personal Protective Equipment (PPE) and tracking worker movement on construction sites. Their approach highlights how vision-based models can support real-time monitoring of safety compliance by recognizing and categorizing visual data relevant to worker safety. In parallel, Ding et al. (2022) leveraged a Vision-and-Language Transformer (ViLT) model for Visual Question Answering (VQA) to detect unsafe behaviors on construction sites. By incorporating both visual and language processing capabilities, their model exemplifies a more advanced integration, enabling nuanced hazard detection through complex reasoning about images and textual queries.

These studies underscore the potential of image-based data for automating hazard recognition in ACC, signaling a shift toward image-driven applications that support operational compliance monitoring. The rapid advancement of vision models—especially with contrastive learning techniques—has further enabled complex scene understanding (Radford et. al, 2021). This recent technique aligns visual data with text for richer interpretation beyond traditional vision-based tasks such as object detection. Recently, Large Language Models (LLMs) have expanded these capabilities, allowing more sophisticated reasoning tasks on image inputs (Touvron et. al, 2023; Liu et. al, 2024). Yet, despite these advancements, the application of such technologies to hazard recognition for image-driven ACCs in building operational environments remains largely unexplored.

In light of these advancements, this study proposes a novel approach using computer vision, particularly vision Large Language Models (vLLM), to automatically identify fire safety non-compliance in buildings during the operational phase. By enabling long-term, dynamic compliance monitoring, this method seeks to assist managers and inspectors in generating compliance reports, reducing the workload for fire safety experts and improving the efficiency of ongoing fire risk management.

2. METHOD

Ensuring fire safety compliance during the operational phase of a building requires a systematic approach capable of identifying dynamic hazards in real time. This study introduces a novel method leveraging vision-Large Language Models to automate hazard recognition and reporting. The proposed approach frames hazard detection as a Visual Question Answering problem, enabling the generation of detailed and context-aware descriptions of fire safety non-compliance. The framework, referred to as Fire Compliance Visual Question Answering (FCVQA), integrates computer vision and natural language processing techniques to analyze multimodal inputs and produce actionable outputs.

The following subsections describe the problem formulation, the FCVQA framework, the model training process, and the evaluation metrics used to assess the model's performance. Each component contributes to the overall objective of automating compliance checks in operational building environments, offering an effective solution to augment traditional expert-based inspection workflows.

2.1 Problem Formulation

To enable automated assessment of potential fire hazards in the operational phase, we deploy a vision-Large Language Model (vLLM) to perform hazard recognition in building environments. This model uses advanced computer vision techniques to evaluate risks based on visual inputs, offering an advanced method for identifying non-compliance with building regulations in real time.

The hazard recognition task is framed as a Visual Question Answering (VQA) problem (Antol et. al, 2015), where the model receives an image of a building environment alongside a textual query about potential fire hazards. The vLLM integrates these inputs to generate a description of any detected risks. For example, given an image with the query, "What fire hazards do you see?" the model may respond with, "The fire exit is obstructed by a large box," or "The fire door is propped open, compromising its function."

Formally, this task can be described as:

$$M(I, Q) \rightarrow A, \quad (1)$$

where M is the vLLM, I is the input image, Q is a natural language question about possible hazards, and A is the generated answer describing any observed fire safety non-compliance issues.

2.2 Fire Compliance Visual Question Answering (FCVQA) Framework

This work introduces the Fire Compliance Visual Question Answering (FCVQA) framework for effective hazard recognition in building operations. The FCVQA framework, as depicted in Figure 1, integrates visual and textual data through a multimodal encoding and decoding process, leveraging the robust reasoning capabilities of large language models (LLMs) for contextual hazard assessment.

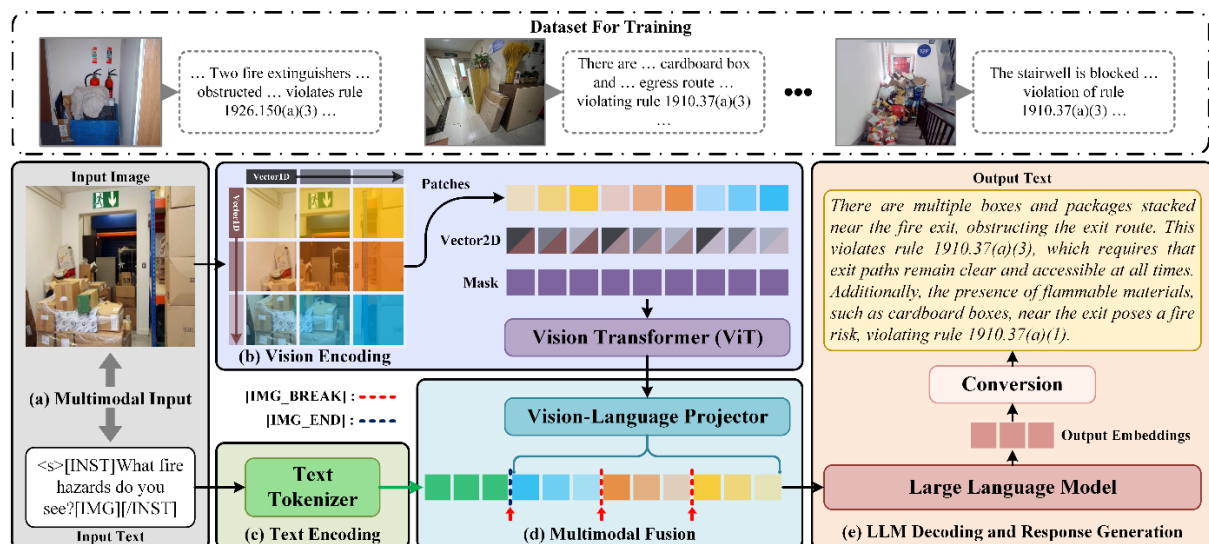


Figure 1. The proposed vLLM-driven FCVQA framework for operational hazards recognition

The FCVQA process includes several key stages:

Vision Encoding: A Vision Transformer (ViT) encodes the visual input, transforming the building environment's image into a high-dimensional latent representation. This representation captures relevant details,

such as object shapes and spatial arrangements, essential for identifying fire hazards.

Text Encoding: The textual query about fire risks is tokenized and encoded to produce a compatible latent representation. This enables the model to interpret specific fire-related questions, such as identifying blocked exits or verifying the functionality of fire doors.

Multimodal Fusion: The encoded image and text representations are combined within a shared latent space, allowing the vLLM to process visual and textual information concurrently. This fusion enables the model to analyze complex visual contexts and correlate them with fire safety questions.

LLM Decoding and Response Generation: Using a large language model decoder, the FCVQA framework processes the combined latent representation to generate a textual response. This output describes any fire hazards detected in the image, providing actionable insights aligned with the original query. For instance, if the image depicts a blocked fire exit, the model would generate a response indicating this specific non-compliance issue.

By leveraging the FCVQA framework, the system can dynamically assess fire safety conditions in real time. This approach enables comprehensive hazard recognition, capturing both obvious and subtle risks within building environments. Examples of hazards that the vLLM can detect include obstructed exits, inadequately stored flammable materials, and disabled fire safety equipment, all of which are essential for ongoing compliance with fire safety regulations. The vLLM's strength lies in its ability to interpret and reason about complex scenes. Unlike conventional object detection models, which only localize predefined objects, vLLMs can understand contextual factors that determine whether an object or scene poses a fire hazard. For instance, a stack of furniture may not be inherently dangerous, but when positioned to obstruct egress, it becomes a regulatory violation. The vLLM's capacity for contextual reasoning is therefore essential for identifying these nuanced risks, allowing for a more sophisticated assessment of fire safety compliance.

2.3 Model Training

The model training approach first follows a vision-language pre-training (VLP) strategy, where larger datasets are utilized to pre-train models on various vision and language tasks prior to specific Visual Question Answering (VQA) training. This VLP strategy, as discussed by Gan et. al (2022), facilitates the transfer of knowledge from a range of related tasks, thereby enhancing the model's understanding of concepts that go beyond the constraints of limited VQA datasets. For VLP, interpretative VQA datasets were reviewed as potential sources of pre-training data. Several popular VQA datasets representing diverse scenarios were combined to enrich the pre-training phase, specifically VizWiz-VQA (Gurari et. al, 2018) and VQAv2 (Goyal et. al, 2017).

Following the pre-training, the model was fine-tuned using a smaller, specialized dataset developed for this study, referred to as the Fire Compliance VQA dataset. This dataset comprises 135 image-text pairs focused on common operational fire hazards in building environments, such as blocked fire exits. This small dataset serves as the foundation for this preliminary experiment to validate the feasibility of the proposed approach. The dataset was split into training, validation, and test sets with an 8:1:1 ratio. Model training was conducted on a single Nvidia A6000 GPU, utilizing a learning rate of 0.001. Each image is paired with the question, 'What fire hazards do you see?' alongside a human-generated response that describes the scene, identifies non-compliance issues, and references relevant Occupational Safety and Health Administration (OSHA) regulations. In this work, OSHA regulations relating to the maintenance, safeguards, and operational features for exit routes, as outlined in Maintenance, Safeguards, and Operational Features for Exit Routes (29 C.F.R. § 1910.37; OSHA, n.d.), and fire protection, as detailed in Fire Protection (29 C.F.R. § 1926.150; OSHA, n.d.), have been selected and used. The response format mirrors a hazard-report style, offering a practical approach to support or potentially replace expert-led building compliance inspections by providing precise, regulatory-aligned insights.

During the training phase, we employed full supervision to optimize the model. For VQA, the sigmoid function predicted character scores \hat{s} between 0 and 1 as probabilities for the answer. We utilized the widely used binary cross-entropy (Teney et. al, 2018) as the guiding loss function. Given a dataset D having n samples, with image $v \in V$, question $q \in Q$ and answer $a \in A$, the goal is to train a model to optimize a mapping function $f: V \times Q \rightarrow \mathbb{R}^{|A|}$. Hence, the answering loss L_{ans} can be formula as follows:

$$L_{ans} = - \sum_q^Q \sum_a^A s_{qa} \log(\hat{s}_{qa}) - (1 - s_{qa}) \log(1 - \hat{s}_{qa}), \quad (2)$$

Where s_{qa} is the true score or target probability for answer a given question q . \hat{s}_{qa} is the predicted probability score for answer a given question q , obtained using the sigmoid activation function on the model's output logits. This score represents the model's confidence in predicting a as the correct answer for q . This loss function L_{ans} calculates the cumulative penalty across all questions and possible answers in the dataset, guiding the model to produce probability scores that align with the true answer distribution for each question.

2.4 Evaluation Metrics

To evaluate the accuracy and effectiveness of hazard identification and compliance flagging, several established

metrics were utilized: BLEU-4 (Papineni et. al, 2002), ROUGE-1, ROUGE-2, and ROUGE-L (Lin,2004). These metrics measure the alignment between the generated descriptions and reference texts, focusing on precision in identifying key compliance terms, such as "fire extinguisher," "fire exit," and specific regulatory codes. In this evaluation, we compare the pre-trained and fine-tuned model using our proposed method against the original, generic vLLM model as the baseline to validate the effectiveness of the proposed approach.

BLEU-4 evaluates the precision of four-gram sequences in the generated text, capturing phrase accuracy critical for detailed compliance reporting. ROUGE-1 measures unigram overlap, verifying that essential terms are included, while ROUGE-2 assesses bigram overlap, ensuring accuracy in phrase sequences. ROUGE-L calculates the longest common subsequence between generated and reference texts, reflecting structural coherence. Formally, these metrics are defined as follows:

$$\text{BLEU-4} = \exp \left(\sum_{n=1}^4 \frac{1}{n} \log p_n \right) \times \text{brevity penalty}, \quad (3)$$

$$\text{ROUGE-1} = \frac{\sum \text{matched unigrams}}{\sum \text{reference unigrams}}, \quad (4)$$

$$\text{ROUGE-2} = \frac{\sum \text{matched bigrams}}{\sum \text{reference bigrams}}, \quad (5)$$

$$\text{ROUGE-L} = \frac{\text{LCS}(\text{generated}, \text{reference})}{\text{length of reference}}, \quad (6)$$

These metrics provide a comprehensive evaluation framework for the model's ability to generate precise, contextually relevant descriptions, essential for automated hazard detection and compliance monitoring.

3. RESULTS

Table 2. Evaluation Metrics Comparison

Method	Average BLEU	ROUGE-1	ROUGE-2	ROUGE-L
Generic vLLM	0.0410	0.3912	0.1086	0.2435
Proposed Method	0.1355	0.5375	0.2590	0.3683

The evaluation metrics for hazard identification performance were compared between the generic vLLM baseline and the proposed method. Table 2 highlights the substantial improvements achieved by the proposed method in both BLEU and ROUGE metrics. The proposed method's average BLEU score reached 0.1355, compared to 0.0410 from the generic vLLM baseline, reflecting enhanced precision in descriptive phrasing. For ROUGE metrics, the proposed method also outperformed the baseline significantly, with ROUGE-1, ROUGE-2, and ROUGE-L scores of 0.5375, 0.2590, and 0.3683, respectively, compared to the baseline's scores. These results underscore the proposed method's greater effectiveness in capturing key terms, maintaining phrase structure, and preserving the overall coherence required for precise hazard identification and compliance monitoring.

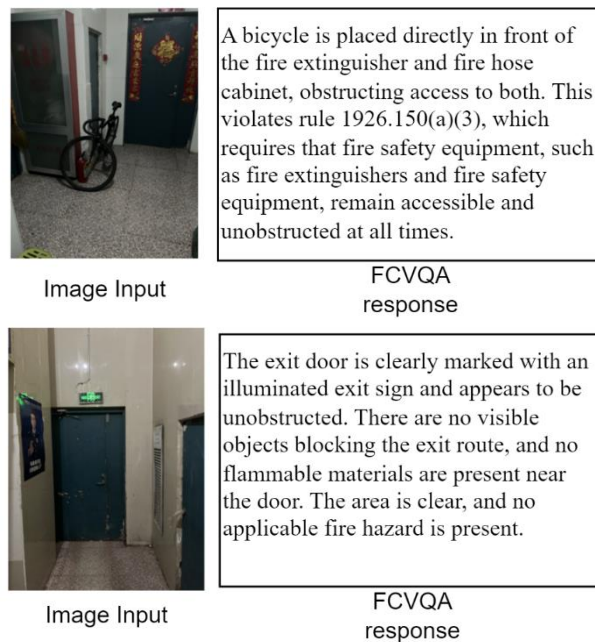


Figure 2. Example output of the developed FCVQA model

Upon closer examination of the model's specific outputs, as illustrated in Figure 2, it is evident that the developed FCVQA model demonstrates strong capabilities in identifying and reporting fire safety compliance issues within building environments. In the first example, the FCVQA model detects that a bicycle is obstructing access to a fire extinguisher and a fire hose cabinet. The model accurately cites OSHA regulation 1926.150(a)(3), which mandates that fire safety equipment must always remain accessible. This output demonstrates the model's ability not only to recognize objects in the scene but also to interpret the situation within the context of fire safety compliance, correlating the obstruction with relevant legal requirements. At the same time, the second example shows the model's capability to identify a compliant scene. The FCVQA model observes that the exit door is clearly marked and unobstructed, with no visible hazards nearby. It concludes that no applicable fire hazard is present, illustrating the model's capacity to affirm compliance in safe environments. This ability to confirm safety in compliant situations is critical for practical deployment, as it enables the model to assist or potentially replace expert-based inspections by offering precise, context-aware assessments of building fire safety compliance.

4. DISCUSSION

This section evaluates the contributions and limitations of the developed model within the broader context of ACC research. First, a comparative analysis highlights the advancements of the proposed method over existing approaches, particularly in terms of complex hazard scene reasoning and automated hazard report generation. Following this, the limitations of the current framework are discussed, along with potential future directions to address these challenges and enhance the model's applicability and robustness. Together, these discussions provide a comprehensive perspective on the strengths and areas for improvement of the proposed method in advancing fire safety compliance monitoring.

4.1 Comparative Analysis

In discussing the capabilities of vision-based approaches for safety compliance monitoring, Table 3 highlights key distinctions between the developed method and prior work by Cheng et al. (2022) and Ding et al. (2022). Specifically, the term "Complex hazard scene reasoning" refers to a model's capacity to interpret intricate hazard situations beyond simple object recognition tasks, such as those related to Personal Protective Equipment (PPE). Traditional CNN-based models, such as those utilized by Cheng et al. (2022), are limited to identifying isolated objects and lack the interpretative depth necessary for complex hazard assessment. In contrast, vision-language models, as applied in Ding et al. (2022) and the developed approach, provide a more nuanced understanding of contextual safety risks, enabling a richer analysis of compliance within complex scenes.

Table 3. Comparison of Vision-Based Safety Compliance Methods

Literature	Complex hazard scene reasoning	Written report generation
Cheng et. al (2022)	N	N
Ding et. al (2022)	Y	N
Developed method	Y	Y

An additional, critical differentiator is the capability for written report generation. While Ding et al. (2022) employed a vision-language model capable of reasoning about hazard scenes, it did not integrate a mechanism for generating fully automated hazard reports. This limitation necessitates expert intervention to interpret the model's outputs and document findings, which can be both time-consuming and prone to human error. Automated report generation, as realized in the developed method, fills this gap by producing structured, regulatory-aligned hazard descriptions that support compliance inspections.

The ability to generate comprehensive, written hazard reports represents a significant advancement in the automation of safety compliance workflows. This feature is essential for either assisting or potentially replacing traditional expert-driven processes, as it reduces the reliance on human inspectors to document compliance findings. By offering a model that interprets hazards with regulatory context and generates formalized reports, the developed approach not only enhances the efficiency of compliance monitoring but also contributes to automated hazard report generation, which is critical for regulatory adherence and continuous monitoring in safety-critical environments.

In summary, the developed model distinguishes itself from previous methods by combining complex scene reasoning with automated report generation, thereby offering a more comprehensive and effective solution for building fire safety compliance monitoring.

4.2 Limitations and Future Directions

Despite the contributions of the developed model, several limitations must be addressed to ensure broader applicability and robustness. A key limitation lies in the dataset used for training and evaluation. The current dataset of 135 image-text pairs, while sufficient for proof-of-concept, is too small and homogeneous to generalize

the model's performance across diverse real-world scenarios. Expanding the dataset to include a wider range of building types, operational conditions, and hazard scenarios, particularly edge cases such as rare fire hazards, would significantly enhance the robustness and applicability of the method. Leveraging larger, publicly available datasets or curated collections tailored for fire safety compliance would further strengthen its predictive capabilities.

Another limitation is the lack of real-world testing under dynamic operational conditions. The framework has not yet been deployed in live building environments, such as high-occupancy residential buildings or industrial facilities. A pilot study in these settings would provide crucial insights into the practical performance, scalability, and utility of the model. Real-world testing would also enable the identification of potential challenges, such as adapting to variations in lighting conditions, spatial layouts, or obstruction types, which are critical for ensuring reliable hazard detection in practice.

Additionally, the computational demands of the model, particularly for real-time hazard detection and report generation, warrant further exploration. It is unclear how the system performs under constrained hardware environments or in challenging conditions, such as low-light settings or poor image quality. Addressing these concerns through hardware optimization or lightweight model adaptations would enhance the framework's usability in resource-constrained scenarios.

The lack of contextual integration also limits the system's ability to fully understand hazards within the broader building-level context. Currently, the developed method focuses on hazard recognition, which, while critical, represents only part of the overall ACC workflow. Hazards such as blocked escape routes or improperly maintained fire doors often require contextual understanding of building design specifications, spatial relationships, and evacuation strategies to assess their true severity. Future work should address these limitations by integrating operational hazards with comprehensive building information, such as BIM or other digital twin technologies. This would enable a transition from isolated hazard detection to end-to-end ACC workflows, offering a more holistic solution for fire safety compliance.

5. CONCLUSIONS

The study presents a novel approach to automated fire safety compliance monitoring for building operational phases, addressing a critical gap in the field of ACC. While previous research has predominantly focused on the design phase, this work emphasizes the importance of continuous monitoring to manage dynamic fire risks during the operational phase. The proposed method leverages advanced computer vision techniques, specifically a vLLM, to assess fire safety compliance in the building operational phase, thus supporting long-term, dynamic fire risk management.

The developed model frames hazard recognition as a VQA task, wherein an image of the building environment is processed along with a text query regarding potential fire hazards. The model effectively identifies hazards, generates descriptive responses aligned with regulatory standards, and provides references that align with OSHA regulations. This dual capability of recognizing hazards and producing detailed, regulatory-referenced reports is essential for enhancing the efficiency of fire safety inspections. By generating automated hazard reports, the model reduces the reliance on human experts to manually document compliance issues, offering a scalable solution for compliance monitoring.

Quantitative evaluation further demonstrates the effectiveness of the proposed method. Significant improvements over a generic vLLM baseline were observed across BLEU and ROUGE metrics, indicating enhanced precision and coherence in hazard descriptions. These metrics underscore the model's ability to capture critical terminology, maintain structured phrasing, and generate accurate, comprehensive responses. Such performance highlights the model's potential to support or even replace expert-driven fire safety compliance inspections.

Nevertheless, limitations remain. The small dataset used for fine-tuning constrains the model's ability to generalize across diverse building types, operational scenarios, and edge cases, such as rare or atypical fire hazards. Expanding the dataset to encompass a broader range of real-world conditions would significantly enhance the robustness and applicability of the model. Additionally, the lack of contextual integration prevents the system from fully understanding hazards within the broader building-level context, such as design specifications, spatial relationships, and evacuation strategies. Currently, the developed method focuses on hazard recognition, which, while critical, is only part of the overall ACC workflow. Future work should address these limitations by expanding the dataset and developing systems that integrate operational hazards with comprehensive building information. Such advancements would enable transition from isolated hazard detection to complete ACC workflows.

Overall, this study contributes an automated and effective solution for fire safety compliance monitoring in operational building environments. By combining complex scene understanding with automated reporting, the proposed method enables a more efficient approach to managing operational building fire risks, offering building managers and risk inspectors a powerful tool to ensure continuous regulatory compliance and ultimately safeguard human lives and property.

ACKNOWLEDGMENTS

This work was supported by the Willis Towers Watson Research Network TECHNIGI-CDT Scholarship.

REFERENCES

- Agrawal, P., Antoniak, S., Hanna, E. B., Chaplot, D., Chudnovsky, J., Garg, S., ... & Wang, T. (2024). Pixtral 12B. arXiv preprint arXiv:2410.07073.
- Antol, S., Agrawal, A., Lu, J., Mitchell, M., Batra, D., Zitnick, C. L., & Parikh, D. (2015). VQA: Visual question answering. In *Proceedings of the IEEE international conference on computer vision* (pp. 2425-2433).
- Beach, T., & others. (2024). Moving automated compliance checking to the operational phase of the building life-cycle: Analysis and feasibility study in the UK. *International Journal of Construction Management*, 1–10. <https://doi.org/10.1080/15623599.2024.2366727>
- Bloch, Y., Zhao, R., & Zhang, J. (2023). Graph-based learning for automated code checking – Exploring the application of graph neural networks for design review. *Advanced Engineering Informatics*, 58, 102137. <https://doi.org/10.1016/j.aei.2023.102137>
- Bosché, F. (2010). Automated recognition of 3D CAD model objects in laser scans and calculation of as-built dimensions for dimensional compliance control in construction. *Advanced Engineering Informatics*, 24(1), 107-118. <https://doi.org/10.1016/j.aei.2009.08.006>
- Chen, D., Chen, L., Zhang, Y., Lin, S., Ye, M., & Sølvsten, S. (2024). Automated fire risk assessment and mitigation in building blueprints using computer vision and deep generative models. *Advanced Engineering Informatics*, 62, 102614. <https://doi.org/10.1016/j.aei.2024.102614>
- Cheng, J. C. P., Wong, P. K.-Y., Luo, H., Wang, M., & Leung, P. H. (2022). Vision-based monitoring of site safety compliance based on worker re-identification and personal protective equipment classification. *Automation in Construction*, 139, 104312. <https://doi.org/10.1016/j.autcon.2022.104312>
- Ding, Y., Liu, M., & Luo, X. (2022). Safety compliance checking of construction behaviors using visual question answering. *Automation in Construction*, 144, 104580. <https://doi.org/10.1016/j.autcon.2022.104580>
- Gan, Z., Li, L., Li, C., Wang, L., Liu, Z., & Gao, J. (2022). Vision-language pre-training: Basics, recent advances, and future trends. *Foundations and Trends® in Computer Graphics and Vision*, 14(3–4), 163-352.
- Goyal, Y., Khot, T., Summers-Stay, D., Batra, D., & Parikh, D. (2017). Making the V in VQA matter: Elevating the role of image understanding in visual question answering. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 6904-6913).
- Gurari, D., Li, Q., Stangl, A. J., Guo, A., Lin, C., Grauman, K., ... & Bigham, J. P. (2018). Vizwiz grand challenge: Answering visual questions from blind people. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3608-3617).
- Jiang, L., Shi, J., & Wang, C. (2022). Multi-ontology fusion and rule development to facilitate automated code compliance checking using BIM and rule-based reasoning. *Advanced Engineering Informatics*, 51, 101449. <https://doi.org/10.1016/j.aei.2021.101449>
- Liu, H., Li, C., Wu, Q., & Lee, Y. J. (2024). Visual instruction tuning. *Advances in neural information processing systems*, 36.
- Lin, C. Y. (2004). ROUGE: A package for automatic evaluation of summaries. In *Text summarization branches out* (pp. 74-81).
- Martins, J. P., & Monteiro, A. (2013). LicA: A BIM based automated code-checking application for water distribution systems. *Automation in Construction*, 29, 12-23. <https://doi.org/10.1016/j.autcon.2012.08.008>
- Malsane, S., Matthews, J., Lockley, S., Love, P. E. D., & Greenwood, D. (2015). Development of an object model for automated compliance checking. *Automation in Construction*, 49, 51-58. <https://doi.org/10.1016/j.autcon.2014.10.004>
- Mistral AI team. (2024). Mistral nemo. Retrieved from <https://mistral.ai/news/mistral-nemo>
- Occupational Safety and Health Administration. (n.d.). Maintenance, safeguards, and operational features for exit routes (29 C.F.R. § 1910.37). U.S. Department of Labor.
- Occupational Safety and Health Administration. (n.d.). Fire protection (29 C.F.R. § 1926.150).
- Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002, July). BLEU: A method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics* (pp. 311-318).
- Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. In *Proceedings of the International Conference on Machine Learning*. Retrieved from <https://api.semanticscholar.org/CorpusID:231591445>
- Teney, D., Anderson, P., He, X., & Van Den Hengel, A. (2018). Tips and tricks for visual question answering:

- Learnings from the 2017 challenge. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4223-4232).
- Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E., & Lample, G. (2023). LLaMA: Open and Efficient Foundation Language Models. ArXiv, abs/2302.13971.
- Zhang, J., & El-Gohary, N. M. (2017). Integrating semantic NLP and logic reasoning into a unified system for fully-automated code checking. *Automation in Construction*, 73, 45-57.
<https://doi.org/10.1016/j.autcon.2016.08.027>
- Zhou, Y.-C., Zheng, Z., Lin, J.-R., & Lu, X.-Z. (2022). Integrating NLP and context-free grammar for complex rule interpretation towards automated compliance checking. *Computers in Industry*, 142, 103746.
<https://doi.org/10.1016/j.compind.2022.103746>
- Fitkau, I., & Hartmann, T. (2024). An ontology-based approach of automatic compliance checking for structural fire safety requirements. *Advanced Engineering Informatics*, 59, 102314.
<https://doi.org/10.1016/j.aei.2023.102314>

A KNOWLEDGE GRAPH BASED COMMON DATA ENVIRONMENT SOLUTION FOR BUILDING SEISMIC LOSS ESTIMATION

Zeyu PAN¹, Jianyong SHI²

1) Ph.D. Candidate, Department of Civil Engineering, Shanghai Jiao Tong University, Shanghai, China. Email: panzeyu@sjtu.edu.cn

2) Ph.D., Assoc. Prof., Department of Civil Engineering, Shanghai Jiao Tong University, Shanghai, China. Email: shijy@sjtu.edu.cn

Abstract: Seismic Loss Estimation (SLE) has become a critical aspect of modern building engineering, aiding in mitigation strategies, real-time disaster response, and post-earthquake reconstruction. The FEMA P-58 method, a performance-based earthquake engineering tool, efficiently links component damage states with engineering demand parameters for comprehensive seismic loss assessment. However, managing the extensive data and semantics required for such evaluations poses challenges. This paper proposes a Knowledge Graph (KG)-based solution, integrating object-based information management principles akin to Common Data Environment (CDE) and Building Information Modeling (BIM). By leveraging KG and digital twin technologies, this approach aims to facilitate dynamic seismic loss estimation, providing stakeholders with a comprehensive view of building performance and enabling efficient data access and analysis.

Keywords: Seismic loss estimation; FEMA P-58; Knowledge graph; Common data environment

1. INTRODUCTION

In modern building engineering practices, seismic loss estimation (SLE) has gradually become an important part across the different stages of buildings. In pre-earthquake stage, such procedure is useful for adopting suitable mitigation strategies or structural designs; during the earthquake event, it can facilitate the decision making for real-time disaster response and relief efforts; after an earthquake event, it can become a good criterion for reconstruction and rehabilitation (Shi et al., 2023; Xu et al., 2019). In the last few years, performance-based earthquake engineering approaches have been proposed to evaluate the performance of certain structures under seismic excitations. Among them, the FEMA P-58 method, a recent product of performance-based earthquake engineering, stands out for its capability to perform fine-grained, full-term seismic loss estimation (ATC, 2012). Its specialty is the capability to efficiently connect the damage state of the component combinations (performance groups, PGs) with specific engineering demand parameters (i.e., peak inter-story drift ratio PIDR, peak floor acceleration PFA). The FEMA P-58 method has been proven to be effective for diverse structural types, including masonry (Zeng et al., 2016), reinforced concrete frame (Baradaran et al., 2013, Del Vecchio et al., 2018, Shi et al., 2023), steel frame with seismic force-resisting systems (Del Gobbo et al., 2018, Yang & Mutphy, 2015).

Such applications require an abundance of data regarding the building of interest, such as the decomposition of different building subsystems, real-time measurements of specific physical quantities, detailed information of both structural and non-structural components, especially for their fragility against specific seismic impacts. Additionally, to effectively impart the building performance to different stakeholders, semantics regarding the seismic loss estimation may need to be further supplemented by the relevant material, such as technical reports or regulatory documents. Moreover, multiple models will be set during the analysis procedure of building seismic loss estimation. The integration of these analytical results is also helpful for the decision makers to have a comprehensive view of the target asset. Despite the wealthiness, challenges occur in managing the relevant data and exploring the underlying semantics. It spawns demand for developing an effective data management environment to make the involved information easy to access and exploit.

In the history of information management for urban assets, including buildings and civil engineering works, common data environment (CDE) and building information modeling (BIM) are two intertwined fundamental concepts. Specifically, according to ISO 19650-1 2018 (ISO, 2018), CDE refers to agreed source of information for any given project or asset, for collecting, managing and disseminating each information container through a managed process. Meanwhile, BIM is the use of a shared digital representation of a built asset to facilitate design, construction and operation processes to form a reliable basis for decisions. In our humble perspective, as illustrated in Figure 1, one of the important common idea underlying these two concepts is the object-based information management paradigm. The kernel of object-based information management paradigm for the built assets is to regard the involved object hierarchy (buildings, building subsystems, components) as the backbone of the relevant information collected from multiple sources. Information such as that within technical reports, real-time measurements from the corresponding sensor networks are required to be integrated to specific objects of interest so that the stakeholders can explore the embedded semantics freely with object-based query engine.

Similar ideas are also introduced by knowledge graph (KG) and digital twin technologies. In 2012, the famous blog posted by Google entitled “Things, not strings” revealed the essence of knowledge graph as managing

the relevant data with object-based paradigms. It proposed a web with semantics which reveals real-world entity interconnections, thereby enabling applications across multiple scenarios, including intelligent search, question and answer (Q&A) systems, personalized recommendation services. Knowledge graph allows users to explore the connections between entities of interest and their potential neighbors, thus broadening their knowledge horizons. Similarly, digital twin technologies, originally proposed for constructing mirrored digital representations for aircrafts, emphasized the importance of object-based information management for different level of details. In such context, information and status of the components from the different levels of system decomposition are integrated and reflected accordingly.

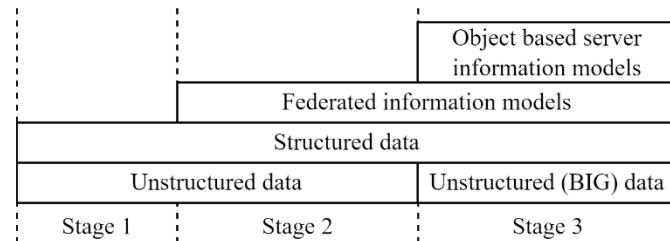


Figure 1. A perspective on stages of maturity of analogue and digital information management: Information Layer (Adapted from ISO 19650-1).

To this end, this paper proposes a KG-based object-based common data environment solution for dynamic seismic hazard assessment based on previous work and the specific data requirements of building seismic loss estimation tasks.

2. METHOD

2.1 KG-based CDE for building seismic loss estimation

In the context of building seismic loss estimation, conventional building information management may encounter three significant challenges: 1) representation of status history of buildings and their corresponding components under seismic events; 2) articulation of the underlying relations between the state of a complex object and the state of its components at various compositional levels, and 3) integration of heterogeneous data from multiple sources related to such applications. These three challenges correspond to three important information management requirements in the application of building seismic loss estimation. The proposed data model, hence, ultimately needs to be flexible enough to satisfy these requirements.

It is acknowledged that each object may possess multiple attributes and states, which can be subject to change over time. Illustrative examples of this phenomenon include the position of an object at different moments in time or the progression of structural damage in a building throughout an earthquake event. The information requirements for the attributes and states of the involved objects vary across different application scenarios and must be adapted to meet specific needs. In addition to attributes and states, the relationship between objects may also change over time, e.g., some dependent subsystem components may not be able to maintain normal functional connections with each other due to the failure of a component. As shown in Figure 2, insufficient redundancy or failure of key components of a system sometimes may lead to cascading failures, resulting in system-level functional paralysis and significant losses, such as the famous cascading failure of the Italian power grid reported by Buldyrev et al., 2010 and the cascading collapse of structures.

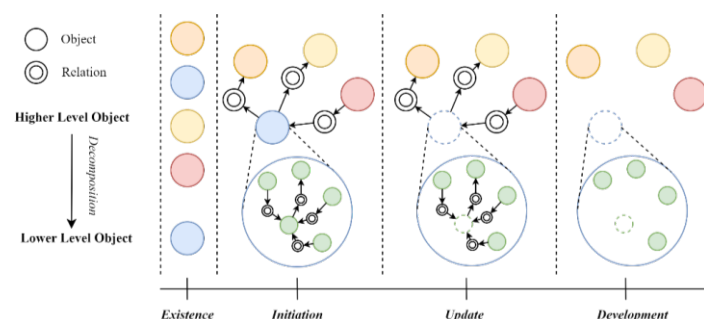


Figure 2. Schematics of cascading failures of interconnected components within a hierarchical system.

In the case of a complex object, it is possible that some of its properties and states may be jointly determined by the states of its various components. These components can be regarded as some kind of functional

relationship. The state of a building, as a complex object, is constrained by the states of its various components. For instance, in the context of an earthquake, the damage state of each building in the region constitutes a pivotal basis for emergency relief decision-making, with the overall state of the building reflected by the damage state of each floor, which in turn is reflected by the damage state of its associated structural components. Subsequent to the earthquake, the estimation of the overall loss to the building, in this case, is primarily determined by the loss of each component, including both structural and nonstructural components.

Relying on a single data source is commonly inadequate to accurately describe a building and its components to different stakeholders. It is also a complicated task to obtain the required information from the CDE for different analytical purposes. For the building seismic damage assessment scenario based on performance analysis, the semantics of describing a column object includes not only the location, such as the floor to which it belongs, and the connectivity of its components, but also its mechanical properties, structural conditions, constraints, ideal representations (refined or simplified representation of the element for structural analysis), the performance group to which it belongs and its corresponding fragility group, the meanings of the indicators describing the current damage state, the value of the performance indicators corresponding to different damage states. The semantics of these properties are usually stored implicitly in the information model and need to be described by domain knowledge from different knowledge sources. Therefore, how to effectively manage these heterogeneous knowledge sources is another practical issue to be considered in similar real-world application scenarios of building information.

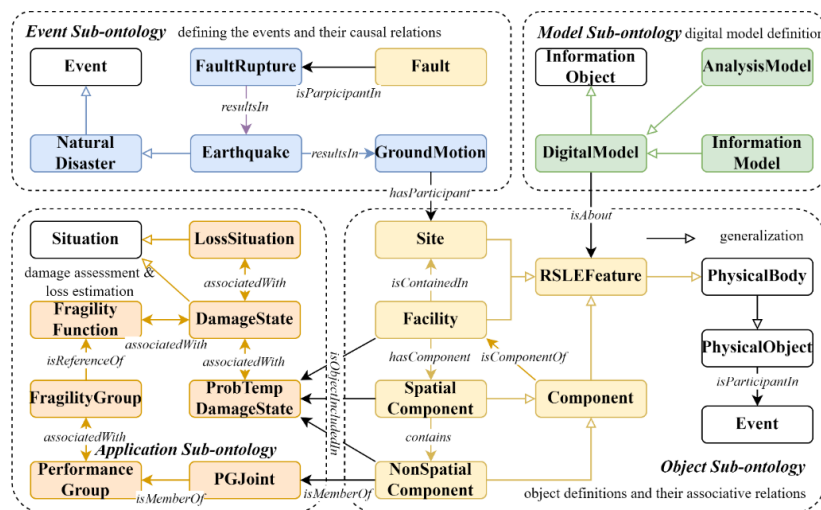


Figure 3. Dynamic regional seismic loss estimation ontology (DRSLEO) proposed in our previous study (Pan et al., 2024).

In our previous study, we have proposed a new domain ontology called dynamic regional seismic loss estimation ontology, DRSLEO in short. It has been established upon the concepts from FEMA P-58 and knowledge organization experiences from multiple outstanding outcomes of knowledge engineering, including a descriptive ontology for linguistic and cognitive engineering DOLCE (Borgo et al., 2021), industry foundation classes IFC (ISO, 2018), building topology ontology BoT (Janowicz et al, 2020). DRSLEO is still a developing ontology and was initially presented to formalize the relevant concepts and relationships within the field of seismic loss estimation.

Figure 3 illustrates the main structure of this ontology. It is composed of four conceptual parts, including event, object, model, and application sub-ontologies. The “event” sub-ontology maintains concepts for defining events and their causal relations, the “object” sub-ontology maintains those for defining objects and different kinds of associative relationships, the “model” sub-ontology maintains those for defining digital models and some other information objects, the “application” sub-ontology maintains those for states or situations that reflect the status of the objects of interest. In this section, details are given about how the knowledge-graph-based CDE can be used to specify the objects of interest and their interactions with their digital representations for respective purposes.

In information science, the term “ontology” is widely accepted as a conceptual representation of the domain of interest. It excels at conceptualizing the interconnections between different concepts. It is also a kind of data model that can be used to organize the data involved by the specific tasks of this application in a logical way. A well-established ontology is regarded as a logical foundation for reasoning upon the instantiated facts. As shown in Figure 4, we adopt DRSLEO to the data integration procedure in forming a knowledge graph based common data environment.

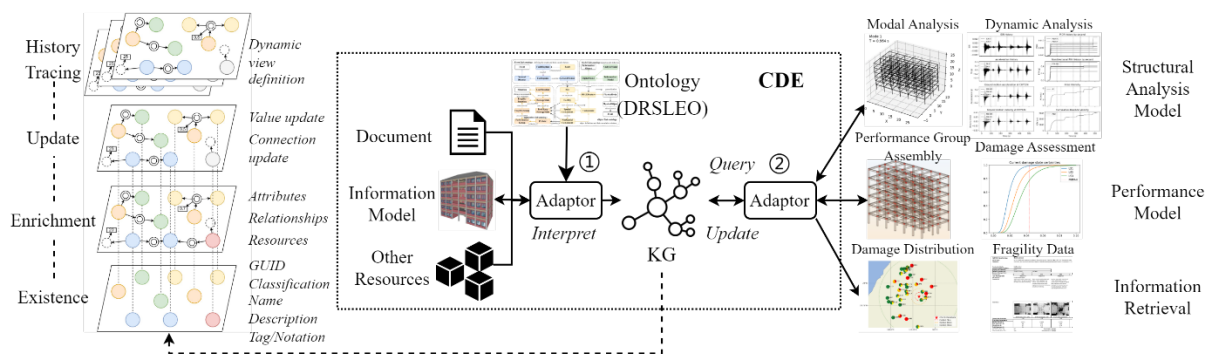


Figure 4. Illustration for the construction and implementation of the proposed knowledge graph (KG) based common data environment (CDE) for building seismic loss estimation.

As shown on the left of Figure 4, to begin with, objects and events involved in this application are identified and classified according to the data requirements of performance-based seismic loss estimation. They are further semantically enriched with the information selectively extracted from both structured and unstructured data (e.g., fragility database, specifications for structural analysis and fragility analysis, document regarding the asset and its sub-systems, ground motion database) and progressively organized into the corresponding knowledge graph. The corresponding knowledge graph in this stage serves as a base graph which reflects the initial state of the target asset. Upon this, update actions may be recalled with additive modifications to the base graph to reflect the status of the asset and its involved components. Each status should be assigned with a timestamp that is crucial to be queried with the specific time or duration of interest. The status of an object may be described by the combination of the valued indicators (e.g., peak inter-story drift ratio) or the existence of specific connections between the individuals, such as the functional relationships between a structural joint and its relevant structural components. Therefore, it is necessary to conserve the modification history as the value or relation regarding the object changes. The retrieved subgraphs for different update time instants can be seen as snapshots of the object during the different stages of the asset. As a result, they together form a dynamic view of asset status.

The first adaptor presented in Figure 4, hence, is responsible for extracting the essential data from multiple data sources and aligning them under the guidance of the corresponding ontological structure. In this procedure, objects of interest are identified and further semantically enriched with their attributes, status, constraints, correlations to other objects. Both structured and unstructured data are selectively passed through this interpretation process. For efficiency, not all data are fragmented into tiny data pieces so that they can be embarrassingly embedded into the corresponding knowledge graph. Reference mechanisms are adopted to address this issue. In one case, part of the semantics of the object of interest is already maintained by another accessible data source, such as, formal description of the classification of a specific object from an organized classification system. In another case, although images and videos regarding the object of interest are indispensable resources to represent the current and historical status of the object, they are not directly stored while referenced in this knowledge graph.

The other adaptor is to deliver the required information for different tasks, such as allocation of structural analysis model and performance model. The analytical results are written back to the knowledge graph through the adaptor once the analysis is completed. There is another task for common information retrieval or object status exploration. In this study, resource description framework (RDF) and the corresponding query language (SPARQL Protocol and RDF Query Language, SPARQL) for organize and manipulate the information of the knowledge graphs. Since the data requirements of the downstream tasks vary, details will be given in the following sections.

Prior to elucidating the construction of the proposed CDE, there is another important concept to delineate further, namely, “object”. An object in this context refers to a thing that is identifiable and capable of describing with a bunch of attributes. The definition of “object” in this context is much narrower than that in IFC, where it refers to the generalization of any semantically thing or process. It is close to “Endurant” in DOLCE’s context. Things such as events and processes are formally defined by other concepts. As the definition implies, the most fundamental attribute is to be identified as an individual and can be described by supplementary information. Each individual that needs to be identified in this context should be associated with a globally unique identifier (GUID), which is also emphasized by IFC. They include physically tangible items such as wall, beam or covering, physically existing items such as room, storey, building. Practically, a fixed 22-character style GUID, suggested by buildingSMART (buildingSMART, 2024), is adopted in this study. It can be created through a reversible transformation of a universally unique identifier (UUID), which is a most frequently used identifier generated by the algorithm of UUIDv4 defined by the specification of RFC 4122.

2.2 Construction of the knowledge graph

As shown in Figure 2 and Figure 4, each knowledge graph can be seen as a web with semantics which

reveals entity interconnections. With the convention of RDF, the graph consists of a series of triplets in the form of (subject, predicate, object). Each triplet is a fact or, more formally, statement regarding the knowledge graph. OWL 2 Web Ontology Language is a formal ontology language for the Semantic Web (Motik et al., 2012). It defines formal representation of the triplet components can be described. In such context, subject should be an individual, either named or anonymous; an object can be either an individual or a specific data value; a predicate, hence, can be regarded as object property when it connects two individuals, data property when it connects the individual with its attribute values, or annotation property when it is used to provide an annotation assertion of the individual, such as providing human-readable label or comment. Both named and anonymous individuals are regarded as nodes in the knowledge graph. Named individuals are those with globally unique identifiers, while anonymous individuals are not needed to be uniquely identified.

Compared to the wide usage of identifier within IFC schema, we narrowed the identifiable features to limited objects in this corresponding task. The criterion is that only the objects required to be frequently queried or referenced by the other individuals, such as wall, beam, column, storey, are named individuals, otherwise, they are treated as anonymous individuals, sometimes, also called blank node in the sense of graph structure. Upon the named individual, for descriptive capacity, each base individual in this task needs to be described with the other intrinsic attributes, such as name, description, tag, notation, and classification, as shown in Table 1. Unlike the GUID, these kinds of attributes are more human-readable, and suitable for indexing the corresponding individuals for query. This procedure is named “existence” definition in this study, as shown in Figure 4. The semantics of the individual can be complemented with its associated classification system outside the knowledge graph if it is possible to access with integrated query.

Table 1. Attribute declaration of BaseIndividual and BaseBlankNode.

Individual type	Prototype	Attribute	Description	Restraint
BaseIndividual	NamedIndividual	id	Unique identifier for the individual	Automatically generated
		name	Name of the individual	Optional[str]
		description	Description of the individual	Optional[str]
		tag	Tag of the individual	Optional[str]
		notation	Notation of the individual	Optional[str]
		classification	Class of the individual in an existing classification system	Optional[str]
		type	Class of the individual defined in the ontology	str
BaseBlankNode	AnonymousIndividual	_namespace	Namespace that the individual belongs to	Namespace IRI
		name	Name of the individual	Optional[str]
		description	Description of the individual	Optional[str]
		type	Class of the individual defined in the ontology	str

Based on these two fundamental concepts, domain-specific concepts proposed in DRSLEO are then extended accordingly. Inspired by BoT and IFC, a built asset is decomposed into several spatial components (e.g., storey, room, zone) and nonspatial components (e.g., structural components, nonstructural components). The built asset can be somehow classified as different types according to different classification systems. However, in ontological definition, it can be only classified as limited types that have similar attributes in common. The decomposition of building subsystems can be extracted from the corresponding information models easily, although the maturity of information about the involved subsystems varies from stage to stage.

With regard to their components, apart from the existence definition, each component type can have its specific attributes or associative relationships. They are essential for constructing a structural analysis model or performance model. For example, mass distribution strongly affects dead load distribution as well as the structural behavior under dynamic excitations. Therefore, the density and dimension of each tangible object will be calculated or at least estimated when it is declared as a component of the built asset. Other information that helps to enrich a component’s semantics in this task includes its location, geometric representation (e.g., body representation or axis representation) and its association with the other components (e.g., decomposition, connectivity, assignment).

Among the associative relationships, connectivity is one of the most important. Connectivity mentioned here is either regarding relationship between components that are physically connected or functionally connected. IFC treats a relationship that describes the spatially containment as a kind of connectivity. Such connectivity is essential for defining the loading path of a structural system from each storey above to the ground. It is also important information for us to determine the performance groups relating to each storey since some fragility groups of the structural components are categorized by structural joints, such as beam-column joints. To be noticed,

the extracted information by the adaptor forms an intricate interconnected web structure before it is converted into a graph structure. Once the information is collected, the instantiation of the knowledge graph of the asset of interest is conducted recursively from one individual to its associated individuals, as shown in the following pseudo code. However, such recursive individual instantiation procedure is prone to cyclic references since the recursion will not stop until stack overflows. Therefore, when organizing the extracted information from the information models and other information sources, it is necessary to make sure the associative relations form an acyclic directed graph.

Algorithm 1 Recursive individual instantiation (with rdflib library and Python programming language)

Input: *individual*: an object that holds information extracted from multiple data sources

graph: an RDF graph used for representing the asset of interest

Output: *iri*: an international reference identifier (IRI) for

```

1: function ToRDF(individual, graph):
2:   this ← individual.iri
3:   if not getattr(individual, "is_declared_in_rdf") then
4:     super(individual).ToRDF(graph)
5:     graph.add((this, RDF.type, DRSLEO[individual.type]))    # class declaration
6:     graph.add((this, DRSLEO["globalId"], Literal(individual.id)))    # guid
7:     if getattr(individual, "name") then
8:       graph.add((this, SKOS.prefLabel, Literal(individual.name)))    # human-readable name
9:     end if
10:    if getattr(individual, "description") then
11:      graph.add((this, RDFS.comment, Literal(individual.description))) # human-readable comment
12:    end if
13:    graph.add((this, RDFS.comment, Literal(individual.tostr))) # auto-generated default description
14:    # declarations for other attributes, case by case, omitted
15:    # declarations for associative relations, here we demonstrate the connectivity relation
16:    for other_individual ∈ individual.connects_to.values() do
17:      other_iri ← other_individual.ToRDF(graph)
18:      graph.add((this, DRSLEO["connectsTo"], other_iri))
19:      graph.add((other_iri, DRSLEO["connectsTo"], this))
20:    end for
21:    # other relations
22:    setattr(individual, "is_declared_in_rdf", True) # this line is activated in the top class method only
23:  return this
24:end function
    
```

3. RESULTS

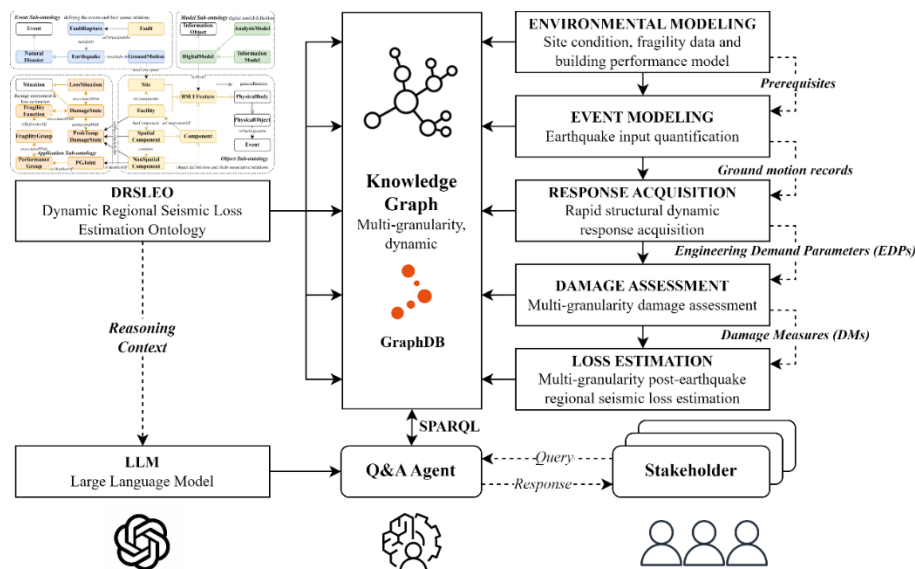


Figure 5. An overview of the workflow of semantic-augmented FEMA P-58 based DRSLE.

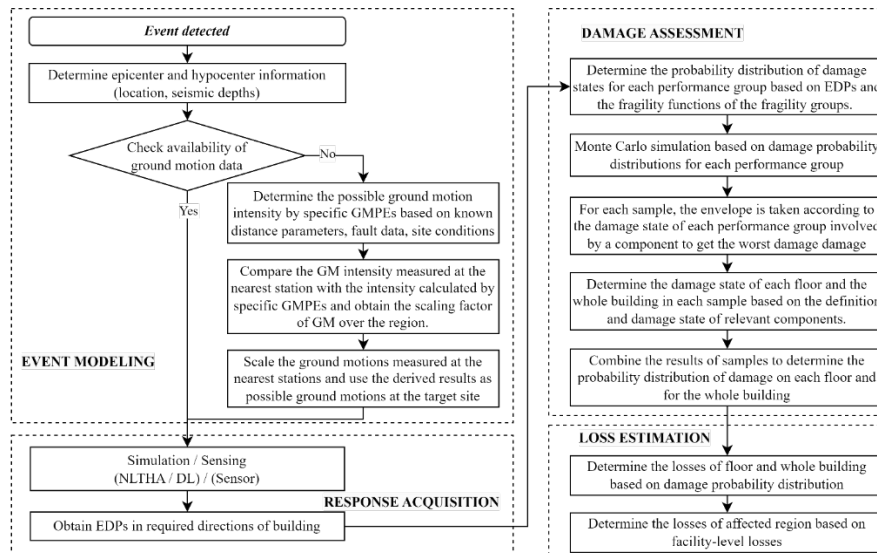


Figure 6. Specification of event modeling, response acquisition, damage assessment, loss estimation in RSLE.

Figure 5 and Figure 6 illustrate the five-step workflow of the dynamic regional seismic loss estimation based on FEMA P-58 and DRSLEO, which encompasses a) environmental modeling (site condition, fragility database, building performance model, etc.); b) event modeling (epicentral location, depths, fault type, rupture plane information, recorded or derived ground motions); c) response acquisition (simulation or sensing); d) damage assessment (Monte Carlo simulation upon the performance groups with EDPs and the fragility functions of their relevant fragility groups); and finally e) loss estimation (multi-granularity estimation for various loss measures based on the damage assessment results). DRSLEO, serving as a unified schema, provides a standardized approach for retrieving both static and dynamic data generated throughout this entire process. During the interaction procedure between the stakeholders with different roles and the bespoke LLM-empowered Q&A agent for DRSLE, the DRSLEO further provides schematic contexts for the exploration of the corresponding knowledge graphs of the relevant information.

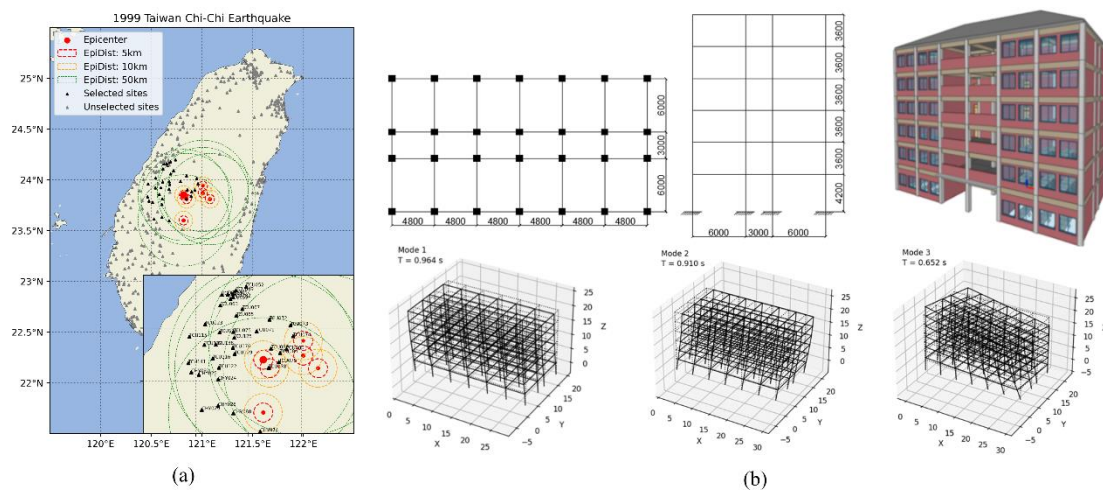


Figure 7. (a) epicentral information and influence circles of mainshock-aftershock sequence of 1999 Chi-Chi earthquake, and distribution of selected and unselected stations in this case study; (b) configuration of the 6-storey RC-OMF educational building (T_1 is 0.964s, T_2 is 0.910s, T_3 is 0.652s).

This scenario is set in 1999 Chi-Chi earthquake event. Ground motion records pertinent to this event have been extracted from the PEER NGA West2 database. The selection of this seismic event is based on two key factors. Firstly, Taiwan had established an extensive seismograph station network prior to the disaster, providing access to a considerable number of ground motion records (over 400 stations) with the majority containing three-dimensional ground motion data. This wealth of data renders it suitable for validating the ontological model's capability to accurately represent the spatial distribution of the loss estimation within the affected region. Secondly,

4. CONCLUSIONS

This research has identified and addressed three critical challenges in the data management of building seismic damage assessment: 1) representation of status history of buildings and their corresponding components under seismic events; 2) articulation of the underlying relations between the state of a complex object and the state of its components at various compositional levels, and 3) integration of heterogeneous data from multiple sources related to such applications. The proposed object-oriented CDE and knowledge graph-based solution aim to provide the necessary flexibility and integration capabilities to tackle these challenges effectively.

While the FEMA P-58 guidelines offer a solid foundation, the proposed approach extends its capabilities by incorporating real-time field data and object state assessments, thereby enhancing the comprehensiveness of seismic loss estimation. The integration of diverse data sources through an object-based framework allows for a more nuanced understanding of building performance during seismic events, facilitating informed decision-making for emergency response and reconstruction efforts.

Nevertheless, the current solution acknowledges the complexity of real-world building systems and the influence of functional components and subsystems on seismic damage assessment. Future work will focus on further refining the data model to capture these intricacies, ensuring a more accurate reflection of building behavior.

In summary, the integration of knowledge graph technology within an object-oriented CDE presents a promising avenue for advancing seismic damage assessment practices. By improving data management and facilitating knowledge discovery, this approach lays the groundwork for more resilient and informed building engineering in the face of seismic hazards.

ACKNOWLEDGEMENTS

This project is funded by the Shanghai Municipal Electric Power Company's technology project "Research on Key Technologies for AI Intelligent Design of 10kV Substations" SGTYHT/24-JS-001.

REFERENCES

- ATC. (2012). *Seismic performance assessment of buildings: volume 1 - Methodology (FEMA P-58-1)*.
- Baradaran Shoraka, M., Yang, T. Y., & Elwood, K. J. (2013). Seismic loss estimation of non-ductile reinforced concrete buildings. *Earthquake Engineering and Structural Dynamics*, 42(2), 297-310.
- Borgo, S., Ferrario, R., Gangemi, A., Guarino, N., Masolo, C., Porello, D., Sanfilippo, E. M., & Vieu, L. (2021). DOLCE: A descriptive ontology for linguistic and cognitive engineering. *ArXiv, abs/2308.01597*.
- buildingSMART. (2024). *IFC GUID*. Retrieved from buildingSMART website: <https://technical.buildingsmart.org/resources/ifcimplementationguidance/ifc-guid/>
- Buldyrev, S. V., Parshani, R., Paul, G., Stanley, H. E., and Havlin, S. (2010). Catastrophic cascade of failures in interdependent networks. *Nature*, 464(7291), 1025-1028.
- Del Gobbo, G. M., Williams, M. S., & Blakeborough, A. (2018). Seismic performance assessment of Eurocode 8-compliant concentric braced frame buildings using FEMA P-58. *Engineering Structures*, 155, 192-208.
- Del Vecchio, C., Di Ludovico, M., Pampanin, S., & Prota, A. (2018). Repair costs of existing rc buildings damaged by the l'aquila earthquake and comparison with FEMA P-58 predictions. *Earthquake Spectra*, 34(1), 237-263.
- ISO International Organization for Standardization. (2018). *Organization and digitalization of information about buildings and civil engineering works, including building information modelling (BIM) – Information management using building information modelling – Part 1: Concepts and principles (ISO 19650-1 2018)*.
- Janowicz, K., Rasmussen, M. H., Lefrançois, M., Schneider, G. F., & Pauwels, P. (2020). BOT: The building topology ontology of the W3C linked building data group. *Semantic Web*.
- Motik, B., Patel-Schneider, P. F., and Parsia, B. (2012). OWL 2 Web Ontology Language Structural Specification and Functional-Style Syntax. Retrieved from W3C website: <https://www.w3.org/TR/owl2-syntax/>
- Pan, Z., Shi, J., and Jiang, L. (2024). A Semantic Augmented Approach to FEMA P-58 based Dynamic Regional Seismic Loss Estimation Application. *Journal of Building Engineering*, 111224.
- Shi, J., Pan, Z., Jiang, L., Chen, P., An, C., and Mulatibieke, N. (2023). Research on a methodology for intelligent seismic performance evaluation and optimization design of buildings based on IFC and ontology. *Engineering Structures*, 288, 116213.
- Xu, Z., Zhang, H., Lu, X., Xu, Y., Zhang, Z., and Li, Y. (2019). A prediction method of building seismic loss based on BIM and FEMA P-58. *Automation in Construction*, 102, 245-257.
- Yang, T. Y., & Murphy, M. (2015). Performance evaluation of seismic force-resisting systems for low-rise steel buildings in Canada. *Earthquake Spectra*, 31(4), 1969-1990.
- Zeng, X., Lu, X., Yang, T. Y., & Xu, Z. (2016). Application of the FEMA-P58 methodology for regional

earthquake loss prediction. *Natural Hazards*, 83(1), 177-192.



Part 10

Information Management in Design and Maintenance

Probability Map Guided Point Rendering Technique for Refined Segmentation of High-Resolution Crack Images

Honghu Chu¹, Weiwei Chen² and Lu Deng^{3,*}

1) Ph.D. Candidate, College of Civil Engineering, Hunan University, Changsha, China. Email: chuhonghu@hnu.edu.cn

2) Ph.D., Lect., Bartlett School of Sustainable Construction, University College London, London, UK. Email: weiwei.chen@ucl.ac.uk

3) Ph.D., Prof., College of Civil Engineering, Hunan University, Changsha, China. Email: denglulu@hnu.edu.cn

Abstract: High-resolution (HR) imaging devices are now widely used for capturing crack images from civil structures, necessitating the development of algorithms for HR image segmentation. However, the traditional refined segmentation of HR images requires substantial GPU resources, which leads to the adoption of the cost-effective point rendering technique for inference. Considering that traditional rendering techniques require the use of coarse masks to guide the rendering points for processing prediction, these coarse masks typically fail to effectively focus the rendering points on the boundary regions of the slender cracks, resulting in ambiguous predictions at crack boundaries. In contrast, we introduce a novel rendering point sampling paradigm that enables the network to focus rendering points on crack boundary regions, guided by the probability maps during the inference phase. This approach significantly improves the segmentation accuracy of crack boundary regions from HR images without increasing computational resource dependence. Experiments on an open-source HR crack image dataset consistently show our method's superiority over state-of-the-art approaches, with final results of 84.24%, 93.78%, and 91.45% on IoU, mBA, and Dice, respectively.

Keywords: Deep learning, Crack segmentation, High-resolution image, Rendering technique.

1. INTRODUCTION

Bridge cracks are a leading cause of many bridge diseases, which can reduce the structural load-bearing capacity of bridges and may trigger structural disasters (Sony et al., 2021; Deng et al., 2022). Therefore, for traffic management departments, it is crucial to identify crack damage timely and accurate to ensure the normal operation of bridges during their service period (Xie et al., 2022). Semantic segmentation methods based on deep learning represent the most promising solution for the automatic and rapid identification of these cracks and have become a key topic in the field of bridge maintenance research (Chu et al., 2022).

In recent years, with the development of imaging technology and the increased requirements for detection, high-resolution (HR) imaging devices have begun to be gradually promoted and applied to collect surface crack images of engineering structures. The advantage of collecting HR images is that they contain richer crack detail information, which can be better used for the assessment of structural safety performance. However, traditional deep learning architectures often need to perform complex convolutional operations to generate new pixels during the decoding stage, based on deconvolution or transposed convolution. These convolutional operations require processing a larger number of data points as the size of the image increases, and the intermediate computational results produced need to be cumulatively stored in GPU memory. This makes the use of current traditional deep learning segmentation methods for accurate segmentation of HR crack images a computationally demanding operation, which is difficult to implement on conventional commercial GPUs, thus significantly limiting the advantages of HR imaging devices in terms of detailed characterisation from being fully utilised in practical tasks.

Unlike traditional encoder-decoder architectures, PointRend, proposed by Kirillov et al. (Kirillov et al., 2020), replaces the conventional decoding architecture with lightweight MLPs (Multilayer Perceptrons) that share weights and perform predictions through point-wise rendering. The MLP designed for point-wise rendering calculates each pixel of the image independently, meaning the network only needs to process a single or a small batch of pixel data at any given time. Therefore, this method is different from traditional upsampling, which requires expanding the entire feature map to a higher resolution all at once, thereby occupying a significant amount of GPU memory. In simple terms, this lightweight MLP performs predictions through point-wise rendering, significantly reducing the segmentation network's dependency on GPU memory while ensuring that the GPU memory does not increase with the size of the inference image.

Meanwhile, compared to traditional methods of dealing with GPU memory constraints by resizing HR images to smaller sizes or cropping HR images into patches for patch-wise inference (Cheng et al., 2020; Wang et al., 2020), the MLP-based point-wise rendering can perform calculations at the original resolution, thus avoiding the potential detail loss caused by proportional resizing and the inconsistencies or stitching artefacts that may occur during the stitching process, thereby better ensuring the coherence and consistency of the prediction results. Overall, the PointRend is considered the optimal solution for addressing GPU memory limitations in high-resolution image segmentation currently available.

However, this point-rendering method was originally designed for conventional targets in natural scenes,

but cracks, as a type of target with elongated topological structures and random distribution, differ greatly from conventional targets. Therefore, directly applying the point-rendering method for segmenting cracks poses challenges. Specifically, the original point-rendering technique, in the decoding phase, uses coarse segmentation masks to guide the rendering points for refined prediction, which is feasible for targets of conventional sizes because the edge details of such targets can be relatively accurately represented in the coarse segmentation masks. However, for targets like cracks that have elongated topological structures and random distribution, a large amount of detail features, including the edges of cracks and tiny crack branches, are already lost in the coarse segmentation masks. This deficiency results in rendering points being unable to effectively focus on these lost detail feature areas, thus producing ambiguous predictions for crack boundaries and missing predictions for tiny crack branches. Figure 1 provides a visual demonstration of the process and results of using coarse segmentation masks to guide rendering points for segmenting crack images. Due to the inaccuracy of the coarse segmentation, the refined rendering points are unable to be effectively guided to the tiny crack branches (the red marked areas in Figure 1(b)) and are instead more often directed to the inaccurately defined coarse segmentation crack edges (the green marked areas in Figure 1(b)). This results in the final predictions obtained through rendering techniques not only missing the tiny crack details but also failing to achieve refined predictions at the crack edges.

It should be noted that, in the field of crack inspection for civil engineering, the precise segmentation of crack edges and the complete identification of tiny crack branches are of indispensable importance for assessing structural integrity, implementing early damage recognition, formulating maintenance strategies, and conducting long-term health monitoring to ensure structural safety.

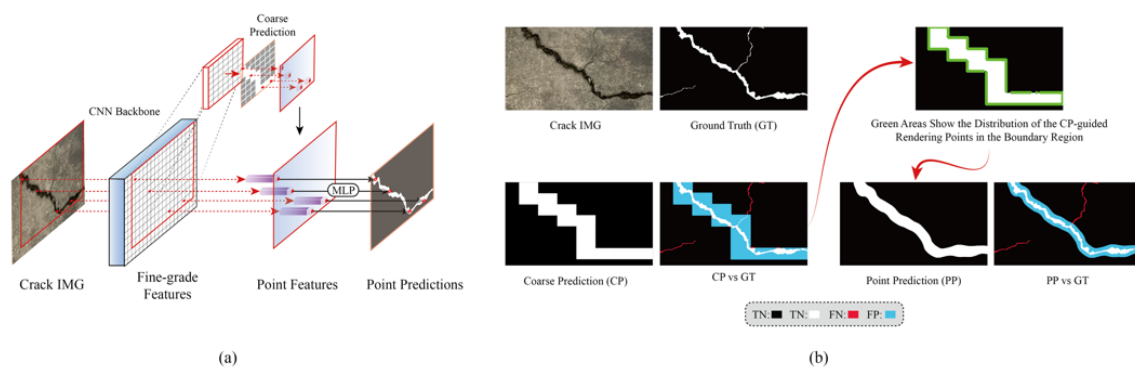


Figure 1. The process of segmenting a randomly selected HR crack image using the original PointRender architecture. (a) Schematic diagram of the PointRender architecture; (b) Coarse segmentation used to guide the rendering points and the final point prediction of the rendering points under its guidance.

To maintain the GPU memory efficiency benefits of the point-rendering method for segmenting HR images while ensuring precise inference at crack boundaries and tiny crack branches, this study proposes a refined rendering approach. This approach utilizes a probability map during inference to guide rendering point sampling. By redefining the intervals of simple and hard samples on the probability map, rendering points can be concentrated from simple samples, such as the background and main body of cracks, towards hard-sample areas like crack edges and tiny crack branches. Consequently, the network achieves refined segmentation for HR crack images on standard commercial GPUs. Our contributions are summarized as:

- For the first time, we incorporate point-rendering technology into HR crack image segmentation, tackling two significant challenges inherent to conventional segmentation methods: the substantial reliance on GPU computational resources and the inadequacy in achieving fine-grained segmentation detail.
- We customized a set of rendering point guidance strategies for the inference phase, which not only reduces the difficulty of training the rendering model but also significantly increases the model's segmentation accuracy for detailed areas such as crack edges and minor crack branches.
- This study proposes a novel paradigm for constructing networks for fine-grained segmentation of HR crack images. It advocates for the replacement of traditional prediction heads with those driven by the probability map and incorporates the multi-scale Transformer architecture for enhanced crack detail extraction.

We conducted experiments on an HR crack image dataset collected in the field and demonstrated that our model achieves state-of-the-art performance in both quantitative and qualitative results, with enhanced fine-grained segmentation capabilities and lower computational resource dependency.

2. Related Work

2.1 Deep Learning-based Crack Inspection

To enhance the detection of cracks' morphological characteristics (direction, edges, and corners), semantic segmentation algorithms rooted in deep learning have been applied to crack image analysis. Initially, FCN and

SegNet served as foundational frameworks (Bang et al., 2019; Liu et al., 2019; Zhang et al., 2019), albeit their repeated downsampling during feature extraction obscured significant fine crack details in the output (Zhang et al., 2019). The introduction of skip connections within U-shaped encoder-decoder architectures addressed these limitations partially, setting a new standard for crack detection (Ronneberger et al., 2015). Despite their advancements, these architectures, largely reliant on CNNs for feature extraction (Huyan et al., 2020; Mei et al., 2020), struggle with modelling long-distance interactions, complicating the segmentation of slender cracks (Khan et al., 2022). Recent shifts towards Transformer architectures (Qu et al., 2022; Shamsabadi et al., 2022; Guo et al., 2023), with their self-attention capabilities for extracting crack features, aim to overcome CNNs' constraints. Yet, these approaches do not adequately prioritize computational resources for critical areas like crack edges, leading to indistinct outcomes on mask boundaries.

2.2 Refined Segmentation

To enhance segmentation accuracy, refinement methodologies were introduced, starting with the amalgamation of Conditional Random Fields (Zheng et al., 2015; Lin et al., 2016; Chen et al., 2017) and Graph Models with deep learning frameworks (Dias & Medeiros, 2018). These initial attempts often relied on basic colour boundaries and did not leverage advanced semantic insights, inadequately addressing contours with minimal contrast to the background. Addressing this, some scholars proposed specialized refinement modules (Peng et al., 2017; Zhang et al., 2019), although their applicability was hampered by the prerequisite of predetermined thresholds. This prompted a shift to-towards the creation of versatile plug-in modules for broader utility. Notable advancements include RefineNet by Lin et al. (Lin et al., 2017), which refines boundaries through multi-path connections that integrate features across layers during network upsampling, and HRNet by Sun et al. (Wang et al., 2020), employing multi-path connections to enhance detail capture via high-resolution representation learning, thereby offering comprehensive image structural insights. Yet, these approaches' reliance on extensive computational re-sources and their substantial parameter counts challenge their application to high-resolution crack image segmentation and model training efficiency.

2.3 Rendering-based Technology for Refined Segmentation

The dilemma of indistinct boundary regions in crack image analysis parallels the jagged edge phenomenon in computer graphics when images are pixelized (Barron et al., 2021; Wu et al., 2021). This domain, crucial for generating and rendering images for applications like gaming and film effects, showcases the pinnacle of computer vision technology through image synthesis, animation, and virtual reality (Kellnhofer et al., 2021; Wang et al., 2021; Tewari et al., 2022). Drawing inspiration from these rendering techniques, Kirillov et al. (Kirillov et al., 2020) introduced PointRender, a novel approach utilizing rendering for enhanced segmentation in natural scenes. PointRender utilizes MLPs with shared weights for point-wise pixel prediction, capitalizing on rendering's precision without burdening GPU memory, thus enabling its application on standard GPUs for high-resolution imagery. Nevertheless, PointRender's dependency on coarse masks for rendering point guidance falls short of accurately detecting slender crack features. Addressing this, our study introduces a refined rendering strategy that maintains PointRender's memory efficiency while improving accuracy on crack peripheries and fine branches through a probability map-guided rendering point approach.

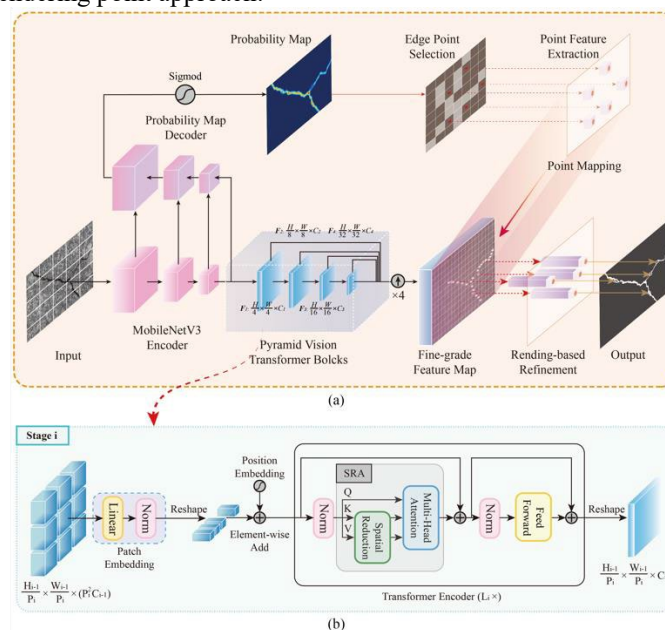


Figure 2. Schematic diagram of the proposed network. (a) The overall architecture; (b) Transformer block with

the SRA embedded.

3. METHOD

The probability map-guided point-rendering crack segmentation network proposed in this study follows the encoder-decoder architecture design pattern, consisting of three main parts: a crack fine-grained feature encoding backbone, a rendering point guidance branch, and a point-rendering-based fine-grained prediction head. The crack fine-grained feature encoding backbone is built from a lightweight encoder and a series of Transformer blocks; the rendering point guidance branch shares a lightweight encoder with the crack fine-grained feature encoding backbone and generates probability maps for guiding rendering points through a traditional decoder; the point-rendering based fine-grained prediction head is built based on the MLP that can perform point-by-point refinement predictions. It should be noted that the rendering point guidance branch and the point-rendering-based fine-grained prediction head together constitute the network's decoder part. Figure 2 visually presents the algorithmic details and computational logic of the proposed framework for HR crack image fine-grained segmentation.

3.1 Crack Fine-grained Feature Encoding Backbone

To ensure that the deep semantic feature maps contain rich crack detail information while improving the model's inference efficiency, this study employs a custom feature extraction encoder that combines a lightweight encoding architecture with an improved pyramid visual Transformer. An image is first input into the front end based on a lightweight encoding architecture, specifically adopting the MobileNetV3 architecture [31]. At this stage, the image undergoes a series of depthwise separable convolution operations that filter each channel of the input image using depth convolutions, followed by pointwise convolutions that fuse the features of these channels. This effectively reduces the model's parameter count and computational complexity while retaining key image features. The output of the lightweight encoding architecture is a set of feature maps that are relatively smaller in spatial dimensions but still retain important visual information of the image.

Subsequently, these feature maps are fed into an improved Pyramid Visual Transformer (PVT) module (Wang et al., 2021). The original PVT module adopts a hierarchical Transformer architecture capable of processing feature maps at different scales, achieving the extraction and fusion of multi-scale features. Within each layer, feature maps are processed through a self-attention mechanism, allowing the model to capture global contextual information and enhance its understanding of image details. Through progressive processing, the PVT module gradually increases the resolution of feature maps while integrating richer contextual information, ultimately producing a set of high-resolution, semantically rich feature maps. Figure 2(b) visualizes the computational logic at each level of the PVT used in this study, which consists of a Patch Embedding layer and a Transformer encoding layer at the L_i level. It is important to note that since this study involves making predictions on HR images, to avoid GPU memory overflow due to excessive computation, the original Transformer encoding layer's multi-head attention layer (MHA) is replaced with a Spatial Reduction Attention layer (SRA). Similar to MHA, SRA still takes Query (Q), Key (K), and Value (V) as inputs. The difference lies in that SRA reduces the spatial dimensions of K and V through spatial reduction operations before the attention operation, thereby significantly enhancing computational efficiency.

In summary, by adopting a feature extraction encoder that combines a lightweight encoder with the visual Transformer, the advantages of both architectures in computational efficiency and feature expression capability are effectively integrated, providing an efficient solution for the extraction of fine-grained crack features.

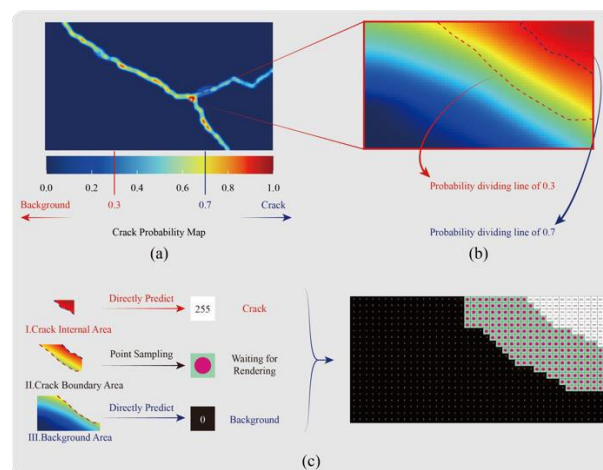


Figure 3. Visual demonstration of the position for the guided rendering point used in the inference phase.

3.2 Point Rendering-based Decoder

To fully decode the fine-grained crack information from the deep semantic feature maps captured by the encoder, the authors improved upon the original PointRend and proposed a decoder that composed of a rendering point guidance branch and a point-rendering based fine-grained prediction head. For the fine-grained prediction head, the authors adopted the same architecture as PointRend because the use of MLPs in rendering technology offers computational efficiency advantages due to weight sharing and point-by-point prediction compared to traditional CNNs. Regarding the branch that guides the rendering points, considering that the rendering point guidance strategy in the PointRend architecture was designed for traditional large-size natural scene targets and is not suitable for tiny crack targets with elongated topological structures, this study conducted specialized customization.

We specifically designed a rendering point guidance strategy based on the probability map for the inference phase to ensure that the model can effectively concentrate computational resources on difficult-to-predict tiny cracks and crack boundary areas. Specifically, we aimed to achieve efficient boundary rendering point guidance through the refined probability map. As shown in Figure 2(a), the probability map is primarily obtained through forward propagation calculations involving a lightweight encoder followed by two convolution blocks, without the need for thresholding or other post-processing steps to determine final category assignments. Therefore, this process is more direct and efficient computationally compared to generating coarse segmentation masks. More importantly, the probability map can reflect the probability of pixels belonging to each category, rather than a simple categorical affiliation. This probabilistic information provides a continuous measure of confidence for each pixel, rather than categorically assigning pixels in a binary manner, thereby preserving more uncertainty and subtle differences about the image area, which is crucial for understanding the nuances and complex scenes within the image. Since crack segmentation is inherently a binary classification task, based on the characteristics of binary classification tasks, this study divided areas on the probability map into three parts: areas with probabilities close to 0 are considered definite background pixels, areas with probabilities close to 1 are considered definite crack regions, and areas with probabilities fluctuating around 0.5 are considered indeterminate areas by the model. These indeterminate areas often consist of pixels with intensity or contrast levels between cracks and backgrounds, primarily concentrated around tiny crack branches and crack boundary areas. Based on the above division principles, only areas on the probability map with probabilities fluctuating around 0.5 undergo refined rendering point sampling during the inference process. Regarding areas on the probability map with probabilities close to 0 and 1, which are considered easily recognizable pixels, they will be directly mapped as background and crack pixels on the prediction mask without further computation, considering the difficulty of recognition and computational efficiency. To visually demonstrate the rendering point guidance method during the inference phase, Figure 3 visualizes the probability map for a randomly selected crack sample. It can be clearly seen on the probabilistic heatmap that the probabilities of the background and main crack areas are concentrated around 0 and 1, respectively. In boundary areas and tiny crack regions, due to issues such as manual annotation errors and insignificant color differences, the probabilities of pixels on the heatmap fluctuate around 0.5. For this reason, this study sets the probability interval for these hard-to-identify boundary and tiny crack pixels between 0.3-0.8. In the subsequent refinement rendering stage, only pixels with probabilities between 0.3-0.8 undergo refined inference. The parameter settings involved in the rendering point guidance strategy during the training and inference phases will be detailed in Section 3.2.

4. Experiments

4.1 Datasets

Gathering substantial crack image datasets is vital for the development and assessment of our segmentation method. Despite the existence of various datasets like Stone331, Bochum CrackDataSet, Deepcrack537, and CRKWH100 (Quan et al., 2023) for civil infrastructure analysis, the intricacies of acquiring and annotating crack images limit dataset sizes and resolution, typically capping at 500 images with dimensions under 600×800 . Such constraints risk model overfitting and lessen computational demand due to lower resolution. To bolster model adaptability across varying real-world cracks, this study amalgamated images from AigleRN, CrackTree260, and Crack500 (Ai et al., 2023), adjusting all to a uniform resolution of 256×256 for cohesive model training and evaluation. This resulted in 800 adjusted images distributed across training, validation, and testing phases with a ratio of 0.6:0.2:0.2 to rigorously assess model efficacy.

Given that higher-resolution images demand more extensive downsampling to fit within GPU memory constraints during inference, they are inherently more susceptible to indistinct boundary delineations than their lower-resolution counterparts, necessitating enhanced refinement. To address this, the study amassed a HR crack image dataset for thorough model assessment. In Changsha's urban context, diverse structural elements such as walls, roads, and foundations were selected for HR image acquisition, employing both tripod-mounted and handheld Nikon D5300 cameras for crisp imagery. A total of 300 6K RAW crack images were compiled. To mitigate the skew from disproportionate sample sizes, crack regions in these images were meticulously cropped,

yielding 60 images across 2K, 4K, and 6K resolutions. Each cropped image was subsequently subjected to detailed annotation for precise model evaluation.

4.2 Evaluation Index

Two commonly used metrics, namely Intersection over Union (IoU) and the Dice similarity coefficient (Dice), were selected to quantitatively evaluate the experimental results. Furthermore, to highlight the performance of the proposed method in boundary areas, the Mean Boundary Accuracy (mBA) introduced in CascadePSP (Cheng et al., 2020) was also used as a metric. The core concept of mBA involves calculating the IoU between the Ground Truth (GT) and the predicted mask within the boundary area.

4.3 Implementation Details

Hardware equipment: All the crack segmentation networks mentioned in this document were completed training under the Ubuntu 18.04 system, using the Pytorch 1.8.0 version as the deep learning framework, with an Intel i7-8700k processor, 32GB of RAM, and an NVIDIA GeForce RTX 3090 GPU with 24 GB of VRAM.

Hyperparameters: To ensure the global optimum of the loss function is found during the training process, Adam, which combines the advantages of momentum and RMSprop, was chosen as the optimizer for the model, with momentum set at 0.9 and weight decay at 1×10^{-4} . The maximum number of training epochs on the low-resolution open-source crack image dataset was set to 800. The batch size was set at 8, with an initial learning rate of 0.001, decaying by 0.0001 every 10 training epochs. After completing the initial training, the same hyperparameter configuration was used to fine-tune the model's performance for an additional 200 epochs using onsite collected concrete crack images, to obtain the final model for subsequent crack segmentation tasks.

4.4 Ablation Study

Ablation Study for the SRA Enhanced PVT: The effectiveness of the improved Pyramid Transformer introduced at the end of the encoder was tested on the test set. Specifically, four typical encoders including ResNet50, DenseNet, MobileNet, Vision Transformer (ViT), and the original Pyramid Vision Transformer (PVT) were selected for performance comparison. The segmentation results of the rendering models with different encoders are summarized in Table 1.

Table 1. Performance comparison of models with different crack feature enhancement backbones introduced during the encoding phase on the test dataset

Feature Extraction Backbone	IoU(%)	mBA(%)	Dice(%)	Parameter (M)
ResNet50	75.48	73.17	86.03	25.6
DenseNet	76.21	73.82	86.50	8.02
MobileNet	73.06	72.39	84.44	4.23
ViT	77.47	74.76	87.30	86.37
PVT	77.81	78.25	87.52	24.51
SRA Enhanced PVT	77.91	78.33	87.58	18.76

By conducting a parallel comparison of the performance of five groups of models utilizing different crack feature enhancement encoders, it can be observed that the networks adopting the Transformer architectures generally outperform those built with CNN architectures, with average improvements in IoU, mBA, and DICE reaching 2.81%, 3.99%, and 1.81%, respectively. This is because, compared to the other three types of models built on the CNN, Transformers can consider all positions within crack images simultaneously through their inherent self-attention mechanism, enabling the model to capture a broader context of crack information. This assists the model in understanding the connections between global and local information of crack features, thereby effectively enhancing the model's crack recognition performance. Further comparison among the three groups of models using the Transformer architecture reveals that the two models adopting PVT outperform the ordinary ViT in recognition performance. Notably, the most significant improvement is observed in mBA, indicating that the Pyramid Transformer has significantly strengthened its ability to capture crack edges and tiny crack details in the global image. This is attributed to the Pyramid Transformer's fusion of multi-scale features across different levels, allowing features from various scales to complement each other. Thus, ensuring the model can utilize both local detail information and global deep contextual semantic information to enhance the representation of crack edges. Lastly, comparing the two groups that employed the PVT architecture, it is evident that after implementing the SRA pro-posed by this study to reduce the dimensions of the original multi-head attention mechanism, the model's number of parameters decreased by nearly 24%, while maintaining high recognition accuracy.

Ablation Study for the Probability Map-based Rendering Point Guidance: To minimize the computational resource consumption during the inference process while maintaining the required inference accuracy, it is necessary to determine a reasonable probability range for areas with uncertain prediction outcomes where the probabilities are concentrated around 0.5. This is because a larger probability range means that more sampling points need to be refined, which, while increasing accuracy, also significantly raises the computational

redundancy of the inference process; conversely, a smaller probability range can enhance inference speed but may result in many tiny cracks and boundary details not being effectively refined, thus severely affecting the refinement level of prediction outcomes. Therefore, it is essential to establish suitable threshold probability values to reasonably control the boundary range. Specifically, two probability values are selected: the critical probability value α between the background and crack boundary areas, and the critical probability value β between the crack boundary and crack internal areas.

For the critical probability value α , this study set three different probability parameters: 0.2, 0.3, and 0.4; for the critical probability value β , three different probability parameters were also set: 0.6, 0.7, and 0.8. These two sets of three different critical probability values define nine different boundary area probability ranges. Table 2 presents the statistical inference results on the test set for models using sampling with these nine different probability ranges.

Table 2. Performance comparison between the traditional decoder and the proposed rendering-based fine-grained prediction head across models trained with various parameterized feature point guidance strategies

Set No.	Probability range for background area	Probability range for boundary area	Probability range for crack internal area	IoU(%)	mBA(%)	Dice(%)
1	(0.0,0.2)	(0.2,0.6)	(0.6,1.0)	77.16	80.39	87.11
2	(0.0,0.2)	(0.2,0.7)	(0.7,1.0)	78.59	81.77	88.01
3	(0.0,0.2)	(0.2,0.8)	(0.8,1.0)	80.04	84.56	88.92
4	(0.0,0.3)	(0.3,0.6)	(0.6,1.0)	78.72	82.34	88.10
5	(0.0,0.3)	(0.3,0.7)	(0.7,1.0)	82.35	86.73	90.32
6	(0.0,0.3)	(0.3,0.8)	(0.8,1.0)	84.24	93.78	91.45
7	(0.0,0.4)	(0.4,0.6)	(0.6,1.0)	78.06	81.32	87.68
8	(0.0,0.4)	(0.4,0.7)	(0.7,1.0)	81.80	83.60	89.99
9	(0.0,0.4)	(0.4,0.8)	(0.8,1.0)	83.03	87.24	90.71

Table 3. Performance comparison between the PointRend guided by various coarse masks and the proposed probability map-guided architecture

Refinement Segmentation Architecture	Source of the Boundary Sampling Guidance	Coarse Segmentation			Refined Segmentation			
		Accuracy			Accuracy			
		IoU (%)	mBA (%)	Dice (%)	IoU (%)	mBA (%)	Dice (%)	
PointRend	Coarse Mask Guidance	FCN-18	68.73	64.36	81.47	77.90	79.79	87.58
		UNet	70.15	67.28	82.46	78.59	80.03	88.01
		DeepLabV3+	73.87	71.12	84.97	79.92	81.03	88.83
		RefineNet	72.10	70.70	83.79	79.45	80.78	88.54
		Swin Transformer	74.85	72.89	85.61	80.72	81.36	89.33
Ours	Probability Map Guidance	Probability interval \in [0.3,0.8]	/			84.24	93.78	91.45

Table 2 shows that sets 4, 5, and 6 (i.e., the experimental groups with the background probability range set between 0.0 and 0.3 achieved relatively superior IoU, Dice, and mBA scores. This is because, compared to the sets with the background probability range set between 0.0 and 0.4, these three parameter settings encompass a wider background sampling area, thereby facilitating the repair of tiny crack details that were not detected in the background. At the same time, the sets with the background probability range set between 0.0 and 0.2 classified too many pixels, which should belong to the boundary area, as background pixels. This resulted in an insufficient number of rendering points guided towards the ambiguous boundary area, unable to fully repair the crack boundary details in that area, and therefore achieved the relatively lower mBA scores. Furthermore, comparing sets 4, 5, and 6 reveals that the highest accuracy in model inference was achieved when the probability range for the crack boundary area was set to its maximum, i.e., when the probability range was between 0.3 and 0.8, with IoU, Dice, and mBA reaching 84.24%, 93.78%, and 91.45%, respectively. This is because the crack main body area, compared to the background and edge areas, is considered a simple sample with a high prediction probability (often exceeding 80% confidence), thus not requiring a too wide probability range. However, as the boundary area serves as a transition zone between the background and the crack main body, where pixel color and contrast often present indistinct conditions, leading to significant fluctuations in prediction probability, a relatively wide probability interval range is needed. Ultimately, the sampling parameter configuration in set 4 was adopted as the

optimal inference phase sampling parameter to guide the model in providing sufficient guidance for rendering points during the inference stage. Indeed, the experimental results also indirectly confirm that the main reason for inadequate crack segmentation accuracy concentrates on the ambiguous boundary area, with the probability range on the coarse segmentation probability map roughly concentrating between 0.3 and 0.8.

4.5 Comparison of the Performance of Models that use the Heatmap and Coarse Segmentation for Guiding Rendering Points during the Inference Phase

To elucidate the benefits of the decoding architecture, which utilizes a probability map for rendering point guidance, over the original PointRend architecture that guided by the coarse segmentation mask, a comparative analysis of their performance was conducted based on the test dataset. Specifically, we selected five mainstream deep learning segmentation architectures with varying degrees of segmentation precision, including FCN-18, UNet, DeepLabV3+, RefineNet, and Swin Transformer, as the networks generating coarse segmentation masks required for predictions by the original PointRend architecture. The network proposed in this study, which guides rendering points with the probability map, participated in the performance comparison using optimal parameters obtained from prior ablation experiments. It is noteworthy that all the coarse segmentation architectures and the refinement segmentation networks were trained under the same configurations in the same deep learning framework with default optimal parameters. Additionally, when making predictions using the trained coarse segmentation models, all HR images were scaled down to have their longer side be 900 pixels to prevent GPU memory overflow caused by excessively high original resolutions.

Experimental results, as shown in Table 3., indicate significant differences in the Coarse segmentation accuracy generated by various coarse segmentation architectures, as observed from rows 2 to 5 from the top. The disparities in IoU, mBA, and Dice scores range from 6.12%, 8.53%, to 4.15%, respectively, from the lowest accuracy with FCN-18 to the highest accuracy with Swin Transformer. However, after applying the original PointRend model for refinement, the differences in refined prediction results become less pronounced, with all five sets of experimental results fluctuating within the ranges of $79.31 \pm 1.4\%$ for IoU, $88.46 \pm 0.79\%$ for mBA, and $88.46 \pm 0.88\%$ for Dice. These findings demonstrate that the original PointRend architecture is in-deed independent of specific coarse segmentation masks, showing good robustness to coarse-grained crack features from different sources. Nevertheless, comparing the final experimental results with the best refinement predictions guided by coarse segmentation masks generated by Swin Transformer in the PointRend group reveals that the method using the probability map for rendering point guidance further improves the accuracy of segmentation results. Notably, the most significant improvement is observed in mBA, more than doubling the increases in IoU and Dice, reaching 12.42%. This outstanding robust performance is largely due to the probability map's ability to reflect the probability of pixels belonging to each category, rather than merely showing simple category membership like coarse segmentation masks. This probabilistic information provides a continuous confidence measure for each pixel, rather than categorically assigning pixels in a binary fashion, thereby preserving ample information on tiny cracks and crack boundaries. Such details allow for the discovery and refinement rendering of these pixels during the inference stage through point-by-point reasoning with the MLP. To further substantiate the validity of these conclusions, Figure 4 visualizes the prediction results under all comparative methods for five randomly selected HR crack images from the test set. The visualizations demonstrate that the predictive masks obtained by the probability map-guided rendering point method outperform those obtained by any coarse segmentation-guided method in recognizing crack edges and tiny cracks, thereby further validating the conclusions drawn from the quantitative results.

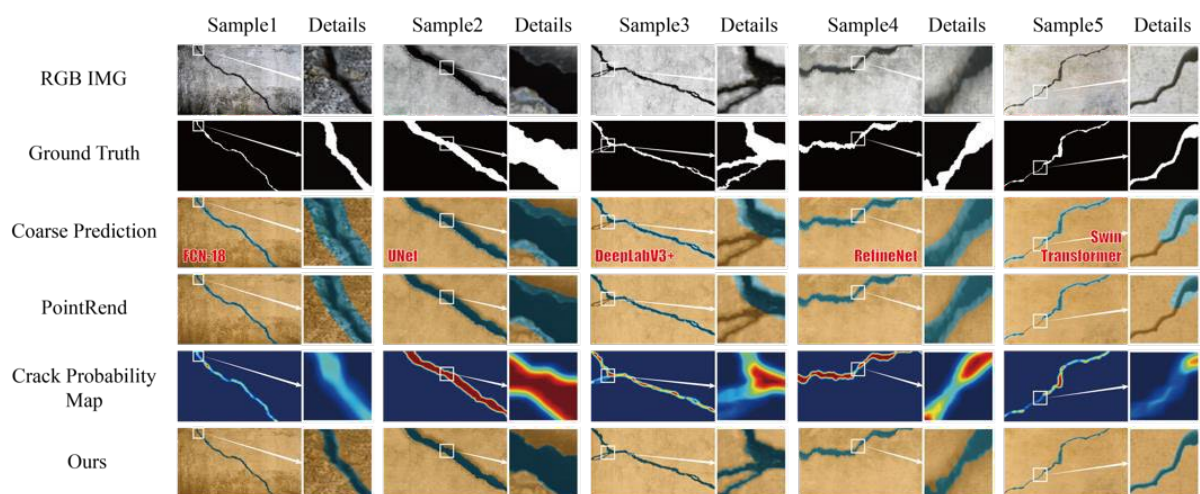


Figure 4. Visualization of fine-grained segmentation results of the PointRend architecture guided by different coarse segmentation masks and the proposed method guided by the probability map.

5. Conclusion

This study pioneers the application of rendering techniques from computer graphics to HR crack image segmentation, enhancing the fine-grained segmentation of crack images through two novel modifications. Initially, a Pyramid Transformer Block with Spatial Relationship Attention (SRA) is implemented within the encoding architecture for effective extraction of deep, detailed crack feature maps. Additionally, a tailored rendering point guidance strategy in the decoder concentrates computational resources on the crack edges and tiny cracks, increasing pixel recognition accuracy for these hard samples. Demonstrated performance on HR crack images confirms this approach as the latest benchmark in the field.

Future endeavors will focus on model pruning and quantization for drone-based deployment, offering bridge maintenance a more dependable crack detection tool for real-world applications, and extending its utility to hydropower projects for surface defect identification.

ACKNOWLEDGMENTS

This work is supported by the Europe Commission project D-HYDROFLEX (No. 101122357), Europe Commission project INHERIT (No. 101123326), the National Natural Science Foundation of China (No. 52278177) and the National Key Research and Development Program of China (No. 2023YFC3806800)

REFERENCES

- Ai, D., Jiang, G., Lam, S.-K., He, P. & Li, C. (2023), Computer Vision Framework for Crack Detection of Civil Infrastructure—a Review, *Engineering Applications of Artificial Intelligence*, 117, 105478.
- Bang, S., Park, S., Kim, H. & Kim, H. (2019), Encoder–Decoder Network for Pixel - Level Road Crack Detection in Black - Box Images, *Computer - Aided Civil and Infrastructure Engineering*, 34(8), 713-727.
- Barron, J. T., Mildenhall, B., Tancik, M., Hedman, P., Martin-Brualla, R. & Srinivasan, P. P. (2021), Mip-Nerf: A Multiscale Representation for Anti-Aliasing Neural Radiance Fields, *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Montreal, Canada, pp. 5855-5864.
- Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K. & Yuille, A. L. (2017), Deeplab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected Crfs, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4), 834-848.
- Cheng, H. K., Chung, J., Tai, Y.-W. & Tang, C.-K. (2020), Cascadepsp: Toward Class-Agnostic and Very High-Resolution Segmentation Via Global and Local Refinement, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, USA, pp. 8890-8899.
- Chu, H., Wang, W. & Deng, L. (2022), Tiny - Crack - Net: A Multiscale Feature Fusion Network with Attention Mechanisms for Segmentation of Tiny Cracks, *Computer - Aided Civil and Infrastructure Engineering*, 37(14), 1914-1931.
- Deng, J., Singh, A., Zhou, Y., Lu, Y. & Lee, V. C.-S. (2022), Review on Computer Vision-Based Crack Detection and Quantification Methodologies for Civil Structures, *Construction and Building Materials*, 356, 129238.
- Dias, P. A. & Medeiros, H. (2018), Semantic Segmentation Refinement by Monte Carlo Region Growing of High Confidence Detections, *Asian Conference on Computer Vision*, Springer, Perth, Australia, pp. 131-146.
- Guo, F., Qian, Y., Liu, J. & Yu, H. (2023), Pavement Crack Detection Based on Transformer Network, *Automation in Construction*, 145, 104646.
- Huyan, J., Li, W., Tighe, S., Xu, Z. & Zhai, J. (2020), Cracku - Net: A Novel Deep Convolutional Neural Network for Pixelwise Pavement Crack Detection, *Structural Control and Health Monitoring*, 27(8), e2551.
- Kellnhofer, P., Jebe, L. C., Jones, A., Spicer, R., Pulli, K. & Wetzstein, G. (2021), Neural Lumigraph Rendering, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, virtual online, pp. 4287-4297.
- Khan, S., Naseer, M., Hayat, M., Zamir, S. W., Khan, F. S. & Shah, M. (2022), Transformers in Vision: A Survey, *ACM computing surveys (CSUR)*, 54(10s), 1-41.
- Kirillov, A., Wu, Y., He, K. & Girshick, R. (2020), Pointrend: Image Segmentation as Rendering, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9799-9808.
- Lin, G., Milan, A., Shen, C. & Reid, I. (2017), Refinenet: Multi-Path Refinement Networks for High-Resolution Semantic Segmentation, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, USA, pp. 1925-1934.
- Lin, G., Shen, C., Van Den Hengel, A. & Reid, I. (2016), Efficient Piecewise Training of Deep Structured Models for Semantic Segmentation, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, USA, pp. 3194-3203.
- Liu, W., Huang, Y., Li, Y. & Chen, Q. (2019), Fpcnet: Fast Pavement Crack Detection Network Based on Encoder-

- Decoder Architecture, pp. arXiv:1907.02248.
- Mei, Q., Gül, M. & Azim, M. R. (2020), Densely Connected Deep Neural Network Considering Connectivity of Pixels for Automatic Crack Detection, *Automation in Construction*, 110, 103018.
- Peng, C., Zhang, X., Yu, G., Luo, G. & Sun, J. (2017), Large Kernel Matters--Improve Semantic Segmentation by Global Convolutional Network, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, USA, pp. 4353-4361.
- Qu, Z., Li, Y. & Zhou, Q. (2022), Crackt-Net: A Method of Convolutional Neural Network and Transformer for Crack Segmentation, *Journal of Electronic Imaging*, 31(2), 023040-023040.
- Quan, J., Ge, B. & Wang, M. (2023), Crackvit: A Unified Cnn-Transformer Model for Pixel-Level Crack Extraction, *Neural Computing and Applications*, 1-17.
- Ronneberger, O., Fischer, P. & Brox, T. (2015), U-Net: Convolutional Networks for Biomedical Image Segmentation, *International Conference on Medical Image Computing and Computer-assisted Intervention*, Springer, Munich, Germany, pp. 234-241.
- Shamsabadi, E. A., Xu, C. & Dias-da-Costa, D. (2022), Robust Crack Detection in Masonry Structures with Transformers, *Measurement*, 200, 111590.
- Sony, S., Dunphy, K., Sadhu, A. & Capretz, M. (2021), A Systematic Review of Convolutional Neural Network-Based Structural Condition Assessment Techniques, *Engineering Structures*, 226, 111347.
- Tewari, A., Thies, J., Mildenhall, B., Srinivasan, P., Tretschk, E., Yifan, W., Lassner, C., Sitzmann, V., Martin - Brualla, R. & Lombardi, S. (2022), Advances in Neural Rendering, *Computer Graphics Forum*, Wiley Online Library, pp. 703-735.
- Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M. & Wang, X. (2020), Deep High-Resolution Representation Learning for Visual Recognition, *IEEE transactions on pattern analysis and machine intelligence*, 43(10), 3349-3364.
- Wang, Q., Wang, Z., Genova, K., Srinivasan, P. P., Zhou, H., Barron, J. T., Martin-Brualla, R., Snavely, N. & Funkhouser, T. (2021), Ibrnet: Learning Multi-View Image-Based Rendering, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, vitural online, pp. 4690-4699.
- Wang, W., Xie, E., Li, X., Fan, D.-P., Song, K., Liang, D., Lu, T., Luo, P. & Shao, L. (2021), Pyramid Vision Transformer: A Versatile Backbone for Dense Prediction without Convolutions, *Proceedings of the IEEE/CVF International Conference on Computer Vision*, vitural online, pp. 568-578.
- Wu, G., Liu, Y., Fang, L. & Chai, T. (2021), Revisiting Light Field Rendering with Deep Anti-Aliasing Neural Network, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9), 5430-5444.
- Xie, X., Cai, J., Wang, H., Wang, Q., Xu, J., Zhou, Y. & Zhou, B. (2022), Sparse - Sensing and Superpixel - Based Segmentation Model for Concrete Cracks, *Computer - Aided Civil and Infrastructure Engineering*, 37(13), 1769-1784.
- Zhang, C., Lin, G., Liu, F., Yao, R. & Shen, C. (2019), Canet: Class-Agnostic Segmentation Networks with Iterative Refinement and Attentive Few-Shot Learning, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, USA, pp. 5217-5226.
- Zhang, J., Lu, C., Wang, J., Wang, L. & Yue, X.-G. (2019), Concrete Cracks Detection Based on Fcn with Dilated Convolution, *Applied Sciences*, 9(13), 2686.
- Zhang, X., Rajan, D. & Story, B. (2019), Concrete Crack Detection Using Context - Aware Deep Semantic Segmentation Network, *Computer - Aided Civil and Infrastructure Engineering*, 34(11), 951-971.
- Zheng, S., Jayasumana, S., Romera-Paredes, B., Vineet, V., Su, Z., Du, D., Huang, C. & Torr, P. H. (2015), Conditional Random Fields as Recurrent Neural Networks, *Proceedings of the IEEE International Conference on Computer Vision*, Santiago, Chile, pp. 1529-1537.

Maintenance management - a module in asset management in a medium-sized German city as an example

Prof. Dr.-Ing. Alexander Buttgereit ¹, and Stefan Gomolluch ²

1) Ph.D., Prof., Department Construction, Geoinformation, Healthcare Technology, Jade University of Applied Sciences, Oldenburg, Germany. Email: alexander.buttgereit@jade-hs.de

2) MBA., Department of Mobility and Civil Engineering, City of Münster, Münster, Germany. Email: Gomolluch@stadt-muenster.de

Abstract: Since the development of public administration management in the 1980s into the New Management Model, the principles of asset management have found their way into the management of public administrations. This enabled the already existing systematic maintenance management of roads to be integrated as an element of asset management. Systematic road maintenance today ensures the provision of a well-functioning infrastructure. The requirements for maintenance management - especially for municipal infrastructure - are becoming increasingly complex. In addition to the purely technical view, the commercial view and life cycle cost considerations have also been incorporated into maintenance management. The available financial resources should be used optimally in such a way as to achieve the useful lives of the assets, ensure the value retention of the fixed assets and implement the new infrastructure requirements. When renewing existing infrastructure today, it is therefore necessary to address issues that will become relevant tomorrow. These include, for example, an ageing society, increased e-mobility, a child and disability-friendly society, flood protection, reduced noise and a reduction in air pollution. These and other aspects are therefore integrated into the existing management system and must also be considered during maintenance. Municipal asset management (AM) is a suitable controlling method for this. This paper intends to show how modern maintenance management of municipal infrastructure as an element of asset management can be achieved by changing the way it is viewed and approached, without generating additional costs in the medium term.

Keywords: Asset management, Maintenance, City, Road infrastructure

1. INTRODUCTION – WHAT IS THE ADDITIONAL BENEFIT OF ASSET MANAGEMENT?

The Pavement Management System (PMS) or systematic road maintenance is an important tool for road authorities to choose the right measure at the right time from a road construction perspective and to estimate the short and medium-term financial requirements. Furthermore, it makes it possible to monitor the success of the measures implemented and the maintenance strategy over longer periods of time (Schmuck, 1987). Long-term considerations and the consideration of changing requirements are not or hardly possible in PM systems. It therefore has a strong operational focus and is only suitable for the management level of an organization to a limited extent.

Asset management is intended to support the management at all levels of an organization in carrying out their leadership and management tasks. It is used for task- and result-oriented controlling under economic aspects based on targets, metrics and key figures. The objectives and requirements defined in asset management are to be achieved through measures and projects. In contrast to project management, asset management makes it possible to consider an asset from the initial planning stage throughout its entire life cycle. While quality management describes the process flow for completing tasks in accordance with predefined goals and standards, asset management is concerned with the organization of tasks in accordance with the predefined goals and standards. Asset management therefore means (strategic) control over the life cycle of an organization's assets.

The requirements for asset management are regulated in E DIN ISO 55001:2017-02. According to DIN ISO 55000 (DIN, 2016), the benefits of asset management over the life cycle can include the following:

- Improved economic performance (maintaining asset values despite cost reduction),
- fact-based decisions on asset-related investments (easier decision-making by weighing up costs, risks, opportunities and performance),
- planned and controlled handling of risks (reducing financial losses, improving occupational health and safety, improving image),
- improved performance and earnings (ensuring performance leads to better services and products),
- proven compliance with regulations,
- enhanced reputation (customer satisfaction increases),
- improved sustainability,
- increased efficiency (the review and the resulting improvement increases effectiveness and leads to the achievement of the organizational goal).

An asset management system can also be used to manage and control the assets of a municipal road infrastructure through a series of business processes for decision-making that promote continuous improvement in

maintenance management.

The following characteristics are therefore decisive for a functioning maintenance management system as part of asset management:

- The nature and purpose of the organization (organizational goals),
- the inventory of all infrastructure assets,
- Knowledge of the age, condition and changes to the assets,
- Availability of information for carrying out a life cycle cost analysis,
- Information to perform risk management analysis,
- information on the development of the financial plan to support investments and
- the development of investment strategies,

to manage the road network over its entire life cycle. Furthermore, a repeatable definition of characteristics is essential to the overall assessment process. It is also necessary to define what is to be achieved in the future.

In order to be able to manage efficiently in terms of AM, it is important to define targets, measures for target fulfillment and associated key performance indicators (KPIs). The values, metrics and KPIs can be used to objectively, quantitatively and qualitatively assess whether the strategic objectives are being achieved. When developing KPIs, the collection of new data should be minimized and the consideration of already collected data should be maximized. To monitor and document success, an annual review is prepared at the end of the year to identify which processes still need to be improved. There are two things to be considered regarding the informative value of key performance indicators. First, the data must be tested to check whether it is suitable for the measurements and calculations. Second, the validity of the results must be checked to ensure that they reflect actual performance.

A performance framework should consider who is affected by the performance of the installations. Here, a distinction is made between three broad groups: external stakeholders (usually more interested in what happens on the network than the average customer; have an interest in the transportation system), customers (anyone who uses parts of the transportation network) and partners (stakeholders who share strategic objectives with the transportation agencies). Partners can help deliver or fund projects.

2. HOW CAN ASSET MANAGEMENT BE SET UP?

Based on the common management function circle (see Figure 1), management models for road maintenance can be developed in a similar way, in which the various tasks for the safety and performance, construction and maintenance of road pavements are combined and linked.

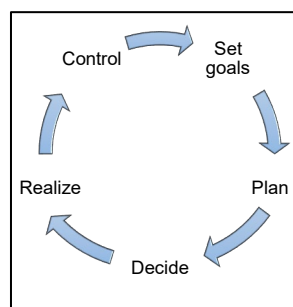


Figure 1. Schematic illustration of the management circle (Münster b, 2020)

Different organizational forms are conceivable depending on the administrative structure, tasks, task weighting and standards. It is therefore a misconception that the asset management standard describes an organizational form or structure. The series of international standards ISO 55000 Asset Management describe the international standard of asset management. They have not yet been implemented in German standards or regulations for road maintenance management. The asset management system is well aligned with the life cycle considerations of the New Control Model for Municipalities (NSM) and New Municipal Financial Management (NKF). This can be used as a guide when setting up an AM system.

In Germany, current systems for systematic road maintenance are based on condition-based target determination. This means that the planning of maintenance activities over the life cycle is geared towards optimizing the condition and achieving the service life.

The management of road maintenance is a special “planning task” in which both the strategic decisions on maintenance measures or measure strategies are made network-wide and the preparation of a maintenance program as well as the preparation and implementation of maintenance measures are carried out on an object-by-

object basis at the operational level. These two levels are linked by a tactical level, which supports the decision-making process to achieve the strategic objectives (Fig. 2).

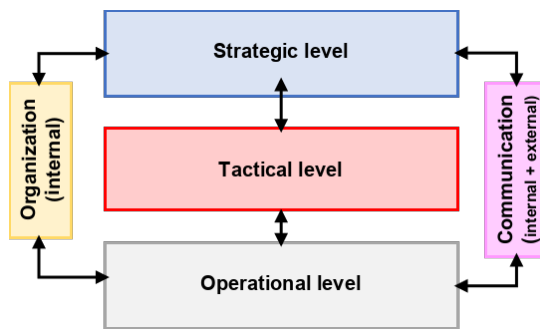


Figure 2. Decision-making levels in asset management AM (Buttgereit et al., 2016)

The basis for this is the road network that has been recorded and prepared in a corresponding database. In order to better integrate financial aspects into construction-related decisions in maintenance management, asset management should also take into account information from asset accounting.

To be able to control efficiently in terms of asset management, it is important to define goals, measurements and key performance indicators. Long-term and short-term management options can be developed based on the defined goals and key performance indicators, which are explained in more detail using an example from the area of finance.

Long-term management (strategic level) should create intergenerational justice, i.e. not place a burden on future generations but be sustainable. To achieve this, it is necessary to achieve an even distribution of resource consumption over the life cycle.

Short-term management (operational level) is primarily concerned with achieving a balanced annual result, i.e. the resource consumption of the financial year is covered by the resource revenue of the same financial year. Efficiency control can be carried out, for example, through time comparisons, operational comparisons or target/actual comparisons. The short-term control of the budget should be carried out with the help of target agreements as part of contract management. The prerequisite for this is that product information on monetary values, quantities and qualities is known, which is mapped and controlled in a management system.

3. APPROACH TO ASSET MANAGEMENT IN MÜNSTER

Considering the previous principles, the asset management of the “Construction and Road Maintenance” department of the City of Münster's Office for Mobility and Civil Engineering will be explained below. For this purpose, it is necessary to determine the organizational goals of the company. The objectives of the Office for Mobility and Civil Engineering, which is responsible for the maintenance of municipal roads, are important. The desired sustainable maintenance management as part of asset management is based on all topics that are relevant for a functioning, future-oriented infrastructure and should consider the three columns of sustainability, ecology, economy and socio-cultural aspects.

Ecological sustainability essentially means actively influencing environmentally harmful factors from the construction and operation of road networks. Influence can be exerted, for example, through traffic avoidance, multimodal offers or the optimization of transport chains, the conservation of natural resources during construction or the avoidance of traffic jams.

The economic sustainability of road maintenance must therefore be geared towards economic efficiency: It must be ensured that the resources expended (investment and maintenance expenditure) are matched by appropriate social and economic benefits (e.g. in the form of time or cost savings or changed service life and reduced maintenance expenditure as part of value retention).

Social sustainability means the organization of socially acceptable transport that can satisfy the existing demand for mobility. In particular, the supply of the population with daily needs, but also education or healthcare, must be taken into account. In addition to the needs of MIT (motorized private transport), the needs of those affected, e.g. residents, must also be integrated.

This approach and perspective thus follow the triple-bottom-line approach of asset management and is cross-organizational, as many organizations have an impact on the asset “street” through their actions. For this reason, numerous strategy papers of the City of Münster must be considered in AM and maintenance management.

Thanks to the implementation of asset management, it is possible to manage road maintenance in a target- and task-oriented manner from an economic perspective. In the medium term, this means a significant increase in the quality of road maintenance. Instruments are thus available to achieve specified or self-imposed goals through active planning, control, implementation and monitoring of our tasks consistently and sustainably over time.

3.1 Derivation of strategic goals at various organizational levels

A central point in the development of asset management is the definition of objectives at a strategic, tactical and operational level. In the public sector, laws and regulations generally form the basis or foundation for action. They often already contain targets and standards from which measurements can then be derived. Consequently, the content of the target system is already fed by these requirements and can be expanded to include local targets that are based on the applicable regulations (legal and technical) (Fig. 3).



Figure 3. Goal levels of the Office for Mobility and Civil Engineering of the City of Münster (Münster b, 2020)

The strategic goals of the Office for Mobility and Civil Engineering have been derived from the long-term corporate goals of the City of Münster and a vision has been formulated, which is visualized in Figure 4.

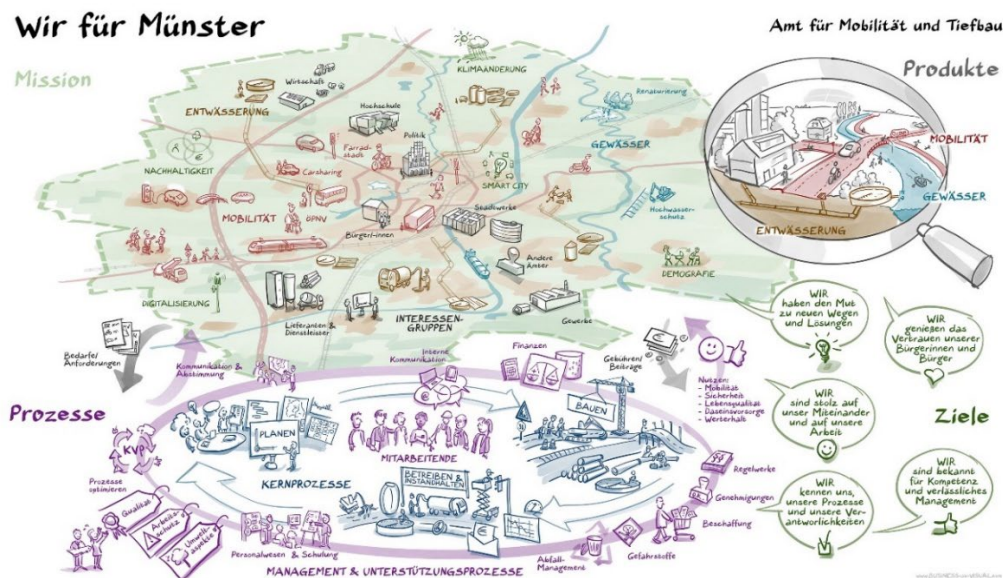


Figure 4. Vision and goals of the Münster Office for Mobility and Civil Engineering (Münster a, 2020)

The goals cover the tasks and responsibilities of the Office for Mobility and Civil Engineering Münster. In the sector of mobility, these include the provision of public traffic areas and facilities, i.e. the planning, construction, maintenance and financing of public roads, paths, squares, bridges, tunnels, noise barriers, lighting, traffic signal systems (traffic lights), parking ticket machines and the parking guidance system, as well as the sovereign tasks of the road construction authority.

Based on the vision and objectives of the department, the following three key objectives have been formulated in the asset management of the “Construction and Road Maintenance” department.

- *We preserve values.*
- *We manage 640 million euros of the city's assets.*
- *We create infrastructure and the conditions for the future.*

These key objectives can be specified in the following strategic goals and express selected requirements for the infrastructure, also in relation to the interested parties (external stakeholders, customers and partners):

Value preservation, mobility, safety, environment/sustainability, collaboration, communication and digitalization (Münster b, 2020).

On the one hand, the trick is to identify the relevant goals, formulate them clearly and focus on them. On the other hand, the quality of an AM is determined by its flexibility, i.e. how it deals with changes, new requirements or competing objectives.

3.2 Development of operational goals and measurements

The strategic objectives of the organizational unit were operationalized and provided with key performance indicators (KPIs) relevant to management (Buttgereit et al., 2016). Some examples are listed below:

Increasing the mobility of every individual is justified by the operational goals of keeping the infrastructure barrier-free for everyone and increasing access to public transportation. Citizens must be connected barrier-free from their front door and integrated into a barrier-free infrastructure at all stages of their lives. They must be free to choose their means of transportation. The goal implies local mobility, i.e. distances of up to two kilometers, such as journeys to public facilities. Measures include, for example, bus stops with raised kerbs, guidance systems for the blind, signal heads for the blind, traffic signal control, special bus lanes, access to electric mobility and bicycle parking facilities at bus stops.

The next goal is environmental compatibility/sustainability. The operational goals of reducing noise and acting more sustainably form the basis of this goal. This involves the careful use of resources, materials and energy as well as everything that surrounds people, affects them and influences their living conditions. Measures such as the construction of noise barriers, the use of noise-optimized asphalt, the use of recycled materials, crossings on main roads, a detour-free footpath network and speed reduction to reduce noise are considered.

Another goal is to maintain the traffic infrastructure. The necessary maintenance and renewal measures are to be carried out in accordance with the selected maintenance strategy. Measured variables such as assets, reinvestment rate, depreciation rate, asset utilization rate, average total and remaining useful life, unscheduled asset disposals and condition values are considered.

Another goal is the safe operation of the facilities. Road safety is an important asset that must be protected. The strategic goal of making the infrastructure safe is aimed at the safe use of the facilities for all road users. Safety is reflected in the operational objectives of increasing road safety, creating capacity, flood protection and speed reduction as a safety concept, road and structure monitoring, traffic light operational readiness and flood protection.

Finally, improving accessibility. This is not primarily about mobility. Rather, the focus here is on the surrounding communities and the associated commuters. They should be able to reach the city center easily and conveniently. The operational objectives of creating a multimodal offering in the city and increasing the general flow of traffic are in line with this strategic objective.

All objectives, key figures and metrics are described and brought together in the strategic asset management plan. This is operationalized in annual asset management plans and programs.

4. A HOLISTIC VIEW OF THE LIFE CYCLE

A holistic approach to asset management means infrastructure management in the sense of responsibility for the entire infrastructure or “asset portfolio” (“sewers, roads, bridges, traffic control and parking guidance systems”) over the life cycle by bringing together the technical, planning and commercial aspects towards intergenerational equity (Fig. 5). It is therefore a sustainable asset management system that is firmly integrated into the overarching control and communication framework of the organization.

Based on network-wide concepts of strategic maintenance planning (strategic level), work programs are derived at the tactical level and summarized in a concrete maintenance program. This concrete maintenance program is then implemented in the form of projects at the operational level.

Infrastructure management requires a proper combination of engineering planning, economic evaluation and political objectives. This addresses different decision-making levels. The political decision-making level decides on plans that set the framework for the road authorities' actions. The technical levels decide on prioritization and action plans based on engineering and economic considerations, while the operational level implements the specific maintenance program in the road network, considering implementation issues and traffic considerations. What all levels have in common is that such decisions and considerations can only be made based on meaningful and transparent data.

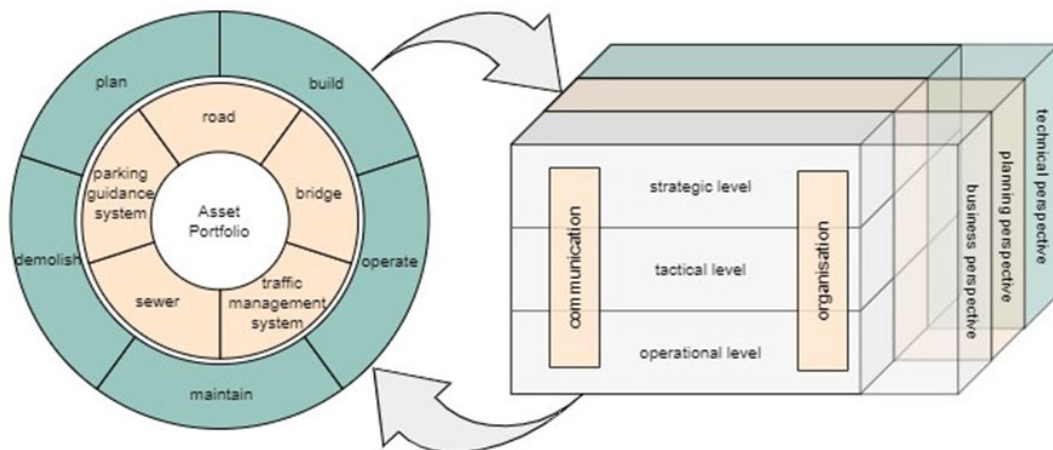


Figure 5. Continuous effect of the maintenance management organization (right) on the asset portfolio to be managed depending on the respective life cycle phase (left)

Various projects have been carried out in Münster in recent years, focusing on the road network. A concept for road maintenance management has been developed at a strategic, tactical and operational level and implemented in the LOGO road information database. The next step is to develop maintenance management for sewers, as significant synergies are seen in the coordinated maintenance planning of roads and sewers. Parallel to this, work is also being carried out on the maintenance management of structures and traffic facilities. Finally, the pumping stations, sewage treatment plants and bodies of water will be considered before finally defining a maintenance management system that covers all facilities. The concepts will initially be tested for small sub-networks. The data required for this, which is still missing or insufficient, will be successively completed. This will then be followed by expansion within the overall network, which in turn requires the collection, completion and continuation of the missing or insufficient data.

At a strategic level, the processes and their factual and content-related design in TIMM were examined under scientific supervision. The analysis was based on the current state of science and research and regarding the current state of the art in terms of quality assurance measures and with a focus on expected future requirements. The long-term financial requirement forecast at the grid level, considering various maintenance strategies over the life cycle, is a key component here (see (FGSV, 2019), (Buttgereit et. al, 2019)). Current developments such as building information management (BIM) and asset management were also taken into account. The condition survey and assessment (ZEB) for depicting the network condition and the consistent use of ZEB data for maintenance management were also identified as important components.

Key performance indicators (KPIs) relevant to management were also defined at the strategic level. The results of the statistical analysis of the status values are referred to as key performance indicators and are also among the most important pillars of any strategic (asset) management. Furthermore, questions were formulated from the perspective of strategic management, some of which are listed below as examples. A distinction must be made between fundamental questions in AM and more specific ones with a focus on maintenance management:

- Are/were stakeholder inputs, concerns and expectations considered?
- Are there planned modifications?
- Is outsourcing an issue?
- How do we pursue the approach of sustainable value retention over the life cycle of the assets?
- How has the condition of the network changed?
- How has the condition developed at section level?
- How will the condition develop in the future if no maintenance measures are carried out?
- How efficient were the individual maintenance programs in the network?
- To what change in condition have they contributed?

At a tactical level, a practice-oriented concept for a network-wide condition survey and assessment (ZEB) was also developed with scientific support. On carriageways of the main roads, the metrological ZEB is continued in a four-year recording cycle. On carriageways of the subordinate road network and on secondary areas, either the visual-image-based or the visual-sensitive ZEB is used. The asphalt pavements of the main roads that have been measured and assessed every four years to date only correspond to approx. 8% of the total road surface, meaning that currently approx. 92% of the road surface has not been recorded and assessed. On the one hand, this means a major financial risk, but on the other hand it also means that there is great potential for improvement.

For this reason, visual image-based sample surveys have been commissioned to complete the data in the future. The aim is to gain insights into the various measurement and evaluation techniques and the use of the results to evaluate the facilities for the inventory. Essentially, cycle paths will be surveyed to gain insights into their condition as well as to provide a basis for future planning (e.g. on cycle routes).

The results of the previous ZEB measurement campaigns in 2009, 2013 and 2017 were quality-assured and synchronized, providing a data basis for numerous further evaluations and analyses. The statistical analyses have provided insights into the effectiveness of the selected conservation strategies (the conservation measures implemented). The data basis thus created forms a necessary basis for the development of a model for forecasting the condition of asphalt pavements in the main road network.

The tactical work and maintenance programs are implemented through projects at the operational level.

5. PROCEDURE IN CIVIL ENGINEERING INFRASTRUCTURE MANAGEMENT MÜNSTER (TIMM)

The theory described in the first part will be put into practice using the example of the strategic objective of value retention. Maintaining the value of the infrastructure is essential for a municipality, as a reduction in assets ultimately leads to a reduction in equity. The preservation of equity and thus the preservation of assets is of central importance for the sustainable fulfillment of municipal tasks and the expectations of our stakeholders.

5.1 Initial situation Development of assets

The starting point was the analysis of the development of assets in civil engineering infrastructure since the introduction of the NKF on 01.01.2008. In 16 years, there has been a 16,4% reduction in assets on the balance sheet. This corresponds to a loss in value of € 252 million. Wastewater disposal assets lost 6% of their original value and transportation areas and facilities lost 26% (Fig. 6).

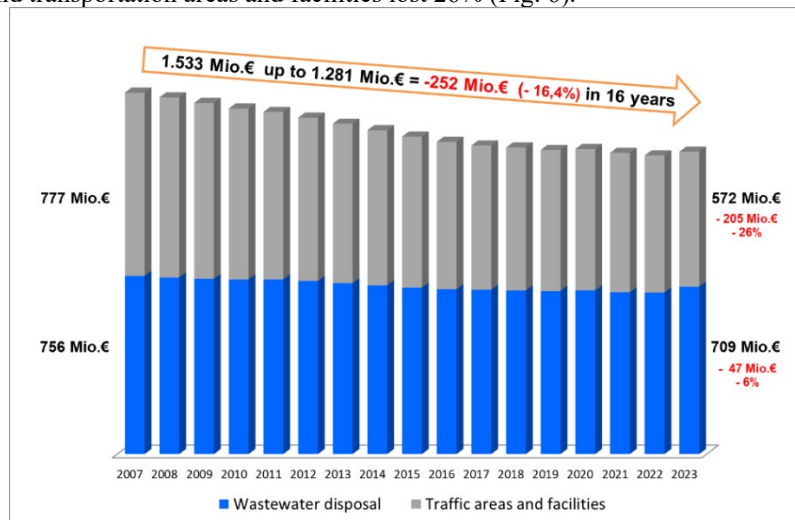


Figure 6. Development of civil engineering infrastructure assets

The greatest loss in value over the past sixteen years was suffered by the municipal road network at 39%. The value of the opening balance sheet of around € 690 million fell by around € 267 million to around € 423 million (Fig. 7).

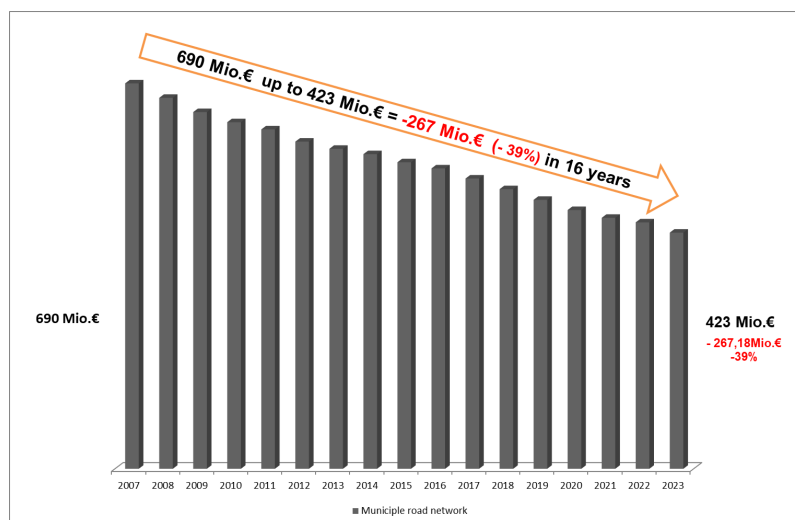


Figure 7. Development of municipal road assets

5.2 What are the reasons for the loss of value? How can countermeasures be taken?

The reason for the high loss in value is the imbalance between depreciation and replacement investments. The loss in value resulting from depreciation cannot be offset by replacement investments. Replacement investments are at a good level and only need to be increased slightly. The main reason for the imbalance is excessive depreciation. This results from insufficient useful lives.

5.3 Which assets are important? What are the most important factors with the greatest leverage?

In the data analysis of the road network, the components of the road network were analyzed. The road network is divided into road categories and the road system is subdivided into carriageway and secondary area. Residential streets make up a significant proportion of the road network, accounting for almost 50% of the total road area. The subordinate network is therefore of particular importance, as are the secondary areas. The data analysis was thus able to identify the important assets of the road infrastructure (Fig. 8).

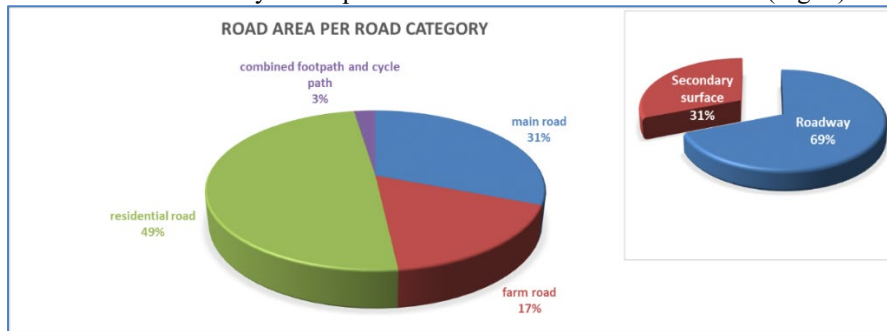


Figure 8. Components of the municipal road network

5.4 Technical versus balance sheet useful life of roads

The “useful lives from optimized maintenance strategy” for roads are to be extended to 40 - 80 years (depending on the road category). The initial assessment according to NKF as of 01.01.2008 was carried out with useful lives for roads of 40 - 60 years as stipulated in the NKF Act. The extension is based on extensive studies of Münster's road network and the experience of recent years. These longer service lives correspond more closely to the technical service lives and should therefore be incorporated into the maintenance strategy. The “Guidelines for the planning of maintenance measures on road pavements” (FGSV, 2001) and the “Ordinance for the calculation of redemption amounts in accordance with the Railway Crossing Act, the Federal Trunk Road Act and the Federal Waterways Act” (BMDV, 2010) also confirm service lives of the individual layers of up to 80 years.

Particularly for the residential street category, which comprises around 50% of the network in Münster, an optimized maintenance strategy should increase the service life to 80 years in line with the underlying sewer. There are clear economic and ecological advantages for the entire road network.

5.5 Effects of extending the useful life

As can be seen from the previous explanations, the technical useful life of the assets plays a key role in maintenance management. The following two simple questions therefore arise: What is the impact of longer useful lives? How do I achieve longer useful lives?

In financial terms, longer useful lives can lead to reinvestments being stretched out, annual road maintenance funds being reduced, and balance sheet depreciation also being reduced. Positive effects from an environmental point of view include lower carbon dioxide emissions, reduced resource consumption, less traffic disruption due to a smaller number of roadworks and less noise emissions due to defective roads, construction work or congestion caused by roadworks.

But how can longer service lives be achieved? At this point, material and construction technology as well as engineering must be analyzed in detail. For example, longer service lives can be achieved by changing material properties or improving quality assurance.

In asphalt road construction, asphalt granulate has been added for years, which comes from roads that are already in use and whose binder has lost its adhesive strength due to the influence of the weather, for example. In order to extend the service life, solutions must be found to improve the adhesive strength. In addition to the addition of fresh bitumen, this can be achieved using rejuvenators. But which one is the right one? Extensive investigations have been carried out in the laboratory, which have subsequently been implemented in pilot projects in Münster. In the meantime, the laboratory tests have shown that a larger quantity of asphalt granulate can be added than in the current regulations, meaning that an additional positive environmental aspect can be achieved by increasing the recycling rate.

Optimized “quality assurance” can be achieved, for example, by ensuring that employees receive regular training. It is conceivable that the previously defined annual number of training days per employee could be increased by 2 days in future. After all, an organization is only able to do its work efficiently through good training

and further training as well as an exchange with other colleagues. Another idea as a metric or key figure would be, for example, to set the target of reducing the number of failed inspections by around 5% per year.

If we add up the above statements and draw a conclusion, we conclude that the measures and fields of action described together serve all three columns of sustainability - ecology, economy and socio-cultural concerns (Fig. 9). It is therefore entirely justified to speak of sustainable asset management in Münster.

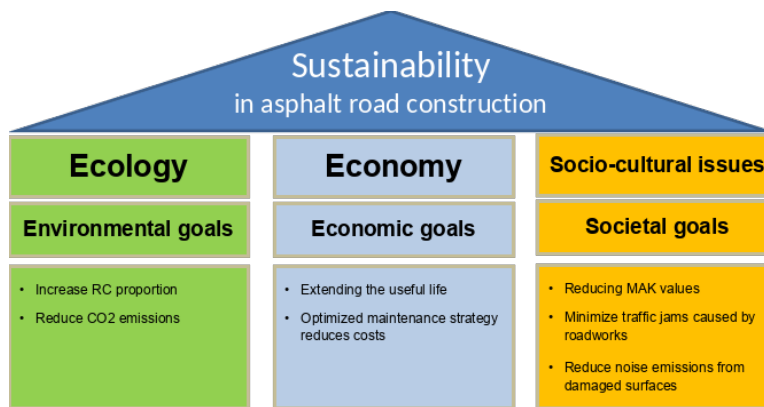


Figure 9. Three columns of sustainability in asphalt road construction

5.6 Coordination of road and sewer service life

As a final example, the cross-asset synchronization of maintenance and renewal schedules for roads and sewers is considered. The aim is to coordinate the planning of measures for both assets. The intersection of the road condition data with the sewer condition data is seen as an essential support here (Fig. 10). Based on individually defined criteria, this makes it possible to form meaningful maintenance sections across all assets so that traffic disruptions in the road network are minimized. This information can also be used to coordinate the work with the utility companies and thus achieve further optimizations. In addition to the ecological benefits, this also has economic advantages.

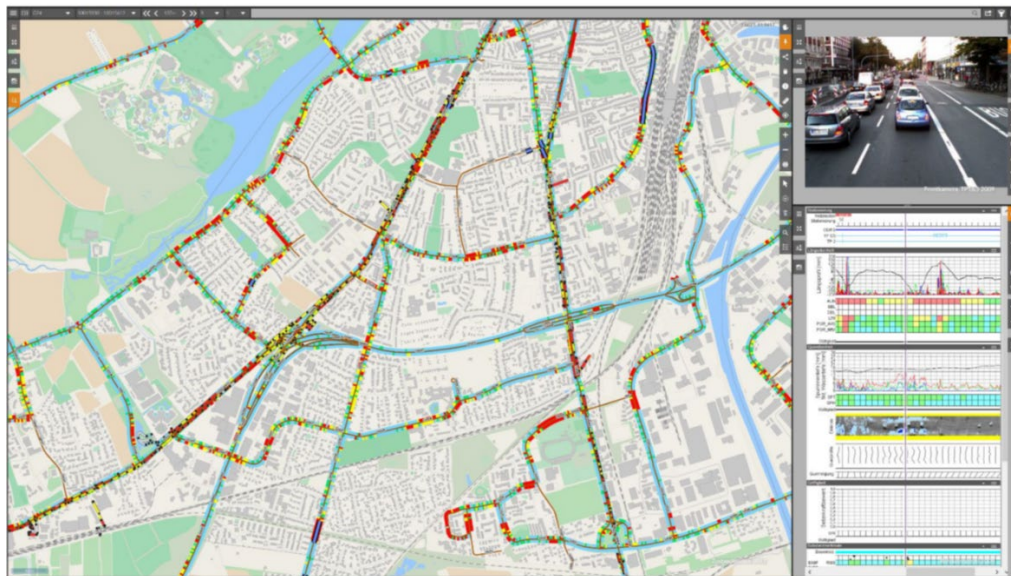


Figure 10. Visualization of road and sewer condition data (Heller, 2019)

4. CONCLUSION AND OUTLOOK

A holistic approach to asset management means infrastructure management in the sense of responsibility for the entire infrastructure or “asset portfolio” (“sewers, roads, bridges, traffic control and parking guidance systems”) over the life cycle by bringing together the technical, planning and commercial aspects towards intergenerational equity. This is therefore sustainable asset management, which forms the framework with organization and communication.

Using the example of the “Road Construction and Maintenance” department of the City of Münster's Office for Mobility and Civil Engineering, it was shown how an AM can be built up to the department level,

considering the city's corporate objectives, and how key figures and metrics can be defined for the municipal infrastructure.

The introduction of the AM has necessitated a readjustment of tasks and standards, which could be explained and communicated to employees in a simple and transparent manner. On the one hand, AM offers managers support in their day-to-day work. On the other hand, strategic planning can be tackled more effectively, the effects of new requirements can be recognized and evaluated more quickly, and any necessary corrections can be made promptly. It quickly becomes clear where, at what point, what risk exists in the completion of tasks or the quality of performance. The introduction of AM has also made it easier to answer questions relating to occupational health and safety or sustainability in road construction, for example, in addition to the further development of systematic road maintenance.

In the opinion of the authors, asset management is a useful tool in public administration. Asset management helps in the development of investment decisions. At the same time, it makes it possible to create quick forecasts and thus enables the administration to react to current political topics or issues. Asset management can also be used to better analyze relevant topics in the future and implement them in the existing working environment. The same applies to the consideration of new technologies.

To be able to answer questions in asset management, it is often necessary to obtain new or additional information. This effort is often shied away from. However, people lose sight of the fact that this new or additional information can also be used to answer more far-reaching questions, e.g. from technical areas - such as smart city or digitalization. The information can also be used to answer financial questions.

For us, asset management means sustainable management in the sense of intergenerational justice, a responsibility for the infrastructure over the entire life cycle in which we bring together the technical planning and commercial aspects as well as the political and legal framework conditions and the expectations of our stakeholders. Ultimately, it is a valuable tool for decision-makers, as it is within their immediate reach, and they can therefore access the information they need for efficient management quickly and easily.

In Münster, strategies and concepts have been developed to supplement the existing regulations, and some have already been tested and implemented in pilot projects. The potential identified here is enormous. Without the courage to carry out pilot projects and experiments, no progress can be made. That is why Münster is continuing to work boldly on innovations.

REFERENCES

- BMDV, (2010). Bundesministerium für Verkehr, Bau und Stadtentwicklung: Verordnung zur Berechnung von Ablösungsbeträgen nach dem Eisenbahnkreuzungsgesetz, dem Bundesfernstraßengesetz und dem Bundeswasserstraßengesetz (Ablösungsbeträge Berechnungsverordnung - ABBV) vom 1. Juli 2010: BMVBS, Ablösungsbeträge-Berechnungsverordnung (BGBl. I S. 856)
- Buttgereit, A.; Gomolluch, S. (2016). Neufassung des PMS, Internes Arbeitspapier des Amtes für Mobilität und Tiefbau der Stadt Münster, unpublished
- Buttgereit, A.; Gomolluch, S. (2019). Das neue Merkblatt über den Finanzbedarf der Straßenerhaltung in den Kommunen – und wie geht es weiter?, Straße und Autobahn, 12/2019, p. 1094–1103
- DIN (2016). DIN ISO 55000, Asset-Management- Übersicht, Grundsätze und Begriffe (ISO 55000:2014), Text Deutsch und Englisch, Entwurf, Stand: August 2016, Beuth Verlag, Berlin
- FGSV, (2019). Forschungsgesellschaft für Straßen- und Verkehrswesen e. V. (2019): Merkblatt über den Finanzbedarf der Straßenerhaltung in den Kommunen
- FGSV, (2019). Forschungsgesellschaft für Straßen- und Verkehrswesen (2001): Richtlinien für die Planung von Erhaltungsmaßnahmen an Straßenbefestigungen RPE-Stra 01, FGSV-Arbeitsausschuss „Systematik der Straßenerhaltung
- Heller, S. (2019). Qualitätssicherung und Feinsynchronisierung der Zustandsdaten dreier ZEB-Kampagnen, unveröffentlichter Bericht für die Stadt Münster
- Münster, (2020a). Amt für Mobilität und Tiefbau Münster: Managementhandbuch, Stand 09/2020
- Münster, (2020b). Amt für Mobilität und Tiefbau Münster: Asset Management -Handbuch, Stand 09/2020
- Schmuck, A (1987). Straßenerhaltung mit System Grundlagen des Managements, Kirschbaum Verlag, Bonn, 1987

EXAMINING FACTORS OF HUMAN-AI TRUST: COMPARING HUMAN- AND AI-RELATED FACTORS

Xinyue Zhou¹ and Sheng Xu²

1) Master, Student, School of Economics and Management, Chang'an University, Xi'an, China. Email: 2023123038@chd.edu.cn

2) Ph.D., Assoc. Prof., School of Economics and Management, Chang'an University, Xi'an, China. Email: sheng.xu@chd.edu.cn

Abstract: With the development of artificial intelligence (AI) technology, human-machine collaboration (HMC) plays an important role in enhancing construction safety. Human trust in AI is the key to the successful implementation of HMC. In this study, the effects of human-related factors (gender, technology acceptance) and AI-related factors (accuracy) on trust were measured and analyzed with controlled experiments simulating a tower crane operation scenario. Twenty-four college students were recruited for the experiment and randomly assigned to two false alarm rate conditions. A remote-controlled tower crane model toy was used to simulate a lifting task. A tablet computer was used to simulate an intrusion alarm monitoring system. Users' initial technical acceptance of the system was assessed via a questionnaire. Subjects' trust scores at the end of each alarm were measured using a trust rating scale. The statistical methods of t-test and two-way ANOVA were used to test the significant relationship between false alarm rate, gender, technology acceptance and trust score. The results show that the false alarm rate is a key factor affecting trust, while gender and technology acceptance and their interaction effects with the false alarm rate are not significant. The study emphasizes the importance of reducing false alarms and improving AI accuracy to enhance user trust.

Keywords: Human-machine collaboration, Human-AI trust, False alarm rate, System reliability.

1. INTRODUCTION

With the rapid development of AI, deep learning, big data, and other information technologies (Duan et al., 2019), machines have transitioned from mechanization and automation to intelligence over the past few decades. With the continuous penetration of intelligent technologies in the construction field, intelligent innovations such as drones, sensors, and smart visualization cameras provide new solutions for remote safety monitoring at construction sites. They play an important role in reducing operator workloads and compensating for the limitations of operator capabilities. However, during hazardous inspections of image data, monitors may face several challenges, including fatigue, stress, and distraction which results in up to half of the potential hazards being overlooked (Chen et al., 2016). To overcome the shortcomings of manual monitoring, advanced computer vision techniques have been developed to enable situation recognition, reasoning, diagnosis, and decision support (Paneru & Jeelani, 2021). Although high levels of automation and intelligence continue to evolve and improve, their value lies in assisting rather than completely replacing human operators. As a result, human-machine collaborative (HMC) systems have emerged as a key factor in improving construction safety and efficiency and play an important role in hazard detection. Among them, human trust in AI is essential for the successful implementation of HMC strategies to improve construction safety (Paneru & Jeelani, 2021).

It has been shown that trust is influenced by both human and AI influences. Human influences often include gender, technology acceptance, and experience. Gender differences may influence how much an individual trusts AI (Hu et al., 2019). The technology acceptance model provides a useful theoretical framework to validate the importance of trust in AI to predict acceptance of AI (Choung et al., 2023). Experience, on the other hand, influences an individual's judgment of the reliability of AI-related influences focusing on accuracy, predictability, and transparency. Accuracy refers to the precision and reliability of the AI in performing a task. Predictability refers to the stability of the AI's behavior and outcomes within the expected range. Transparency refers to the visibility and understandability of the AI's workings and decision-making process to the user (Parasuraman & Riley, 1997). Together, these factors determine user trust in the AI. Therefore, to enhance users' trust in AI systems, these factors need to be considered and analyzed comprehensively to ensure that the system's performance and user experience are optimized.

Intelligent monitoring and alerting systems are one of the important applications of AI in the construction field. When an intelligent monitoring and alerting system detects unsafe human behavior or an unsafe state of a machine, it issues an alert prompting the operator to take corrective action. However, differences in individual human characteristics and biases against AI may trigger human-machine trust issues, resulting in the misuse or abandonment of intelligent systems. In addition, most alert systems set low thresholds because the cost of missed alarms (e.g., casualties) is usually much higher than the cost of false alarms. This results in most alerts being "false alarms" with negative consequences. (Zhou & Liao, 2023). First, because the final decision is made by humans, frequent false alarms cause humans to double-check the information to ensure that the alarm is indeed false, which

can be distracting and consume unnecessary energy (Okpala et al., 2020). Second, after frequent false alarms, people may experience "crying wolf" syndrome, which can lead them to ignore alarms (including true alarms), ultimately resulting in fatal accidents and system abandonment (Woods, 2019). Therefore, to prevent potential problems, AI should not only be used as a tool to assist humans but also to support the process of human trust.

Therefore, this study conducted a human-machine trust experiment for a remotely controlled tower crane hoisting task by simulating a tower crane hazardous area intrusion alarm system scenario. The human-related factors in this research focused on gender and technology acceptance. The AI-related factors is focused on accuracy. Through controlled experiments, the factors were analyzed and compared. The AI accuracy information is provided in the form of "alarm + monitoring", where the alarm system provides immediate warning and the monitoring system continuously provides real-time operating environment data, which closely fits the operating scenarios of tower crane drivers under intelligent technology. Statistical analysis methods including T-test and two-way ANOVA are used. This study aims to reveal the factors of that influence human-machine trust and provide a new perspective on human-machine trust in Chinese HMCs in the construction industry.

2. LITERATURE REVIEW

All forms of interaction between humans and machines are collectively referred to as human-machine interaction. The machine can refer to automated or autonomous systems, autonomous agents, robots, algorithms, or AI (Xiong et al., 2022). For example, in order to improve the safety of tower crane operations, an intelligent identification and warning system that realizes accurate monitoring and real-time warning of dangerous areas through image acquisition, intelligent identification, coordinate warning, and other technologies (WU et al., 2024). The system is a new type of robot that can be used to perform a single task. Many construction robots are manufactured to perform a single type of work, such as autonomous excavators (Kim et al., 2019). Human-robot collaboration, as a type of human-robot interaction, can be used to perform complex tasks through physical contact or non-contact collaboration between humans and machines (Hentout et al., 2019). In this process, humans are responsible for exercising dexterity and making decisions, while machines take on tasks that are not suitable for direct human execution, such as high-precision manipulation (Yang et al., 2022). In the construction industry, this type of human-machine interaction can be utilized in a variety of ways. This cooperation between humans and machines is particularly important in the construction industry, as it involves complex task execution and risk management (Janssen et al., 2019). As the level of AI increases, so does the autonomy of smart machines, which may affect the trust of workers, making it particularly important to study trust in the built environment (Alikhani et al., 2023).

Trust can be defined as a strong belief in another person's intentions or will, by following their words, expressions, decisions or actions (Gupta et al., 2020). As humans are increasingly required to interact with AI, automated systems, trust becomes an important factor in synergistic interrelationships. In this context, sufficient trust can mediate the relationship between humans and automated systems (Yagoda & Gillan, 2012). Trust is a subjective experience with three components: dispositional trust, situational trust, and learned trust (Hopko et al., 2023). Dispositional trust is influenced by demographic, personality, and social characteristics. Previous research has shown that gender may influence trust building, with males being more inclined than females to trust robots (Hu et al., 2019). Acceptance and perception of robot behaviors are influenced by operator gender (Kuo et al., 2009). Venkatesh has emphasized the important role of gender differences in technology acceptance research in several empirical studies, where men are more sensitive to the perceived usefulness of technology while women value the perceived ease of use of technology, as well as significant gender differences in subjective normative factors in technology acceptance (Venkatesh et al., 2000). Low acceptance reduces human-computer collaboration and team performance. Acceptance is related to many factors in human-machine collaboration, such as machine performance, transparency, interpretability, and human characteristics (Gursoy et al., 2019; Kraus et al., 2020). Situational trust includes internal human factors (e.g., fatigue) and external factors (e.g., accuracy). The higher the system accuracy, the more operators trust the automated system and the higher the human-machine collaboration performance (Hoesterey & Onnasch, 2023). Acquired trust is related to the operator's expertise and past experience, and this trust changes dynamically (Hoff & Bashir, 2015). Thus, the overall perception of trust in a human-machine collaborative environment depends on human-related factors, AI-related factors, and the interactions between them.

Although there are many papers studying the factors influencing trust, there are still some research gaps. First, there are discrepancies between the findings presented in different papers. These inconsistencies may be related to the small sample size and the nature of the different research questions. Therefore, further research on various factors (e.g., the effect of validation factors on trust or examining mediators and moderators in human-computer trust) is needed. Second, it is unclear whether the results of previous studies can be replicated in the construction field. Due to the specific nature of the industry, the effects of previously proposed influences on trust need to be revisited during the construction process and, more importantly, more construction-related variables need to be considered.

Therefore, research on the effects of human and AI influences on trust in human-machine collaboration in the construction domain is still in its infancy. There is a gap in the understanding of the impact of gender,

technology acceptance and system accuracy on dynamic trust, and in this context, the focus of this study is to understand how gender, technology acceptance and system accuracy as human-AI factors affect trust in human-computer collaboration, and to explore the impact of different factors on dynamic trust.

3. METHODOLOGY

3.1 Subjects

Twenty-four undergraduate and graduate students, 12 males and 12 females, with an age range of 20 to 25 years ($M = 23.2$ years, $SD = 1.4$ years) were openly recruited for this experiment at Chang'an University. All subjects had normal hearing, normal or corrected-to-normal vision, and were in good health. All subjects were randomly assigned to one of two false alarm rate (33.33% and 66.67%) experimental conditions, with 12 subjects in each condition, including 6 males and 6 females. There was no significant difference in the age of the subjects between the two experimental conditions and they had never participated in a similar experiment. All subjects received a gift in return at the end of the experiment.

3.2 Equipment and Materials

The remote-control tower crane model toy used in this experiment is mainly composed of a lifting arm, balance arm, luffing trolley, lifting hook, and workbench (as shown in Figure 1a). The total height of the model is 132cm, the arm length is 90cm, and the working table rotates 360°. Using 2.4GHz wireless remote control can control the front and rear movement of the luffing trolley, the rotation of the lifting arm and the up and down of the rope. In this experiment, the remote-controlled tower crane model toy was used to simulate the operation of a tower crane in a real construction site. Subjects used the remote control to control the tower crane to complete the tasks of handling objects and placing objects. The model can help the subjects establish a connection with the actual situation. Test and record their behavioral responses and trust scores when performing the tasks under different experimental conditions.

A HUAWEI MatePad tablet with a screen size of 11.5 inches was used to play the tower crane hazardous area intrusion alarm simulation monitor. The tablet was placed in front of the operator's right-hand side, simulating the placement of an actual tower crane hook monitoring panel. The clips and alarm tones for the tower crane hazardous area intrusion alarm simulation monitoring were publicly available on the Internet. The video clips were segmented using the Python programming language and related libraries. The true and false alarms were randomly reorganized according to the experimental condition ratios (33.33% and 66.67% false alarm rate). The generated long video consists of 10 video clips, of which 6 video clips are alarm clips, with a total duration of 4 minutes and 32 seconds.

Two false alarm rate conditions, 33.33% and 66.67%, were selected in this study based on a review of the existing literature and an understanding of the practical application context. Wang et al. study human-machine trust in the aircraft engine fire alarm system scenario with a false alarm rate set to 33.33% and all alarm sequences considered (Wang et al., 2022). Many studies do not have a uniform and fixed standard for setting the false alarm rate and accuracy of the system but rather set the relative ratios according to the needs of the study (Li et al., 2023). In complex environments such as construction sites, security monitoring systems often need to set lower thresholds to avoid major accidents that may be caused by missed alarms, which also leads to higher false alarm rates. In this experiment, the number of alarms in one experiment is 6. The ratio of the number of false alarms is 1/3 for a low false alarm rate and 2/3 for a high false alarm rate. Therefore, these two ratios represent relatively low and high false alarm scenarios, respectively. These ratios were chosen to simulate different scenarios in the real world, where 33.33% of the false alarm rates are closer to good performance under ideal conditions, while 66.67% reflect some of the challenges that may exist under current technological conditions. By comparing the user responses in these two extreme scenarios, it is possible to better understand how the false alarm rate affects trust building during human-computer interaction.

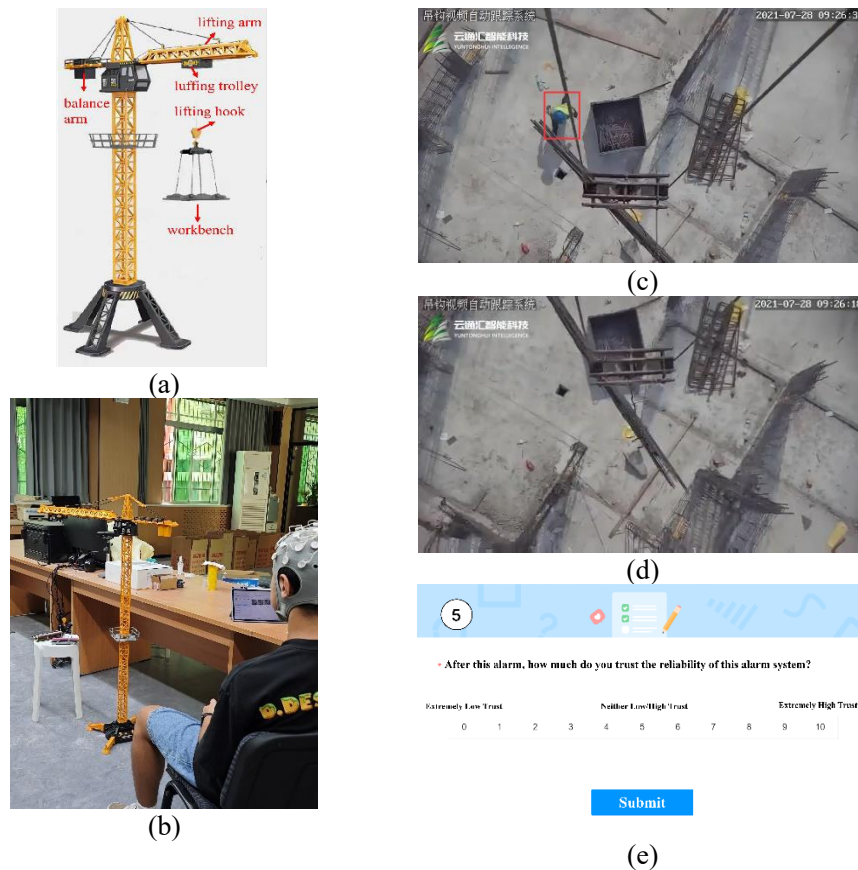


Figure 1. (a) Remote-control tower crane model (b) Experiment setup (c) True alarm (d) False alarm (e) Trust rating scale

3.3 Trust Measures

Self-report (questionnaire and trust rating scale) was used in this study. The "Tower Crane Hazardous Area Intrusion Alert System Technical Acceptance" questionnaire was used to collect demographic information prior to the experiment and to assess the initial technical acceptance of the system by the subjects. This acceptance was based on the system's performance, reliability, and the user's perception of the system's usefulness, ease of use, and behavioral attitudes and intentions. For the experiment, subjects verbally reported trust ratings after determining the authenticity of each alert to measure the subjects' trust scores at the end of each alert.

3.4 Experimental Design

This experiment utilized a 2 (gender) \times 2 (false alarm rate) between-subjects design. All subjects were required to sign an informed consent form and divided into two groups by random draw for the experiment with different false alarm rates, with a balanced gender ratio in each group. Subsequently, subjects were asked to complete a questionnaire on the technical acceptability of the tower crane hazardous area intrusion alarm system. Prior to the official start of the experiment, subjects were informed in detail about the experimental procedure and precautions, and familiarized with the operation of the remote-control handle of the remote-controlled tower crane model toy, which lasts for 3 to 5 minutes.

At the beginning of the experiment, subjects were required to maneuver a remote-controlled tower crane for the task of lifting objects, which were loaded and unloaded by the experimenter. An analog monitor of the tower crane's hazardous area intrusion alarm was randomly played. When the simulated monitor sounded an alarm, subjects were required to check the monitor screen. It was a true alarm if the monitor screen showed that a worker was passing by (as shown in Figure 1c), and a false alarm if the monitor screen showed that no worker was passing by underneath the crane hook (as shown in Figure 1d). A trust rating scale then appeared on the screen and subjects were asked, "After this alarm, how much do you trust the reliability of this alarm system?" (as shown in Figure 1e). To avoid a lengthy question, it was explained to subjects prior to the experiment that the trust rating would be based on cumulative alarm experience as a whole, not just the current alarm. Ratings ranged from 1 (extremely low trust) to 10 (extremely high trust). Subjects were required to verbally report the results of the ratings. After completing one rating, the lifting task was continued. Subjects need to wait for the next alarm until the end of this experiment.

This study utilized a multiple-experiment approach. Each subject was required to complete three experiments in order to extend the total experimental duration and to ensure that more data points were collected, thereby increasing the reliability and validity of the results and reducing errors arising from the influence of chance factors (e.g., mood swings and distraction of the subjects) on the results of a single experiment. The length of each experiment was approximately 5 minutes. There was a 2-minute break between each experiment in which subjects were asked to relax. The total duration of the experiment was approximately 30 minutes.

3.5 Data Analysis

(1) Technology Acceptance

Before the experiment, subjects completed a 13-item questionnaire on "Technical Acceptance of Tower Crane Hazardous Area Intruder Alert Systems". The items included five dimensions: knowledge of the system (e.g., "I understand how the intrusion alarm system works"), perceived ease of use (e.g., "The interactive information of the intrusion alarm system is clear and easy to understand"), perceived usefulness, behavioral attitudes, and behavioral intentions. To ensure the quality of the questionnaire, after the initial design of the questionnaire, a small sample was collected for Cronbach's coefficient alpha reliability analysis and Bartlett's spherical test validity analysis. In this study, a technology acceptability score was obtained by summing all scores for each item and dividing by the total number of items. The mean value of technology acceptance was calculated for all subjects, and based on this mean value, the subjects were categorized into "high technology acceptance" and "low technology acceptance" groups.

Table 1. Results of reliability analysis

Dimension	Cronbach Alpha	Normalized term-based clone Bach Alpha	Item count
Knowledge of the system	.861	.866	2
Perceived ease of use	.855	.855	2
Perceived usefulness	.820	.828	3
Behavioral attitudes	.833	.834	3
Behavioral Intentions	.885	.889	3

(2) Trust Score

The average score of all trust ratings in the three experiments was used as the trust score for each subject as in Equation (1) (Chauhan et al., 2024). Where r_{ij} denotes the j_{th} trust rating in the i_{th} experiment.

$$\text{Trust score} = \frac{\sum_{i=1}^3 \sum_{j=1}^6 r_{ij}}{3 \times 6} \quad (1)$$

(3) Methods of Analysis

The dataset was grouped according to false alarm rate, gender, and technology acceptance, respectively. After the Shapiro-Wilk test was used to initially assess the normal distribution information of the dataset, the significance of each of the three for the difference in the trust scores was compared using the two-independent samples t-test. The dataset was divided into two subgroups according to gender and technology acceptance under two groups of false alarm rates, representing the male and female groups, the high technology acceptance group and the low technology acceptance group, respectively. The significance of the differences in trust scores by gender and technology acceptance under the two types of false alarm rates was compared using two-way ANOVA, respectively.

4. RESULTS

The results in Table 1 show that the questionnaire exhibited good reliability (Cronbach's alpha > 0.8) and validity ($p < 0.001$). The mean value of technology acceptance for all subjects was 4.05 with a standard deviation of 0.55. Based on the grouping, 12 data were available for the high and low technology acceptance groups respectively.

The Shapiro-Wilk test performed on the data sets indicated that all data sets satisfied normal distribution. Across the datasets, high and low false alarm rates were significantly different from trust scores as shown in Table 2. The t-test results showed that trust scores in the low false alarm rate group (7.107 ± 1.493) were significantly higher than trust scores in the high false alarm rate group (5.116 ± 0.725), $t = 4.155$, $p < 0.001$, 95% CI (0.975 ~ 3.007). This suggests that false alarm rate is an important factor influencing trust scores and that low false alarm rate may be associated with higher perceptions of trust. This may be due to the fact that a low false alarm rate reduces false alarms and improves the accuracy of the system, which in turn enhances the user's trust in the system.

There was no significant difference between gender and trust scores, as shown in Table 3. The t-test results showed that the difference was not statistically significant at $t = -0.717$, $p = 0.481$. The interaction effect between

false alarm rate and gender was not significant, as shown in Table 5. The results of two-way ANOVA test showed that the main effect of the false alarm rate was significant ($F=16.366$, $p<0.001$, partial $\eta^2=0.450$). In contrast, the main effect of gender was not significant ($F=0.850$, $p=0.368$), nor was the interaction effect between false alarm rate and gender ($F=0.006$, $p=0.941$). This suggests that gender may not be a significant factor influencing trust scores in this sample. This result may be due to the more balanced distribution of gender in the sample and the fact that gender differences may have less of an effect on technology trust in the student population.

There was no significant difference between technology acceptance and trust scores, as shown in Table 4. The t-test results showed that the difference was not statistically significant at $t=0.145$ and $p=0.886$. The interaction effect between false alarm rate and technology acceptance was not significant, as shown in Table 5. The results of the two-way ANOVA test showed that a significant main effect of false alarm rate ($F=16.232$, $p<0.001$, partial $\eta^2=0.448$), while the main effect of technology acceptance was not significant ($F=0.238$, $p=0.631$). The interaction of false alarm rate and technology acceptance effect was also not significant ($F=0.189$, $p=0.669$), suggesting that initial technology acceptance had little effect on trust scores in this study. It indicates that the effects of these two factors on trust scores are independent. This may be related to the fact that the experimental subjects, who were all students and had no experience with the tower crane hazardous area intrusion alarm system, did not differ significantly in initial technology acceptance and therefore had a limited effect on trust scores.

Table 2. Difference in human trust score between low false alarm rate and high false alarm rate

Human trust measure	Low false alarm rate		High false alarm rate		T value	p value	95%CI
	Mean	SD	Mean	SD			
Trust score	7.107	1.493	5.116	0.725	4.155	$p<0.001^*$	0.975~3.007

Table 3. Difference in human trust score between male and female

Human trust measure	Male		Female		T value	p value	95%CI
	Mean	SD	Mean	SD			
Trust score	5.884	1.603	6.338	1.495	- 0.717	0.481	- 1.766~0.858

Table 4. Difference in human trust score between high technology acceptance and low technology acceptance

Human trust measure	High technology acceptance		Low technology acceptance		T value	p value	95%CI
	Mean	SD	Mean	SD			
Trust score	6.157	1.660	6.065	1.468	0.145	0.886	- 1.234~1.419

Table 5. Two-way ANOVA analysis

Variable: trust score					
	Sum of squares	Degree of freedom	Mean squares	F	p value
False alarm rate *Gender	0.008	1	0.008	0.006	0.941
False alarm rate *Technology acceptance	0.280	1	0.280	0.189	0.669

5. DISCUSSION

This study provides insights into the factors influencing human-machine trust by simulating a tower crane hazardous area intrusion alarm system scenario. The finding indicates that AI accuracy, especially false alarm rate, was shown to be a significant factor affecting trust scores. The experimental results showed that the false alarm rate was significantly correlated with trust scores, and the trust scores of the low false alarm rate (33.33%) group were significantly higher than those of the high false alarm rate (66.67%) group. This may be due to the fact that the low false alarm rate reduces false alarms and improves the accuracy of the AI, which in turn enhances the user's trust in the AI. This finding emphasizes the importance of reducing false alarms and improving AI accuracy when designing and implementing tower crane alarm systems.

Gender is not a significant factor in trust scores, which may be related to the balanced gender distribution in the sample and the low impact of gender differences in the student population on technology trust. This finding contradicts some previous studies that suggest that gender may play an important role in technology trust (Hu et al., 2019). Potential reasons for this difference may include factors such as sample composition and experimental design characteristics. The participants in this experiment were all in the college student population, and the age range was concentrated between 20 and 25 years old. The younger generation is generally open to new technologies, which may have attenuated the effect of gender differences in the traditional sense. In addition, none of the subjects had prior exposure to similar tower crane hazardous area intrusion alarm systems, implying that

their initial technology acceptance was more consistent, further minimizing the differences due to gender. Although no significant gender effect was observed in this study, this does not mean that the same conclusion would be reached in other populations or different application scenarios. Future research should consider expanding the sample size and examining more diverse contextual factors to fully assess how gender and other personal characteristics work together to shape trust formation during human-computer collaboration.

The main effect of technology acceptance was not significant, which may be related to the fact that the experimental subjects were all students and had no experience with tower crane hazardous area intrusion alarm systems. The difference in initial technology acceptance was not significant and therefore had a limited effect on trust scores. The interaction effect between false alarm rate and technology acceptance was also not significant, suggesting that these two factors had independent effects on trust scores. This result suggests that the effect of false alarm rate on trust may be consistent across levels of technology acceptance.

This experiment also recorded the subconscious behavior of whether subjects chose to believe after hearing an alarm, while EEG signals were collected as subsequent analysis data. Objective data from EEG were used to further analyze the dynamics and neural mechanisms of trust.

6. CONCLUSIONS

This study reveals that the false alarm rate is a key factor influencing tower crane drivers' trust in man-machines, while gender and technology acceptance have insignificant effects in this study. These findings provide new perspectives on the study of human-machine trust in the construction industry, especially in the scenario of tower crane drivers' operations with intelligent technology. The findings emphasize the importance of focusing on reducing false alarms and improving system accuracy when designing and implementing tower crane alarm systems in order to enhance users' trust in the system. In addition, the effects of gender and technology acceptance on trust may vary depending on sample characteristics, and future research needs to validate this in a wider population and explore other factors that may affect trust, such as individual experience, cultural background, etc. With further research, we can better understand the dynamics of human-machine trust and provide more effective strategies for safety management in the construction industry.

ACKNOWLEDGEMENT

This research is supported by Shaanxi Province (Program No. 2024GH-GHJD-16), Fundamental Research Funds for the Central Universities (Program No. 300102234618, 300102234501, 211923200005).

REFERENCES

- Alikhani, H., Le, C., Jeong, H. D., and Damnjanovic, I. (2023). Sequential Machine Learning for Activity Sequence Prediction from Daily Work Report Data, *Journal of Construction Engineering and Management*, 149 (9).
- Chauhan, H., Pakbaz, A., Jang, Y., and Jeong, I. (2024). Analyzing Trust Dynamics in Human-Robot Collaboration through Psychophysiological Responses in an Immersive Virtual Construction Environment, *Journal of Computing in Civil Engineering*, 38 (4).
- Chen, J., Song, X., and Lin, Z. (2016). Revealing the "Invisible Gorilla" in construction: estimating construction safety through mental workload assessment, *Automation in Construction*, 63, 173-183.
- Choung, H., David, P., and Ross, A. (2023). Trust in AI and Its Role in the Acceptance of AI Technologies, *International Journal of Human-Computer Interaction*, 39 (9), 1727-1739.
- Duan, Y. Q., Edwards, J. S., and Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data - evolution, challenges and research agenda, *International Journal of Information Management*, 48, 63-71.
- Gupta, K., Hajika, R., Pai, Y. S., Duenser, A., Lochner, M., and Billingham, M. (2020). Measuring Human Trust in a Virtual Assistant using Physiological Sensing in Virtual Reality, *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*.
- Gursoy, D., Chi, O. H., Lu, L., and Nunkoo, R. (2019). Consumers acceptance of artificially intelligent (AI) device use in service delivery, *International Journal of Information Management*, 49, 157-169.
- Hentout, A., Aouache, M., Maoudj, A., and Akli, I. (2019). Human-robot interaction in industrial collaborative robotics: a literature review of the decade 2008-2017, *Advanced Robotics*, 33 (15-16), 764-799.
- Hoesterey, S., and Onnasch, L. (2023). The effect of risk on trust attitude and trust behavior in interaction with information and decision automation, *COGNITION TECHNOLOGY & WORK*, 25 (1), 15-29.
- Hoff, K. A., and Bashir, M. (2015). Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust, *HUMAN FACTORS*, 57 (3), 407-434.
- Hopko, S. K., Mehta, R. K., and Pagilla, P. R. (2023). Physiological and perceptual consequences of trust in collaborative robots: an empirical investigation of human and robot factors, *Applied Ergonomics*, 106.
- Hu, W. L., Akash, K., Reid, T., and Jain, N. (2019). Computational Modeling of the Dynamics of Human Trust

- During Human-Machine Interactions, *IEEE Transactions on Human-Machine Systems*, 49 (6), 485-497.
- Janssen, C. P., Donker, S. F., Brumby, D. P., and Kun, A. L. (2019). History and future of human-automation interaction, *International Journal of Human-Computer Studies*, 131, 99-107.
- Kim, K., Kim, M., Kim, D., and Lee, D. (2019). Modeling and velocity-field control of autonomous excavator with main control valve, *Automatica*, 104, 67-81.
- Kraus, J., Scholz, D., Stiegemeier, D., and Baumann, M. (2020). The More You Know: Trust Dynamics and Calibration in Highly Automated Driving and the Effects of Take-Overs, System Malfunction, and System Transparency, *HUMAN FACTORS*, 62 (5), 718-736.
- Kuo, I. H., Rabindran, J. M., Broadbent, E., Lee, Y. I., Kerse, N., Stafford, R. M. Q., and MacDonald, B. A. (2009). Age and gender factors in user acceptance of healthcare robots, *RO-MAN 2009 - The 18th IEEE International Symposium on Robot and Human Interactive Communication*.
- Li, Y., Zhang, L., Huang, Q., and Ma, S. (2023). Impact of System Transparency Information on Human-Machine Trust and Collaborative Decision Making, *Packaging Engineering*, 44 (20), 25-33.
- Okpala, I., Parajuli, A., Nnaji, C., and Awolusi, I. (2020). Assessing the Feasibility of Integrating the Internet of Things into Safety Management Systems: a Focus on Wearable Sensing Devices, *Construction Research Congress 2020 - Computer Applications*, 259-268.
- Paneru, S., and Jeelani, I. (2021). Computer vision applications in construction: current state, opportunities & challenges, *Automation in Construction*, 132, 103940.
- Parasuraman, R., and Riley, V. (1997). Humans and automation: use, misuse, disuse, abuse, *HUMAN FACTORS*, 39 (2), 230-253.
- Venkatesh, Morris, and Ackerman. (2000). A Longitudinal Field Investigation of Gender Differences in Individual Technology Adoption Decision-Making Processes, *Organizational behavior and human decision processes*, 83 (1), 33-60.
- Wang, Y. F., Guo, J. B., Zeng, S. K., Mao, Q. R., Lu, Z. P., and Wang, Z. K. (2022). Human-Machine Trust and Calibration Based on Human-in-the-Loop Experiment, *2022 4TH INTERNATIONAL CONFERENCE ON SYSTEM RELIABILITY AND SAFETY ENGINEERING*.
- Woods, D. (2019). First & Last Interview: Boeing 737 Max accidents reveal past results on Automation Surprises.
- Wu, L., Li, H., Li, D., Wu, Y., Liu, P., and Xue, X. (2024). Crane danger zone intrusion warning based on computer vision, *China Safety Science Journal*, 34 (07), 139-145.
- Xiong, W., Fan, H., Ma, L., and Wang, C. (2022). Challenges of human-machine collaboration in risky decision-making, *Frontiers of Engineering Management*, 9 (1), 89-103.
- Yagoda, R. E., and Gillan, D. J. (2012). You Want Me to Trust a ROBOT? The Development of a Human-Robot Interaction Trust Scale, *International Journal of Social Robotics*, 4 (3), 235-248.
- Yang, G., Zhou, H., and Baicun, W. (2022). Digital Twin-driven Smart Human-machine Collaboration: Theory, Enabling Technologies and Applications, *Journal of Mechanical Engineering*, 58 (18).
- Zhou, X. and Liao, P.-C. (2023). EEG-Based Performance-Driven Adaptive Automated Hazard Alerting System in Security Surveillance Support, *Sustainability*, 15 (6).

A DOMAIN KNOWLEDGE-ENHANCED LARGE VISION-LANGUAGE MODEL FOR CONSTRUCTION SITE SAFETY MONITORING

Chak-Fu Chan¹, Xiaowen Guo², Peter Kok-Yiu Wong³, Jolly Pui-Ching Chan⁴, Jack C.P. Cheng⁵, Pak-Him Leung⁶ and Xingyu Tao⁷

1) M.Phil. Candidate, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: cfchanay@connect.ust.hk

2) Ph.D. Candidate, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: xguoaw@connect.ust.hk

3) Post-doctoral Fellow, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: cekywong@ust.hk

4) Electrical and Mechanical Engineer, Drainage Services Department, the Government of the Hong Kong Special Administrative Region, Wanchai, Hong Kong Island, Hong Kong, SAR. Email: pcchan02@dsd.gov.hk

5) Professor and Associate Head, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: cejcheng@ust.hk

6) Co-Founder, AutoSafe Ltd. Email: issacleung@autosafe.ai

7) Research Assistant Professor, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: xtaoab@connect.ust.hk

Abstract: To address the industry-wide and policy-driven requirements toward construction site safety monitoring, this paper develops a virtual assistant agent based on a large vision-language model (VLM), integrated into on-site surveillance camera system for real-time identification and alerting of unsafe worker behaviors. First, we designed a semi-automatic image-text labeling pipeline, employing in-context learning to enhance data annotation efficiency. Then, we established a two-stage curriculum learning paradigm to deeply embed construction domain knowledge into the VLM, which is eventually embedded into a real-time video analytical engine for safety compliance inspection and interactive visual question answering. The system has been deployed on a real construction site, with around 90% accuracy in identifying violations of work-at-height safety regulations.

Keywords: Construction Site Safety Monitoring, Data-Efficient Fine-tuning Strategy, Domain-Tailored Large Vision-Language Model, Multi-modal Safety Compliance Checking, Virtual Construction Safety Assistant

1. INTRODUCTION

The construction sector is recognized as one of the most dangerous sectors worldwide, with a history of high accident and casualty rates. From 2013 to 2019, the construction sector in Hong Kong has consistently recorded the highest death rates compared to the other 14 primary industry sectors (Labour Department, 2018, 2019). Similarly, in 2019, the United States reported that approximately 20% of work-related deaths took place at construction sites (Occupational Safety and Health Administration, 2019). Many of the casualties are related to non-compliance with construction safety rules. Construction sites are characterized by safety risks that arise from complex interactions between numerous workers and machinery. To prevent safety risks and serious injuries, it is crucial to detect and analyze any unsafe practices among workers in real-time. This can involve monitoring workers' activity on-site and ensuring they're complying with the safety rules. Traditional site safety monitoring mainly involves regular on-site safety inspection by management personnel and safety inspectors, which is labor-intensive and easily leads to overlooking unsafe behavior. In recent years, analyzing construction images or videos by computer vision (CV)-based deep learning (DL) methods have been widely studied for identifying unsafe construction objects and behavior (Cheng et al., 2022), issuing real-time alerts or more in-depth safety analyses.

However, existing CV-based safety analysis characterized by Convolutional Neural Network (CNN) models has been faced with several problems. First, developed CV models are specialist models only well-trained for a small subset of detection tasks, thus limited by the narrow domain of knowledge. For example, an object detection model trained to perform decently in PPE detection may fall short in fall or injury detection. Second, the architecture of CNN-based models leads to limited embedded semantic information (e.g. object category and localization), while extracting semantic information (e.g., construction activities, interactions between different construction objects) from construction images is an essential step for further CV-based application in construction management (Paneru & Jeelani, 2021; Y. Wang et al., 2022). More high-level tasks like safety compliance checking require a high-level and comprehensive semantic understanding of the on-site scene.

Other researchers have integrated the developed CV models with some other Natural Language Processing (NLP) techniques like pre-defined knowledge graphs for further compliance checking (Fang et al., 2020; Tang et al., 2020). However, processing information separately with vision and language modules may be inefficient and time-consuming. Moreover, during the process of transferring safety information across different modalities, features not extracted in the vision modality will inevitably be lost and not represented in the language modality.

The recent advancement in Large Language Model (LLM), based on Generative Pre-trained Transformer

(GPT), has shown tremendous potential for human-like reasoning and conversation. To enable multi-modal tasks like visual question answering (VQA), Large Vision-Language Model (VLM) is further developed with an encoder-decoder architecture, where the encoder processes visual information and the decoder generates textual representation (Chen et al., 2023). An illustrative example is the GPT-4V developed by OpenAI (OpenAI et al., 2024), which has demonstrated human-like reasoning capabilities in combining natural language, texts, and images into VQA processes.

Enabled by the GPT-based pre-trained large vision and language backbones, VLMs can extract both intricate visual and linguistic features, undergo deep fusion between them and generate detailed image descriptions, and even perform multiple types of instruction-following tasks such as multi-round conversation that require sophisticated visual-semantic understanding and reasoning. The strong capability of these pre-trained VLMs made them a strong candidate in many traditional vision-language tasks like image captioning, visual grounding, and VQA in a zero-shot or few-shot setting without any fine-tuning. Compared with CV-NLP-based methods, VLMs enable deep fusion between vision and language features for more comprehensive safety analysis.

While pre-training of VLMs from a large corpus of image-text pairs can align the visual encoder with the language backbone's word embedding space and achieve all-round performance in general tasks, fine-tuning can be performed on a much smaller but more supervised instruction-following dataset to adapt the model to domain-specific context. However, there remain challenges when adapting a pre-trained VLM to tasks related to construction safety analysis and management via fine-tuning.

1) Fine-tuning of VLMs requires a vast amount of labeled and high-quality data. The construction industry is known for limited open access to structured data for DL training, and currently, there is a lack of public datasets for effectively fine-tuning the VLMs toward diverse construction safety tasks. However, the fine-tuning of LLM may require thousands of images, together with high-quality instruction-response annotations, which is a resource and labor-consuming task, compared to those of traditional VQA and image-captioning datasets which may just consist of a few words as the response for each image.

2) Similar to CNN-based models, VLMs are mainly pre-trained on internet-crowdsourced images, which have notably different image quality from those captured on real construction sites by surveillance cameras. There is a significant data distribution shift when trying to adapt a VLM for real-site downstream tasks. Moreover, the fine-tuning of VLMs in the construction safety domain depends on both images and high-quality instruction-response sets for learning domain-specific knowledge. Efficient learning of domain knowledge with limited vision and text data remains a critical research gap that needs to be addressed by special fine-tuning strategies.

2. METHOD

To address these fundamental challenges in adapting VLMs for construction safety applications, we propose a comprehensive framework that systematically tackles both the data scarcity issue and the domain adaptation problem. Our methodology, illustrated in Figure 1, encompasses three innovative modules specifically designed to overcome the identified limitations:

- (1) A semi-automatic image-text data labeling pipeline to enhance the data preparation efficiency;
- (2) A two-stage curriculum learning framework to integrate domain knowledge into the VLM;
- (3) A real-time analytical system for automatic incident reporting and interactive VQA.

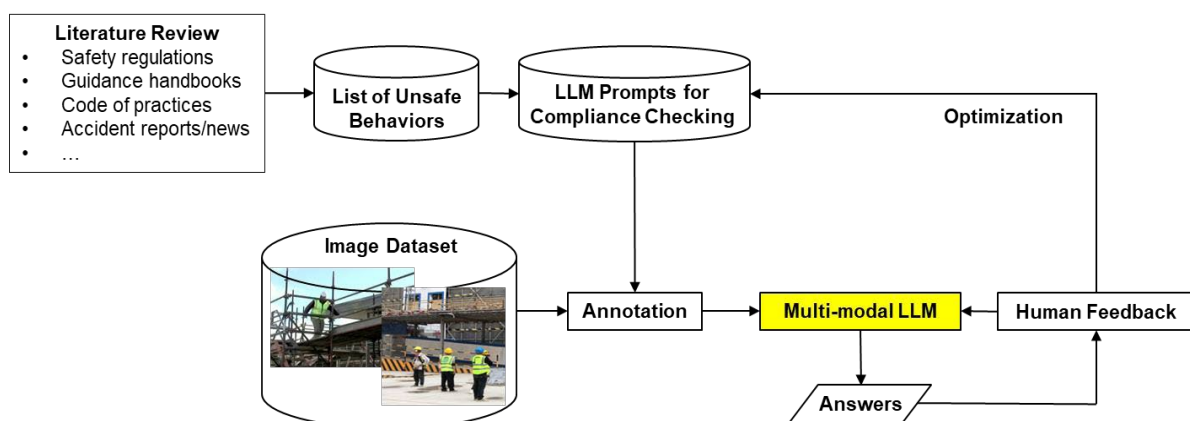


Figure 1: Research Methodology for Developing a VLM-based Safety Monitoring System

2.1 Semi-automatic Data Labeling Pipeline

This section describes the multi-step process to create a high-quality, multi-modal dataset for fine-tuning the VLM for safety compliance and monitoring on construction sites. A semi-automatic pipeline is developed to facilitate image-text data labeling, driven by an in-context learning framework, as illustrated in Figure 2.

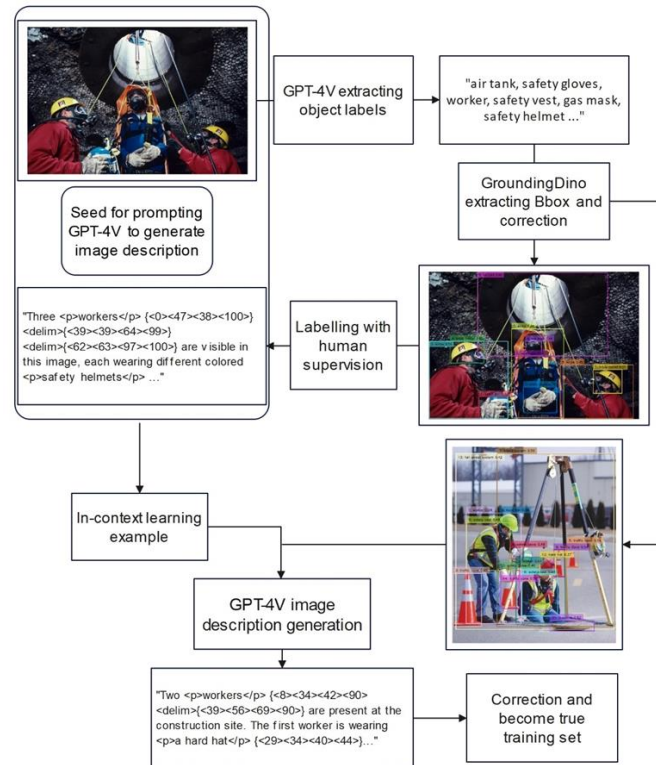


Figure 2: A Self-instructed Pipeline for Semi-automatic Image-text Data Labeling

Specifically, the semi-automatic data labeling pipeline will perform several tasks:

- 1) Automatic Label and Bounding Box Generation: GPT-4V published by OpenAI is utilized to generate preliminary descriptions (pseudo labels) of each image in batches. Figure 2 shows the assisting pipeline utilizing GPT-4V for effectively extracting candidate object labels or phrases and also a strong open-set object detector named GroundingDINO (S. Liu et al., 2024) for extracting candidate bounding boxes for identified objects.
- 2) Manual Correction of Pseudo-Labels: The automatically generated pseudo-labels and bounding boxes will be manually corrected to ensure accuracy and reliability.
- 3) Seed Image Selection: A seed image will be chosen from the images of each scenario (e.g., confined space as shown in Figure 3), and the description of the seed image will be hand-crafted based on the corrected bounding box labels either as grounded captioning (detailed descriptions) or safety compliance analysis examples.

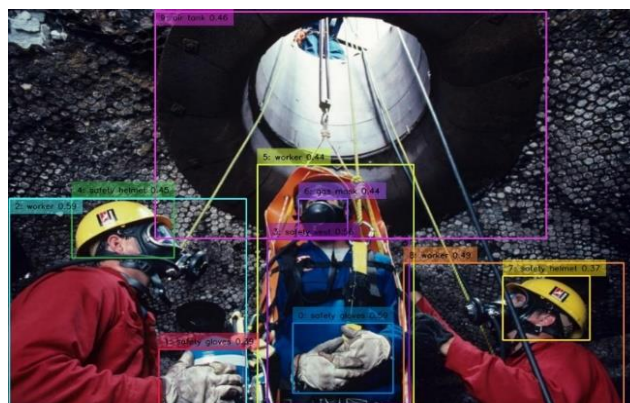


Figure 3: A Seed Image Generated by GPT-4V+GroundingDINO

- 4) In-context Learning: The annotated seed images will be included in the in-context learning prompt of GPT-4V to generate candidate image descriptions for remaining images. The in-context learning prompt for automatic dataset generation is shown below:
- 5) Polishing Candidate Descriptions: The candidate image descriptions generated by GPT-4V will be

further polished to create more desired responses in the training dataset triplets. These triplets will contain the images, instructions, and the final, polished responses. Lastly, the candidate image description will be further polished to the true desired responses in the training dataset triplets containing images, instructions, and responses. The styles of the instructions across different images will be changed to introduce higher diversity. This diversification of the instructions can help the VLM models learn more robust and generalizable capabilities.

2.2 Two-stage Curriculum Learning Framework for Effective VLM Training

The architecture of CogAgent (Hong et al., 2023) is chosen as our baseline model, as illustrated in Figure 4, CogAgent is built on a pre-trained Vision Language Model (VLM), specifically the CogVLM 17B, which is an open-source and state-of-the-art large vision language model.

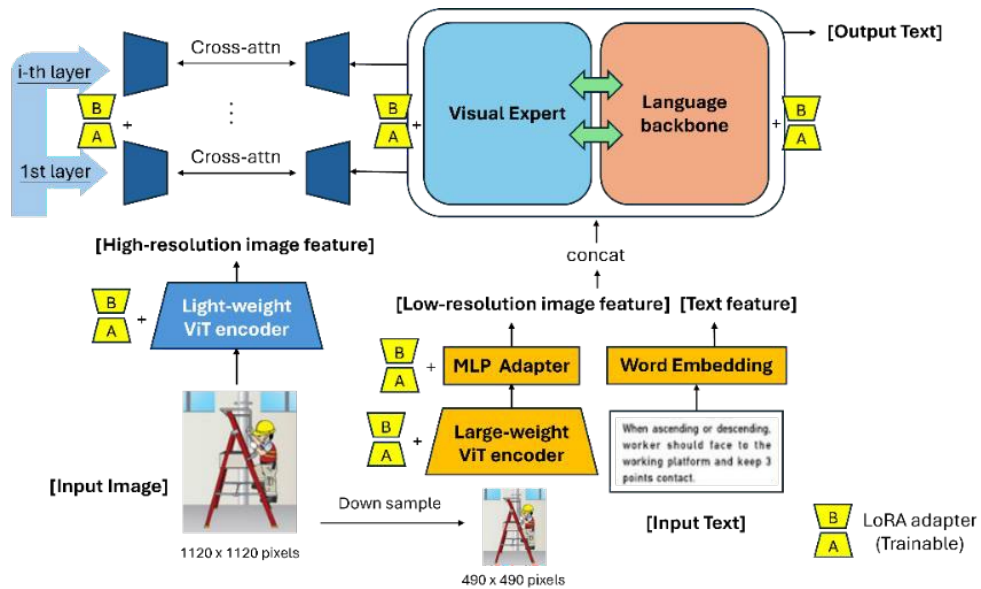


Figure 4: Architecture of the Proposed VLM Architecture (Revised from CogAgent (Hong et al., 2023))

It uses EVA2-CLIP-E as the encoder for low-resolution images (490×490 pixels), and an MLP adapter to map its output into the feature space of the visual-language decoder. The visual-language decoder consists of both vision experts and language backbone for effective deep fusion between image feature and language feature. The decoder processes a combined input of the low-resolution image feature sequence and text feature sequence, and autoregressively outputs the target text. However, the original CogVLM can only accommodate images of relatively low resolution (224 or 490), which is insufficient for screen resolution of computers or smart devices is typically 720p (1280 × 720 pixels) or higher. To address this, a high-resolution cross-module is introduced in their CogAgent model architecture, which not only maintains efficiency with high-resolution images but also offers flexible adaptability to a variety of visual-language model architectures. The high-resolution cross-module acts as a new branch for higher-resolution input, accepting images of size 1120 × 1120 pixels. Unlike the original low-resolution input branch, the high-resolution cross-module adopts a much smaller pre-trained vision encoder (visual encoder of EVA2-CLIP-L in our implementation, 0.30B parameters), and uses cross-attention of a small hidden size to fuse high-resolution image features with every layer of VLM decoder, thus reducing the computational cost.

In terms of the attention procedure, each layer's attention module is formulated in Equations (1) and (2):

$$X'_i = \text{MSA}(\text{layernorm}(X_{in}^i)) + X_{in}^i \quad (1)$$

$$X_{out}^i = \text{MCA}(\text{layernorm}(X'_i), X_{hi}) + X'_i \quad (2)$$

MSA and MCA respectively represent multi-head self-attention with visual expert and multi-head cross-attention. The cross-attention with high-resolution images can be perceived as a complement to the features of low-resolution images, thereby effectively utilizing the previous pre-trained model in low resolution. CogAgent exhibits exceptional performance on benchmarks that assess referring expression comprehension (REC), such as RefCOCO, RefCOCO+, and RefCOCOg, which is comparable to grounding specialists like GroundingDINO.

Current studies adopted a staged learning scheme to train VLMs more effectively. For instance, MiniGPT-v2 was trained with three stages (pretraining, multi-task training and multi-modal instruction tuning) for better visual-textual feature alignment (Chen et al., 2023). Inspired by such principle, a two-stage curriculum learning strategy is proposed to equip our VLM with domain knowledge and capabilities in multiple safety-related tasks:

1) The first stage involves pre-training of our model on construction site images with grounded captioning labels, for learning construction-specific domain knowledge given the labeled construction site objects.

2) The second stage focuses on fine-tuning the VLM toward site-specific safety rules to gain task-specialized performance in safety compliance inspection from images.

This two-stage paradigm ensures that our VLM not only understands the nuances of the language but also adapts to the specific requirements of the task. The dataset decreases in volume but increases in supervision as the training stage progresses. The final stage of instruction tuning aimed to enhance the instruction following and conversation ability of CogVLM.

We also designed a curriculum learning scheme to align the base CogAgent model with advanced safety inspection ability. The first stage of the training mainly involved training the model on grounded captioning datasets of construction images to equip it with construction domain knowledge. Afterwards, the second stage of the training aimed to enhance the inspection ability of the model by learning from specific safety rule compliance/violation cases.

A list of essential work-at-height behaviors from construction documents, including the Safety Manual established by Drainage Services Department (2018), the Work-at-Height Safety Handbook published by Development Bureau and Construction Industry Council (2019) of Hong Kong are reviewed to extract some key safety rules for working at height. The list of work-at-height behaviors are:

a. Powered-Operated Elevating Working Platforms (PEWP)

- i. Wear full body safety harness with its lanyard anchored to a specified anchorage point

b. Metal Scaffolds

- i. When erecting, altering dismantling of scaffolds or it is impracticable to erect a safe working platform or provide safe access and egress, the use of full body safety harness attached to secure anchorage point or an independent lifeline is required

c. Light-Duty Working Platform/ Mobile Working Platform

- i. When ascending or descending, worker should face to the working platform and keep 3 points contact
- ii. Only three types of platforms are allowed to carry out work-at-height tasks, including hop-up platform, step platform and mobile platform
- iii. The surrounding of working platforms should be kept free from waste and miscellaneous materials

d. Floor Opening/Edge Protection

- i. Provide guard-rails and toe-boards at the floor edge
- ii. Provide secure coverings with warning signs at the floor opening.
- iii. Provide guard-rails, toe boards and warning signs at the floor opening
- iv. Whist installing, alternating or dismantling fall protection facilities at the floor edge, opening and windows, suitable fall arresting system should be provided to workers

2.3 Real-time on-site analytical system

The VLM-based safety compliance monitoring system operates in two different modes: (1) automatic reporting and (2) interactive chatbot. In Mode 1, the real-time video from each camera is streamed continuously to the computer, where video frames are sampled at regular intervals to be fed into the VLM for inference (10 seconds in our field trials). Then, the VLM automatically generates appropriate answers to the image. On the other hand, Mode 2 can be user-triggered via a simple button to extract the current frame at any time. Based on that frame, safety officers can input any question to prompt the VLM to interactively generate personalized answers.

An alarm mechanism is developed to generate descriptive alert messages to safety officers. As illustrated in Figure 5 (left), in case any safety rule violation is identified, such “Failed” safety compliance is reported via an alarm from the onsite computer, where the safety officers are notified in real-time. A descriptive alert message is then prompted out, concisely summarizing which specific safety rules are violated. In case no violation is observed from an image, as illustrated in Figure 5 (right), it passes the safety compliance checking by our VLM which results in a simple message “No Safety Risk” output without triggering the alarm system.

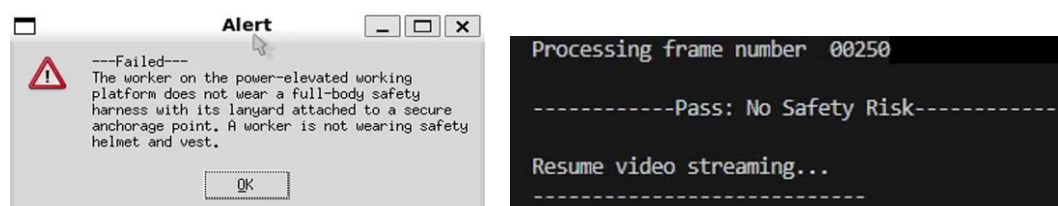


Figure 5: Illustrations of With-Alert and No-Alert Messages Generated by our Alarm System

In addition, an incident logging system is developed to store all the processed images and generated output. These data are properly indexed/labeled, which constitute a well-managed database for record-keeping, which also allows safety officers to query historical incidents and search for associated images for verification. As illustrated in Figure 5, each processed frame is indexed with a number (e.g. “00250” denotes the 250th frame of a video stream), which serves as the key for future queries. All these indexed frames are stored, each of which is then labeled with the final output generated by the VLM, which serve as the value for future queries.

3. RESULTS AND DISCUSSION

3.1 Dataset and Experimental Setup

The developed VLM system is deployed at a real construction site in the Shek Wu Hui Sewage Treatment Works. This site is a secondary sewage treatment plant occupying 9.4 hectares of land and handling 81,000 m³ of sewage per day produced by a population of 300,000 in Sheung Shui and Fanling Districts.

The dataset for fine-tuning the CogAgent model consists of 1,500 construction site images, after applying augmentation techniques including random cropping, horizontal flipping and color space alternation. The collected images are then annotated with the assistance of the pipeline mentioned in Section 2.1. As shown in Figure 6, the raw images are first processed by the proposed methodology to obtain accurate bounding boxes for objects with precise and diverse textual descriptions. The textual descriptions are based on the list of safety questions summarized. Some features/objects in the images are highly relevant to certain safety rules (e.g. scaffolds, working platforms), thus are specifically highlighted in the annotations to incorporate such knowledge into our VLM.

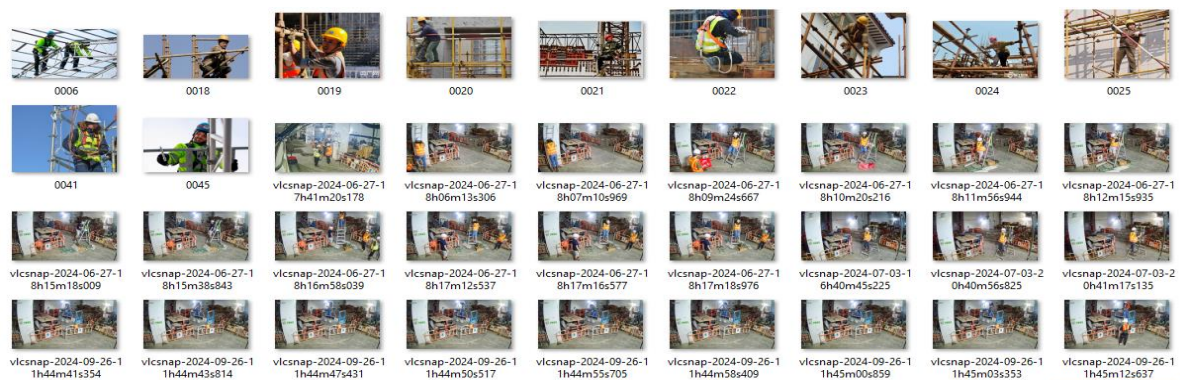


Figure 6: Snapshot of the Collected Image Dataset for VLM Fine-tuning and Testing

The data split is 6:4, i.e. 60% (900 images) for fine-tuning and validation, the remaining 40% (600 images) for testing. The fine-tuning was done with a single Nvidia A40 GPU with 48 GB of available VRAM. For the training configuration, Low-Rank Adaptation (LoRA) is used for Parameter-Efficient Fine-Tuning (PEFT) while all the original weights are frozen. 5 types of LoRA modules, each with a rank of 64, and a LoRA alpha value of 128, are injected into the visual language backbone, the high-resolution cross-attention layers, the large ViT encoder, the MLP adapter and the lightweight ViT encoder respectively. The VLM was trained with a batch size of 2 for 2 epochs for each of the two stages, with a warmup ratio of 0.1, an initial learning rate of $3e-5$, and a cosine learning rate scheduler. The lora dropout is set to 0.1 and the weight decay is set to 0.05. After finetuning, the VLM was quantized to 8 bit and deployed in one RTX-4090 GPU with 24 GB of VRAM.

3.2 Quantitative Evaluation

The collected images are fed into our VLM to carry out VQA. The performance of safety compliance checking is evaluated with the protocol summarized in Table 1. In the context of safety compliance checking in this project, two possible outcomes are defined: (1) *Compliance – Pass*, where no safety rule is violated, and (2) *Compliance – Failed*, where a particular safety rule is violated. Therefore:

- True Positive (TP) means that an image does contain violation scenario(s), and the VLM correctly identifies them. Note that the evaluation is counted on per-rule basis rather than per-image basis.
- False Negative (FN) means that an image does contain violation scenario(s), but the VLM cannot identify them and wrongly output “Pass – No Safety Risk”.
- False Positive (FP) means that an image has no safety violation, but the VLM wrongly outputs “Failed” for a particular safety rule.
- True Negative (TN) means that an image has no safety violation, and the VLM correctly outputs “Pass – No Safety Risk”.

These metrics form the foundation for our two primary evaluation measures, Sensitivity and Specificity, defined in Equations (3) and (4):

- *Sensitivity* denotes the percentage of violation scenarios correctly identified by the VLM.
- *Specificity* denotes the percentage of compliant scenarios correctly passed without any false alert.

Table 1: Proposed Evaluation Metrics for Safety Compliance Checking

<i>Ground-truth \ Output</i>	<i>Compliance – Failed</i>	<i>Compliance – Pass</i>
<i>Compliance – Failed</i>	<i>TP</i>	<i>FN</i>
<i>Compliance – Pass</i>	<i>FP</i>	<i>TN</i>

*The values denote the number of safety rules being checked one-by-one

$$\text{Sensitivity (\%)} = \frac{TP}{TP + FN} \quad (3)$$

$$\text{Specificity (\%)} = \frac{TN}{TN + FP} \quad (4)$$

In the context of construction safety monitoring, these metrics carry significant practical implications. Sensitivity measures the system's ability to correctly identify actual safety violations, such as missing PPE or unsafe behaviors at heights. This metric is particularly crucial as missing violations (false negatives) could lead to serious accidents or fatalities. Complementarily, specificity measures the system's accuracy in correctly identifying safe conditions while avoiding false alarms. High specificity is essential for maintaining operational efficiency, as false positives can cause unnecessary work interruptions and diminish trust in the monitoring system.

Our analysis reveals significant improvements through the two-stage curriculum learning framework. The baseline model initially showed sensitivity and specificity rates of 76.3% and 74.3% respectively. Our enhanced framework substantially improved these metrics to 84.7% and 91.7%. Statistical validation through Fisher's Exact Test yielded p-values of 0.0132 for sensitivity and <0.0001 for specificity, confirming the statistical significance of these improvements (both <0.05).

The detailed results in Tables 3 and 4 demonstrate the tangible impact of our approach, showing an increase of 77 correct samples (25 TP + 52 TN), representing approximately 13% of the testing set. This improvement reflects enhanced accuracy in safety compliance identification and validates the effectiveness of our integrated approach for construction safety monitoring.

Table 2: Quantitative Results among the Baseline and Proposed Methods

Method	Sensitivity	Specificity
Baseline	76.3% (71.1% ~ 81.0%)	74.3% (69.0% ~ 79.2%)
Two-stage Learning	84.7% (80.1% ~ 88.6%)	91.7% (87.9% ~ 94.5%)

*The ranges inside brackets denote the results at 95% Clopper-Pearson confidence intervals

Table 3: Contingency Table for Sensitivity of Safety Compliance Checking

Method	<i>TP</i>	<i>FN</i>	Total
Baseline	229	71	300
Two-stage Learning	254	46	300

Table 4: Contingency Table for Specificity of Safety Compliance Checking

Method	<i>TN</i>	<i>FP</i>	Total
Baseline	223	77	300
Two-stage Learning	275	25	300

3.3 Qualitative Evaluation

Based on the safety rules listed in Section 2.2, the output generated by our VLM (for compliance “Pass” and “Failed” respectively) are illustrated below.

Case 1: Working on Light-duty Mobile Platform (Safe Ascending/Descending Required)

- “Pass”: In Figure 7, the worker ascending/descending the step platform is facing directly to the platform and maintaining three-point contact with it. Hence, the safety compliance checking correctly results in a “Pass”.
- “Failed”: As shown in Figure 8, the worker is facing away from the step platform, and his hands are not holding the step platform safely. In that cases, the alarm system is triggered with an alert message prompted out, stating the safety rules being violated.

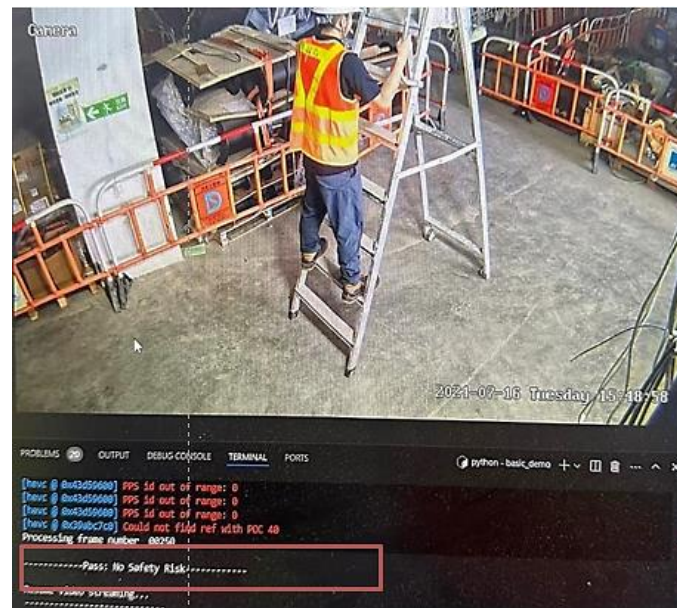


Figure 7: Result of Facing to the Platform and Three-point Contact Compliance (“Pass”)

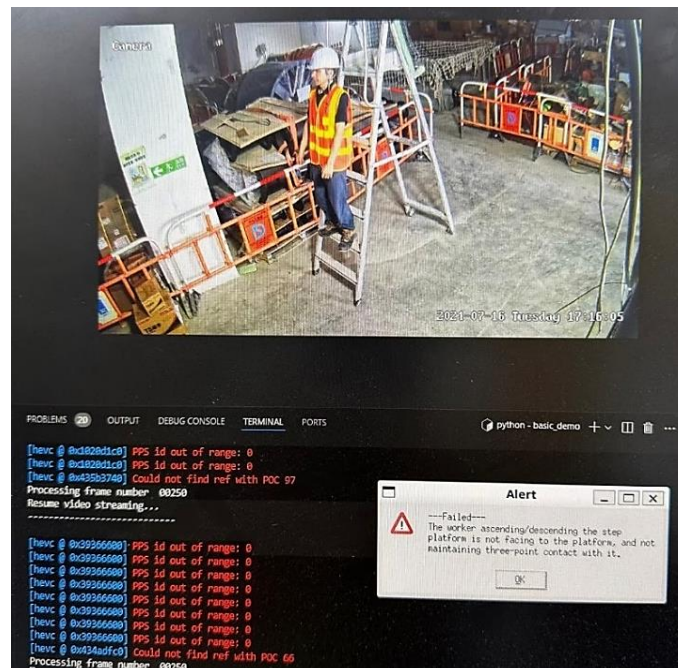


Figure 8: Result of Facing to the Platform and Three-point Contact Compliance (“Failed”)

For monitoring worker safety when climbing the mobile platform, our VLM can analyze the orientation of the worker’s face and body relative to the mobile platform. However, the identification of three-point contact with the mobile platform is slightly more challenging. The worker’s limbs may be partially occluded when climbing the mobile platform (e.g. when facing to the right, his left hand and left leg are occluded respectively by his body and right leg). More systematic prompt engineering may be explored in the future to further enhance the identification accuracy.

Case 2: Working on PEWP (Safety Harness Required)

- “Pass”: As shown in Figure 9, the worker standing on top of the PEWP is wearing a safety harness, with its lanyard attached to a secure anchorage point. The VLM simply prints “Pass: No Safety Risk” on the screen to verify his safety compliance, without triggering any alarm.
- “Failed”: As shown in Figure 10, the worker on the PEWP is not wearing a safety harness. Our VLM correctly identified such safety violation and generated the corresponding alert message.

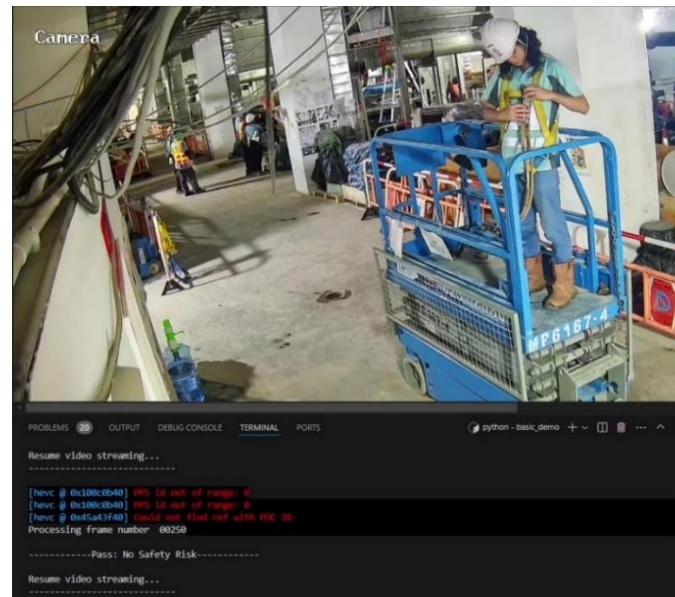


Figure 9: Result of Safety Harness and Anchorage Compliance (“Pass”)

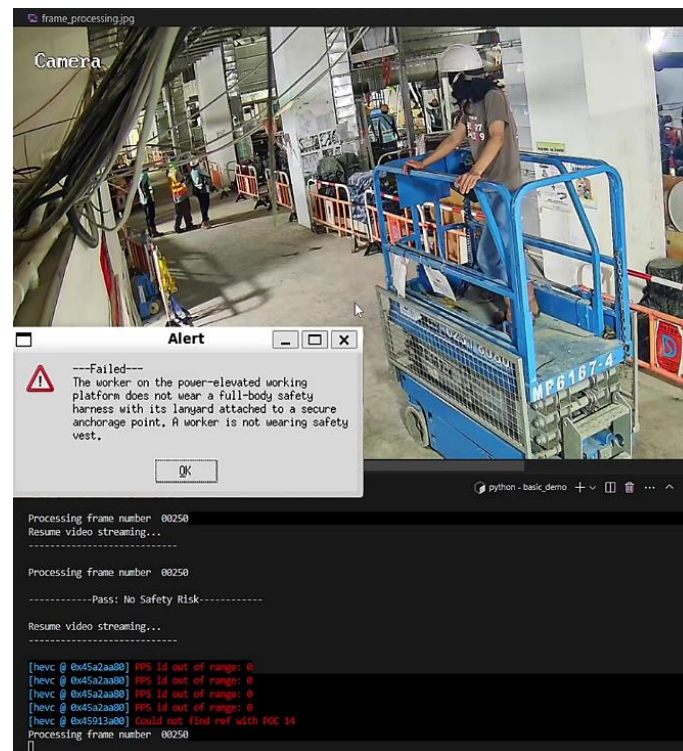


Figure 10: Result of Safety Harness and Anchorage Compliance (“Failed”)

Overall, our VLM can accurately identify the presence/absence of safety harness when being worn by a worker. However, it remains challenging to determine whether the harness is attached to a secure anchorage point (in the site trial, attaching to the PEWP is considered safe), due to the very thin rope and small size of the anchor. More crucially, the anchor is usually occluded by the worker or the PEWP itself, making the anchorage detection inaccurate. In the future, more sophisticated strategies of image/video analytics, such as small-object attention mechanism, can be further incorporated into our framework, to enhance the harness detection robustness.

Nevertheless, the preliminary results show that the CogAgent model show some hallucination when transferring the visual information into textual information. This may be due to the frozen visual encoder and shallow alignment method in the model architecture during the model training. Further investigation may be needed on the multi-modal feature alignment capability and contextual awareness toward safety monitoring.

4. CONCLUSION AND FUTURE WORK

This paper addresses critical challenges in adapting VLMs for construction safety monitoring through a comprehensive framework. Our semi-automatic pipeline, combining GPT-4V and GroundingDINO for image captioning, successfully overcame data scarcity. Our two-stage curriculum learning framework demonstrated remarkable effectiveness in domain knowledge integration, achieving 84.7% sensitivity and 91.7% specificity in work-at-height safety compliance checking, substantially improved over the baseline method.

These findings have significant implications for both research and industry. Our framework provides a scalable solution for adapting AI systems to specialized domains with limited data availability, while the demonstrated success in real-world deployment establishes a practical pathway for automating construction safety monitoring. The high accuracy achieved in safety compliance inspection suggests potential for widespread adoption in construction site management, enabling more proactive and efficient safety protocols. Future research will focus on enhancing the system through architectural modifications for improved multi-modal feature alignment and contextual awareness. The framework's generalizability will be validated across diverse safety rules, and be extended to real-time video analytics for automated safety monitoring practices.

ACKNOWLEDGMENTS

The authors would like to acknowledge the project (No. *DEMP/2023/25*) funded by *Drainage Services Department, the Government of the Hong Kong Special Administrative Region, Wanchai, Hong Kong Island, Hong Kong, SAR*, for providing support to this research. The authors declare no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

- Chen, J., Zhu, D., Shen, X., Li, X., Liu, Z., Zhang, P., Krishnamoorthi, R., Chandra, V., Xiong, Y., & Elhoseiny, M. (2023). *MiniGPT-v2: Large language model as a unified interface for vision-language multi-task learning* (arXiv:2310.09478). arXiv. <https://doi.org/10.48550/arXiv.2310.09478>
- Cheng, Jack C. P., Wong, P. K.-Y., Luo, H., Wang, M., & Leung, P. H. (2022). Vision-based monitoring of site safety compliance based on worker *re*-identification and personal protective equipment classification. *Automation in Construction*, 139, 104312. <https://doi.org/10.1016/j.autcon.2022.104312>
- Fang, W., Ma, L., Love, P. E. D., Luo, H., Ding, L., & Zhou, A. (2020). Knowledge graph for identifying hazards on construction sites: Integrating computer vision with ontology. *Automation in Construction*, 119, 103310. <https://doi.org/10.1016/j.autcon.2020.103310>
- Hong, W., Wang, W., Lv, Q., Xu, J., Yu, W., Ji, J., Wang, Y., Wang, Z., Zhang, Y., Li, J., Xu, B., Dong, Y., Ding, M., & Tang, J. (2023). *CogAgent: A Visual Language Model for GUI Agents* (arXiv:2312.08914). arXiv. <https://doi.org/10.48550/arXiv.2312.08914>
- Labour Department. (2018). *Occupational Safety and Health Statistics 2018*. Occupational Safety and Health. <https://www.labour.gov.hk/eng/osh/pdf/Bulletin2017.pdf>
- Labour Department. (2019). *Occupational Safety and Health Statistics 2019*. Occupational Safety and Health. https://www.labour.gov.hk/eng/osh/pdf/archive/statistics/OSH_Statistics_2019_eng.pdf
- Liu, H., Li, C., Li, Y., & Lee, Y. J. (2024). *Improved Baselines with Visual Instruction Tuning* (arXiv:2310.03744). arXiv. <https://doi.org/10.48550/arXiv.2310.03744>
- Liu, S., Zeng, Z., Ren, T., Li, F., Zhang, H., Yang, J., Jiang, Q., Li, C., Yang, J., Su, H., Zhu, J., & Zhang, L. (2024). *Grounding DINO: Marrying DINO with Grounded Pre-Training for Open-Set Object Detection* (arXiv:2303.05499). arXiv. <https://doi.org/10.48550/arXiv.2303.05499>
- Occupational Safety and Health Administration. (2019). *Commonly Used Statistics | Occupational Safety and Health Administration*. OSHA Data & Statistics. <https://www.osha.gov/data/commonstats>
- OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V., Baltescu, P., Bao, H., Bavarian, M., Belgum, J., ... Zoph, B. (2024). *GPT-4 Technical Report* (arXiv:2303.08774). arXiv. <https://doi.org/10.48550/arXiv.2303.08774>
- Paneru, S., & Jeelani, I. (2021). Computer vision applications in construction: Current state, opportunities & challenges. *Automation in Construction*, 132, 103940. <https://doi.org/10.1016/j.autcon.2021.103940>
- Tang, S., Roberts, D., & Golparvar-Fard, M. (2020). Human-object interaction recognition for automatic construction site safety inspection. *Automation in Construction*, 120, 103356. <https://doi.org/10.1016/j.autcon.2020.103356>
- Wang, W., Lv, Q., Yu, W., Hong, W., Qi, J., Wang, Y., Ji, J., Yang, Z., Zhao, L., Song, X., Xu, J., Xu, B., Li, J., Dong, Y., Ding, M., & Tang, J. (2024). *CogVLM: Visual Expert for Pretrained Language Models* (arXiv:2311.03079). arXiv. <https://doi.org/10.48550/arXiv.2311.03079>
- Wang, Y., Xiao, B., Bouferguene, A., Al-Hussein, M., & Li, H. (2022). Vision-based method for semantic information extraction in construction by integrating deep learning object detection and image captioning. *Advanced Engineering Informatics*, 53, 101699. <https://doi.org/10.1016/j.aei.2022.101699>

Exploring Construction 5.0 Paradigms in the AEC Sector of New Zealand: Conceptual Foundations and Implications for Practice

Xichen Chen^{1*}, Ali Ghaffarianhoseini², and Amirhosein Ghaffarianhoseini³

1) Ph.D., Lecturer, School of Future Environments, Auckland University of Technology, New Zealand. Email: xichen.chen@aut.ac.nz

2) Ph.D., Professor, School of Future Environments, Auckland University of Technology, New Zealand. Email: ali.ghaffarianhoseini@aut.ac.nz

3) Ph.D., Professor, School of Future Environments, Auckland University of Technology, New Zealand. Email: amirhosein.ghaffarianhoseini@aut.ac.nz

Abstract: The increasing emphasis on integrating technological advancements with human-centered and sustainable practices highlights the paradigm shift toward Construction 5.0 (C5.0) in the architecture, engineering, and construction (AEC) sector. Despite its potential, investigations about C5.0's key pillars, practical implications, and adoption challenges remain limited, with much existing research focusing on conceptual frameworks or literature reviews. This study addresses these gaps through an empirical investigation, incorporating insights from a focus group of 17 industry practitioners to explore C5.0's key pillars, core features, technological enablers, and implications. The findings highlight three core features of C5.0: human-centricity, sustainability-driven practices, and collaborative intelligence. Seventeen emerging digital technologies were identified as critical enablers of C5.0, with artificial intelligence/machine learning, digital twins, and collaborative robots ranked as the most impactful technologies. These technologies support 31 application domains and enable AEC organizations to achieve enhanced productivity, innovation, sustainability, worker safety and well-being, and competitive advantage. Under the enhanced sustainability category, improved compliance with environmental regulations and increased capacity to meet client demands for sustainable practices were emphasized as key outcomes. This study contributes both theoretically and practically to the understanding of C5.0. Theoretically, it defines the key pillars, core features, and technological enablers of C5.0, bridging gaps in the existing literature and advancing the academic discourse on the evolution to C5.0. Practically, it offers a roadmap for integrating critical technologies with human-centered and sustainability goals, enabling AEC practitioners to prioritize investments effectively. Future research should expand empirical studies to conduct a cost-benefit evaluation of C5.0 technologies and explore C5.0's impact on project management methodologies, stakeholder collaboration, and organizational strategy development.

Keywords: Industry 5.0 (I5.0), Construction 5.0 (C5.0), Digital Technology, Architecture, Engineering and Construction (AEC), Concept, Implication.

1. INTRODUCTION

The architecture, engineering, and construction (AEC) industry is a major contributor to global economic growth, accounting for approximately 13% of global GDP (McKinsey & Company, 2020). However, it remains one of the least digitized sectors and continues to face challenges such as low productivity, high waste generation, and persistent safety concerns (Abioye et al., 2021; World Economic Forum, 2023). Further, the industry is responsible for nearly 40% of global greenhouse gas emissions (Ürge-Vorsatz et al., 2020), emphasizing a need for transformative practices to meet global sustainability targets. Addressing these challenges requires a transition towards greener and more proactive approaches in AEC practices, driven by advancements from technological revolutions that have evolved from mechanization to digitalization under the Construction 4.0 (C4.0) paradigm (Chen et al., 2023).

C4.0 introduced digital tools and solutions such as building information modeling (BIM), the Internet of Things (IoT), and digital fabrication, revolutionizing construction through automated and smart processes (Sawhney et al., 2020). However, C4.0 primarily focuses on efficiency, productivity, precision, and cost-effectiveness, while lacking a holistic approach to sustainability, particularly in addressing long-term environmental and social impacts. Inspired by the principles of Industry 5.0 (I5.0), the emerging Construction 5.0 (C5.0) paradigm highlights human-centricity, sustainability, and resilience (Yitmen et al., 2023). C5.0 seeks to transform the sector by integrating advanced technologies such as artificial intelligence (AI), digital twins, and collaborative robotics (cobots) to enhance operational efficiency, support environmental stewardship, and prioritize worker well-being (Müller, 2020).

The increasing focus on integrating technological advancements with human-centered and sustainable practices indicates a shift towards adopting C5.0 principles. For instance, Marinelli (2023) identified human-robot collaboration (HRC) as a central theme, while Yitmen et al. (2023) emphasized C5.0's role in promoting resource efficiency and client-tailored solutions. Similarly, Tunji-Olayeni et al. (2024) examined C5.0's potential for improving health, safety, and sustainability outcomes in construction projects, and Bello et al. (2024) investigated how transitioning from C4.0 to C5.0 can strengthen supply chain resilience through a human-centric approach.

Although C5.0 is regarded as a critical future direction for the AEC sector, several gaps persist in terms

of practical implementation. These include a limited understanding of its key pillars, their practical implications for the sector, and the challenges involved in its adoption. Realizing the full potential of C5.0 requires collaborative efforts in education, technological integration, and cultural transformation. However, most current studies are predominantly conceptual or review-based, focusing on theoretical frameworks rather than empirical evidence from industry practices (e.g., Cisneros-Gonzalez et al., 2024; Ikudayisi et al., 2023). While early studies have examined the technological and economic dimensions of C5.0 (Biswas et al., 2024; Rane, 2023), ethical and societal concerns, such as data privacy and job displacement, remain underexplored. Further, despite the sustainability benefits highlighted in recent studies (e.g., Bello et al., 2024; Hu et al., 2023), there is a lack of comprehensive metrics or frameworks to evaluate the long-term environmental, social, and economic impacts of C5.0 technologies. Against this backdrop, this study investigates the foundations and implications of C5.0 in the AEC sector through an empirical approach that engages industry stakeholders. Specifically, it seeks to answer the following research questions:

1. What are the key pillars of C5.0, including its core features and technological enablers? How are these technologies implemented, and how do they align with its core features?
2. What implications does the C5.0 paradigm hold for AEC organizations, and how will it affect their project management practices?

2. LITERATURE REVIEW

2.1 Industry 5.0 and Construction 5.0 Paradigms

The I5.0 concept emerges as a strategic initiative to address the limitations of Industry 4.0 (I4.0) while advancing digital and green transitions in the manufacturing sector. According to the European Commission (Müller, 2020), I5.0 introduces a human-centered and value-driven perspective, prioritizing technologies that align with ethical principles and societal needs. It emphasizes a balance between technological advancement and the well-being of people and the environment. Ivanov (2023) highlights I5.0 as a framework connecting resilience, sustainability, and human-centricity to create adaptable and regenerative industrial ecosystems. According to Zizic et al. (2022), I5.0 integrates stakeholder-driven socio-technological changes, focusing on sustainable production, workplace inclusion, and ethical considerations. Maddikunta et al. (2022) describes I5.0 as employing human creativity in interaction with intelligent and efficient machines, enabling hyper-customization and resource-efficient manufacturing. Core technologies include cobots, digital twins, wearable technology, and 5G/6G networks, which allow efficient human-machine collaboration and adaptive industrial processes (Chen et al., 2024a). Ghobakhloo et al. (2022) further addresses the focus of I5.0 on human dignity and workplace equality while deploying cognitive cyber-physical systems (CPS) to achieve sustainable industrial practices. Xu et al. (2021) positions I5.0 as an evolution that combines human intelligence and advanced technologies to enhance inclusivity, resilience, and innovation in manufacturing systems. These perspectives establish I5.0 as a transformative model, advancing beyond automation to balance technological progress with societal and environmental priorities.

C5.0 builds upon C4.0 by integrating the human-centric, collaborative, and sustainability-focused principles of I5.0, positioning it as the AEC industry's counterpart to I5.0 (Marinelli, 2023; Yitmen et al., 2023). C5.0 combines human expertise with advanced technologies such as digital twins and cobots to enhance efficiency, resilience, and tailored project delivery (Nahavandi, 2019; Yitmen et al., 2023). Central to this evolution is HRC, which addresses complex and hazardous tasks, enabling safer and more precise operations in dynamic environments (Ohueri et al., 2024). Furthermore, C5.0 inherits I5.0's emphasis on sustainability and resilience, with applications in waste management, supply chain adaptability, and environmental forecasting to meet societal and ecological goals (Bello et al., 2024; Tunji-Olayeni et al., 2024). By integrating advanced technologies with human ingenuity, C5.0 redefines construction processes to address the industry's unique challenges while promoting sustainable and collaborative innovation.

C5.0 emphasizes HRC through cohesive interaction enabled by advanced communication protocols and real-time data exchange, addressing dimensions of human well-being and the societal impact unacknowledged in previous paradigms (Ohueri et al., 2024). Grounded in sustainability, human-centricity, and resilience, C5.0 marks a shift from the focus of C4.0 on automation to a more holistic and technology-driven framework supported by empirical modeling (Marinelli, 2023; Yitmen et al., 2023). Although applications are primarily at the prototype stage, they lay a strong foundation for practical implementation (Chen et al., 2024b). For example, integrated systems utilizing human digital twins, smart sensors, and AI enhance worker safety and well-being by monitoring workers' physical, mental, and emotional states (Davila-Gonzalez & Martin, 2024). Similarly, virtual-real interaction models using 3D modeling, BIM, and digital twins are being validated for various construction stages (Wang et al., 2022). These developments emphasize the transformative potential of C5.0 in advancing sustainable construction practices and workforce development.

Despite its transformative potential, the implementation of C5.0 faces considerable technological, organizational, and cultural challenges. Limited digitalization, insufficient R&D investment, and fragmented workflows hinder the cohesive integration of advanced technologies like IoT, AI, robotics, and BIM, particularly

in dynamic and uncontrolled construction environments (Brozovsky et al., 2024; Chen et al., 2024c; Marinelli, 2023). Workforce readiness is also a critical challenge, requiring extensive upskilling and interdisciplinary training to effectively use cobots, exoskeletons, and neuro-responsive systems (Almusaed et al., 2024; Hadi et al., 2023). Additionally, resistance to change, cultural apprehensions about automation, and ethical concerns surrounding privacy and data security further complicate adoption (Musarat et al., 2023; Rane, 2023). Sustainability, a core pillar of C5.0, demands integrating lean principles and circular economy practices like design for disassembly and material passports, which remain underdeveloped (Hadi et al., 2023). Collaboration between academia and industry is essential to align technological advancements with practical applications and promote skill development (Brozovsky et al., 2024). Generative AI and advanced systems offer opportunities to enhance decision-making, collaboration, and occupant-centered designs, but scalability and ethical challenges persist (Almusaed et al., 2024; Rane, 2023). Overcoming these barriers is critical for the AEC sector to fully realize the sustainable, human-centric, and resilient transformation envisioned by C5.0.

2.2 Implications of Construction 5.0

C5.0 technologies aim to revolutionize operational efficiency, decision-making, and sustainability by integrating advanced tools such as AI, IoT, robotics, and digital twins (Almusaed et al., 2024; Bello et al., 2024). Adopting C5.0 technologies requires a practical understanding of their applications, benefits, limitations, and the connection between human expertise and advanced systems. Despite significant progress in developing robotic systems, AI-driven tools, and blockchain-based frameworks, challenges persist in bridging the knowledge gap between practitioners and technology developers, particularly in dynamic and fragmented construction environments (Hu et al., 2023; Ikudayisi et al., 2023; Rane, 2023).

Table 1 summarizes recent studies on C5.0 technologies, highlighting the diverse technologies explored and their implications across technological, social, economic, and environmental dimensions. Technologies such as AI, machine learning (ML), digital twins, and BIM demonstrate significant potential to improve safety, efficiency, and resource optimization but face barriers such as technical immaturity, workforce readiness, and integration costs. Other technologies, including cobots, HRC, Human-CPS (HCPS), and generative AI, remain underutilized due to limited practical knowledge and misconceptions about their capabilities. The findings emphasize the need for multi-disciplinary collaboration and targeted training to bridge these gaps and ensure the progressive adoption and implementation of C5.0 practices.

Table 1. Summary of studies on C5.0 in the existing literature

Source	Article Type	C5.0 Technology	Implication
Almusaed et al. (2024)	Questionnaire Survey	AI, BIM, Brain-Computer Interface, Digital Twins	Enhances comfort and well-being; streamlines processes; optimizes energy use; reduces emissions.
Bello et al. (2024)	Literature Review	AI, Cobots, IoT	Improves supply chain resilience; enhances safety; reduces physical strain; increases efficiency; supports sustainability.
Biswas et al. (2024)	Literature Review	BIM, Digital Twins, ICT, IoT, ML	Enhances lifecycle management; improves precision; encourages collaboration through advanced data integration.
Cisneros-Gonzalez et al. (2024)	Systematic Literature Review	Automation and Robotics, Cobots, Wearable Devices, Prefabrication	Boosts productivity and precision; reduces waste; streamlines processes; minimizes reliance on human labor.
Hadi et al. (2023)	Literature Review	Cobots, Lean-Offsite Construction, Exoskeletons	Improves efficiency; reduces delays and costs; emphasizes design for disassembly and resource recovery.
Hu et al. (2023)	Theoretical Analysis; Questionnaire Survey	Green Intelligent Building Materials	Reduces waste and emissions; improves cost-efficiency; enhances convenience and comfort for end-users.
Ikudayisi et al. (2023)	Literature Review	BIM, Blockchain, Digital Twins, Modular Integrated Construction	Facilitates integration; reduces waste; enhances collaboration and transparency; supports sustainable design.
Jiménez Rios et al. (2024)	Systematic Literature Review	AI, Digital Twins, HRC, Renewable Energy, Smart Materials	Improves cultural heritage conservation; enhances monitoring systems; addresses climate challenges; supports ethical and sustainable practices.
Marinelli (2023)	Literature Review	HRC	Increases precision and safety; promotes ergonomics; reduces costs; optimizes resource use.

Musarat et al. (2023)	Literature Review	AI, Big Data, Cobots, Digital Twins, IoT, Smart Materials	Improves efficiency and collaboration; enhances safety; reduces costs; supports eco-friendly practices.
Ohueri et al. (2024)	Systematic Literature Review	AI, HRC, VR	Enhances safety in hazardous processes; reduces risks and costs; promotes material reuse and recycling.
Rane (2023)	Literature Review	Generative AI	Accelerates decision-making; improves communication; optimizes costs and efficiency; supports sustainability.
Saradara et al. (2024)	Literature Review	Blockchain, Smart Relational Contracts	Promotes circularity and collaboration; improves stakeholder integration; reduces waste.
Tunji-Olayeni et al. (2024)	Literature Review	6G, AI, Big Data, BIM, Blockchain, Cobots, CPS, Digital Sensors, Wearable Devices	Enhances safety and waste management; improves environmental forecasting; emphasizes collaboration and sustainability.
Yitmen et al. (2023)	Questionnaire Survey	HRC, HCPS	Promotes human-machine collaboration; supports societal benefits; reduces waste; improves resource efficiency and sustainability.

3. RESEARCH METHODOLOGY

This study adopts a qualitative focus group method to understand the key pillars of C5.0 and examine its implications for AEC organizations and practices. The focus group approach is particularly well-suited for investigating emerging topics where existing knowledge of a subject is limited. Its interactive nature encourages the exchange of perspectives based on participants' prior knowledge while enabling the co-construction of new insights informed by their experiences (O.Nyumba et al., 2018). Given the varied needs, challenges, and levels of technological maturity across organizations and professional backgrounds, this method offers a more comprehensive understanding of how industry professionals perceive the core features of C5.0. It captures their literacy regarding C5.0 technologies while revealing firsthand accounts of the real-world challenges, barriers, and gaps professionals face in adopting C5.0 technologies. Figure 1 illustrates the research process aligned with the predefined research questions, including research design, sampling, data collection, and analysis.

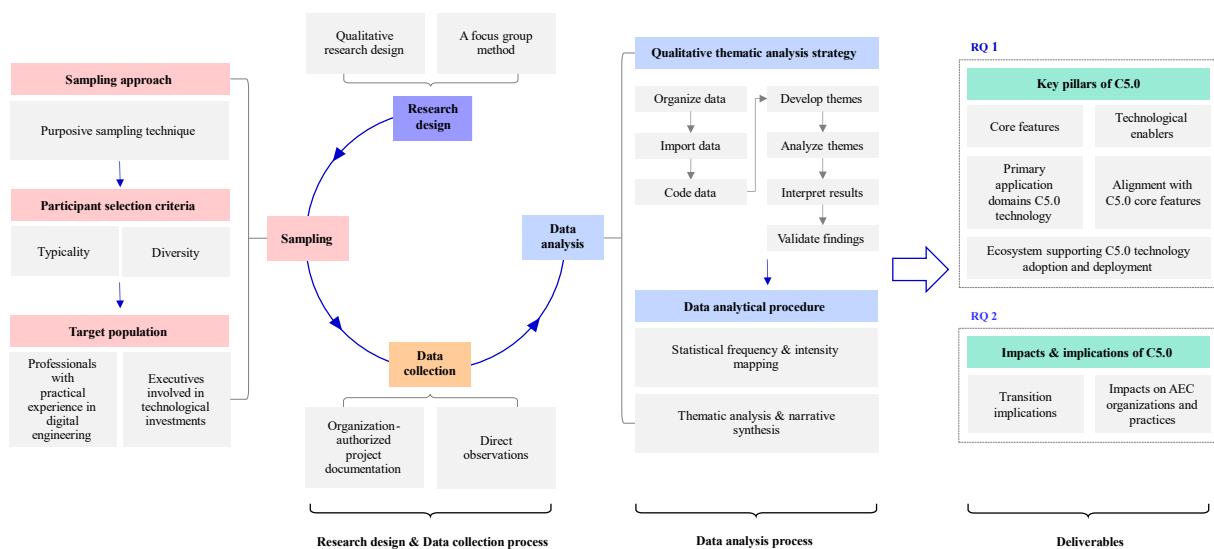


Figure 1. Overview of the research process

A purposive sampling strategy was employed to select participants, targeting professionals actively engaged in emerging technology implementation within AEC projects or organizational roles and firm leaders responsible for technological investments. This sampling approach ensured the inclusion of individuals capable of providing both operational and strategic insights. To achieve a representative and diverse sample, participants were further screened based on their expertise with advanced technologies and their roles across various organizational contexts, such as firm types and project domains.

The study drew on data from a broader research project examining the transition from C4.0 to C5.0. In

the earlier phase of the project, background information on participants, such as job roles, professional experience, company profiles, and technology adoption practices, was gathered. For the current phase, the contact database was screened to identify eligible candidates. Potential participants were then emailed with an explanation of the study's objectives and invitations to participate. Based on positive responses, 17 participants were selected for focus group observations. Figure 2 illustrates the demographic composition of the finalized focus group participants.

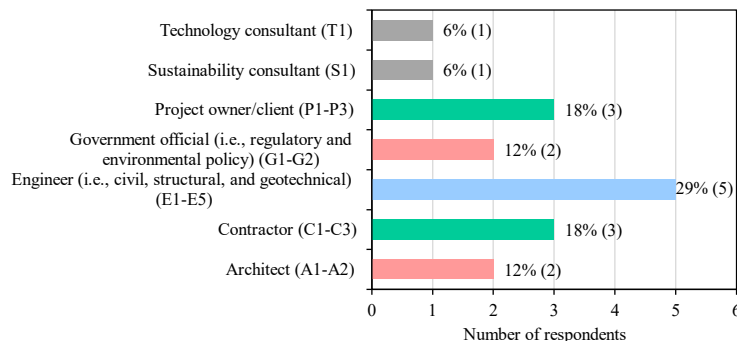


Figure 2. Demographic profile of the participants

The focus group session covered two main themes: (1) the key pillars of C5.0 and (2) its implications for AEC organizations and practices. Regarding the first theme, participants shared insights on three core areas: (i) their understanding of C5.0's core features, (ii) the enabling technologies they identified as critical for advancing the C5.0 paradigm, and (iii) the success factors they identified as critical for promoting the transition from C4.0 to C5.0. For the second theme, participants provided feedback on the anticipated benefits of transitioning to C5.0 for their organizations and the potential impacts of adopting C5.0 technologies on their projects.

A mixed-methods approach was employed for data analysis, integrating both quantitative and qualitative techniques. The quantitative analysis focused on mapping the frequency and intensity of technologies within the C5.0 framework, examining their implementation across application areas, integration with C4.0 technologies, and deployment outcomes aligned with C5.0 core objectives. Subsequently, qualitative thematic analysis and narrative synthesis were conducted to explore and describe the applications and impacts of these technologies within C5.0 environments.

4. FINDINGS AND DISCUSSION

4.1 Conceptual Foundations of Construction 5.0

(1) Core Features

Figure 3 presents the key pillars of C5.0, synthesizing insights from the 17 participants. Building on the foundation of C4.0, C5.0 introduces three core features: human-centricity, sustainability-driven practices, and collaborative intelligence. These features represent a paradigm shift in addressing long-standing challenges in the AEC industry. Human-centricity focuses on improving worker well-being and safety by integrating technologies such as wearable devices, cobots, and HCPS (Cisneros-Gonzalez et al., 2024; Yitmen et al., 2023). Participant P3 remarked, *“One of the significant benefits of wearable devices lies in their ability to adapt to workers’ needs, providing real-time data to prevent accidents and ensure well-being.”* Similarly, Participant E2 highlighted the transformative potential of cobots, stating, *“Collaborative robots are game-changers for reducing physical strain in repetitive tasks. They don’t replace people. They augment people’s capabilities, especially in high-risk operations.”* Extended reality (XR) technologies further support this human-centric feature by offering immersive training simulations and virtual site inspections. As respondent C1 observed, *“VR or AR gives workers hands-on experiences without taking any risk. It bridges knowledge gaps and ensures safety compliance.”*

The sustainability-driven aspect of C5.0 prioritizes minimizing the environmental impact of construction processes while optimizing resource use. Digital twins, IoT-enabled smart systems, and 3D/4D printing promote energy efficiency, waste reduction, and circular construction practices (Ghobakhloo et al., 2022; Jiménez Rios et al., 2024). Participant E5 noted, *“Digital twins could give us a clearer view of where we can cut wastes and save resources.”* Participant P2 emphasized the role of advanced machinery: *“It’s not just labor savings with autonomous equipment. It’s also fewer emissions and less wasted material.”*

Collaborative intelligence emphasizes the integration of advanced technologies with human expertise, enabling interconnected systems that enhance decision-making and coordination. For instance, AI/ML, cognitive computing, and advanced communication networks (e.g., 5G and 6G) support real-time data sharing, decision-making, and dynamic planning (Tunji-Olayeni et al., 2024; Musarat et al., 2023). Participant E4 highlighted the transformative nature of this integration: *“The integration of AI and 5G could make the collaboration between*

human and machine smoother. Workers can control autonomous machinery remotely, safely, and precisely without any delay.” Participant G2 shared, “Collaboration across teams will become easier with IoT and communication networks. Everyone is aligned because the data is instant and accessible.”

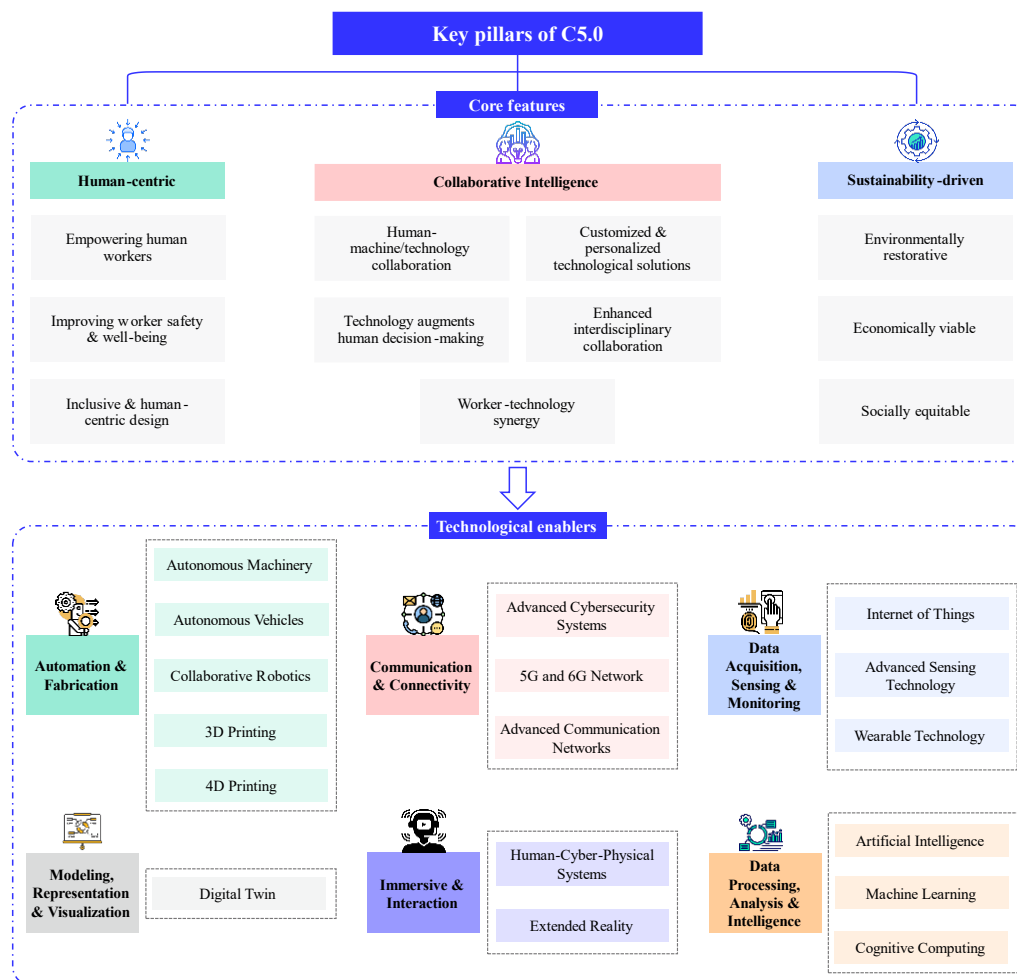


Figure 3. Key pillars underpinning C5.0

(2) Technological Enablers

The transition to C5.0 introduces advanced technologies to redefine the AEC industry through human-centric design, sustainability-driven practices, and collaborative intelligence. Figure 3 presents some main technological enablers and the associated application potential. Figure 4 summarizes the detailed application areas of these technologies, ranked by the criticality based on focus group feedback. Specifically, the use of AI and ML enhances predictive maintenance, safety monitoring, and automated defect detection and quality control, ensuring worker safety and promoting collaborative intelligence of humans and machines. The findings align with other regional surveys and global review studies, such as Almusaed et al. (2024), Bello et al. (2024), and Rane (2023), which emphasize the role of AI in facilitating informed decision-making throughout the project lifecycle. Similarly, digital twins enable virtual training, real-time lifecycle optimization, and scenario simulation, contributing to safety management, waste reduction, and energy efficiency. This is consistent with the review results of international studies by Biswas et al. (2024), highlighting the integration of digital twins, BIM, and IoT to provide precise data-driven insights for advancing sustainable project management processes. Cobots are particularly transformative in reducing physical strain on workers, enhancing operational safety and efficiency, and supporting construction process optimization. These outcomes are supported by Marinelli (2023), which emphasizes cobots' ability to promote worker ergonomics and productivity through a bibliometric analysis of international literature. Among the three core features of C5.0, human-centric emerges as the most prominent, with most identified technologies aligning closely with this feature. This supports the findings of Jiménez Rios et al. (2024), which identified human-centricity as the most distinguishing feature of I5.0 compared to I4.0 through a systematic literature review. Sustainability remains a core theme in C5.0, with technologies like 3D/4D printing and autonomous machinery enabling waste minimization and emissions reductions. Tunji-Olayeni et al. (2024)

similarly identified these technologies as pivotal in advancing sustainable construction practices through reviewing international literature. Collaborative intelligence, another core feature of C5.0, is enabled by cobots, IoT and advanced sensors, advanced communication networks (5G/6G), cognitive computing, and advanced cybersecurity. As addressed by Musarat et al. (2023) through the review of international literature, collaborative intelligence encourages interconnected ecosystems, bridging human creativity with machine intelligence to meet the evolving industry demands. These insights invite further discussion on overcoming adoption barriers, developing ethical frameworks, and maximizing C5.0's potential to transform the AEC sector.

















































Technology		Main Application Area in C5.0	Alignment with C5.0 Core Features		
Application Areas of C5.0 Technologies	 AI and ML	Predictive maintenance and risk mitigation			
		Automated defect detection and quality control			
	 Digital Twin s	Virtual training			
		Project lifecycle management			
		Environmental impact simulation			
	 Cobots	Perform repetitive tasks to reduce worker strain			
		High-risk task automation			
		Precise assembly processes			
		Interactive workflows for process optimization			
	 HCPS	Safety monitoring and risk prediction			
		Resource management and decision -making			
	 IoT and Advanced Sensors	Real-time worker safety monitoring			
		Smart building energy optimization			
		Seamless communication among connected systems			
		Reliable communication for remote operations			
	 5G, 6G, and Advanced Communication Networks	Energy-efficient communication			
		Support real-time human-machine collaboration			
	 Wearable Technology	Health condition monitoring			
		Worker movements and task progress tracking			
	 Extended Reality	Immersive training simulations			
		Virtual site inspections and remote monitoring			
	 Autonomous vehicles, cranes, and heavy machinery	Immersive design and visualization			
		Reduce human involvement in high -risk tasks			
		Operation optimization			
	 3D/4D Printing	Seamless integration for real-time site adjustments			
		Personalized building designs for human comfort			
	 Cognitive Computing	Precise manufacturing			
		Rapid prototyping and real-time design adjustments			
	 Advanced Cybersecurity Solutions	Dynamic and intelligent decision support			
		Protect sensitive worker and operational data			
		Detect and prevent vulnerabilities in collaborative CPS			

Figure 4. Summary of C5.0 technologies and their applications as reported by industry representatives

4.2 Impacts and Implications of C5.0 for AEC Organizations and Practices

As the 17 industry stakeholders highlighted, the transition from C4.0 to C5.0 would substantially improve the AEC sector in productivity, sustainability, human-technology collaboration, and worker safety and well-being. The transition implications are presented in Table 2. 86% of respondents noted that C5.0 improves workflows, optimizes resource allocation, and enhances operational efficiency, consistent with the findings of Hermann et al. (2016) on the benefits of digital transformation in construction. Further, 71% highlighted enhanced decision-making capabilities, enabling agile responses and real-time adjustments, echoing the content-centric synthesis results of literature by Ghobakhloo et al. (2022), which addressed the importance of data-driven decision-making in I5.0. However, only 36% reported increased customization in construction solutions that adapt to client needs and environmental conditions, reflecting findings from Müller (2020) on the complexities of integrating human-centric solutions in C5.0. Enhanced sustainability emerged as one of the most significant implications of the transition, with all respondents acknowledging C5.0's role in improving compliance with environmental regulations and meeting client demands for sustainable practices, aligning with the emphasis on sustainability in the systematic literature review by Davila-Gonzalez & Martin (2024). Furthermore, 93% noted reduced costs related to material usage, waste management, and energy consumption, indicating environmental and financial

benefits. Additionally, 43% addressed increased attractiveness to clients and investors in prioritizing green and socially responsible projects, reflecting the growing demand for sustainable construction solutions.

Regarding human-technology collaboration, 79% of respondents observed that delegating repetitive and hazardous tasks to machines improves workforce safety and productivity. This reflects the findings of Yitmen et al. (2023), based on surveys conducted in Europe, North America, and the Middle East, highlighting the potential of robotics and automation to enhance efficiency and minimize human involvement in dangerous tasks. 57% observed a shift in worker roles toward higher-level problem-solving, suggesting that workers are moving toward more strategic roles that combine data-driven insights with human expertise. Regarding worker safety and well-being, 64% recognized that transitioning to C5.0 could reduce the risk of injuries, minimize workplace hazards, and lead to a healthier and more motivated workforce. This suggests that adopting advanced safety technologies and delegating dangerous tasks to machines are key drivers in improving worker safety and well-being. These findings align with Marinelli (2023), which emphasizes the role of collaborative technology in improving workplace safety by reducing human exposure to hazardous conditions.

Table 2. Summary of the impacts and implications of the C5.0 transition highlighted by industry representatives

Impacts and Implications	Description	Respondent (N=17)
Improved Productivity and Innovation	Streamlined workflows, optimized resource allocation, and increased operational efficiency.	15 (86%)
	Enhanced decision-making capabilities that allow agile responses and real-time adjustments.	12 (71%)
	Greater customization in construction solutions that adapt to client needs, environmental conditions, and project specifications.	6 (36%)
Enhanced Sustainability	Improved compliance with environmental regulations and greater capacity to meet client demand for sustainable practices.	17 (100%)
	Reduced costs associated with material usage, waste management, and energy consumption.	16 (93%)
	Increased attractiveness to clients and investors prioritizing green and socially responsible projects.	7 (43%)
Improved Human-Technology Collaboration	Enhanced workforce safety, reduced fatigue, and improved productivity by assigning repetitive or hazardous tasks to machines.	13 (79%)
	Shift in worker roles toward higher-level problem-solving that combines data-driven insights with human expertise.	10 (57%)
	Creation of opportunities for upskilling and developing a more competent workforce.	4 (21%)
Enhanced Worker Safety and Well-being	Reduced risk of injuries, minimized workplace hazards, and a healthier and more motivated workforce.	11 (64%)
Competitive Advantage	Opportunities to position companies as industry leaders capable of managing complex, large-scale projects that require innovative, sustainable solutions.	4 (21%)

5. CONCLUSIONS

While C5.0 is recognized as a critical direction for the AEC sector, gaps remain in understanding its conceptual foundations and practical implications. Specifically, existing research in literature primarily focuses on conceptual frameworks or literature reviews, with limited empirical evidence from industry practices. This study addresses these gaps by empirically examining the foundations and implications of C5.0, incorporating insights from 17 industry practitioners through a focus group approach. It investigates C5.0's key pillars, core features, and technological enablers, analyzes the alignment of technology implementation with C5.0 goals, and explores the paradigm's implications for AEC organizations and project management practices.

The results identify three core features of C5.0: human-centricity, sustainability-driven practices, and collaborative intelligence. A total of 17 emerging digital technologies are highlighted as key enablers of the transition to C5.0, with 12 technologies deemed critical to achieving its objectives across 31 application domains. Among these, (1) AI/ML, (2) digital twins, and (3) cobots were identified as the most critical C5.0 technologies. The study also identifies five major implications for AEC organizations transitioning to C5.0: improved productivity and innovation, enhanced sustainability, strengthened human-technology collaboration, improved worker safety and well-being, and competitive advantage. Under the "Enhanced Sustainability" category, all industry representatives emphasized improved compliance with environmental regulations and a greater capacity to meet client demand for sustainable practices as key outcomes.

This research contributes theoretically and practically to understanding C5.0's pillars, technologies, and implications. Practically, identifying critical technologies enables AEC organizations to prioritize investments in

tools essential for achieving C5.0 objectives. By linking technologies to specific C5.0 features, the study provides practitioners with a clear roadmap for integrating these tools to meet human-centric and sustainability goals. Theoretically, the study advances the understanding of C5.0 by defining its key pillars, core features, and technological enablers, bridging gaps in the existing literature. It also provides a foundation for exploring the interaction between technological advancements and human-centered principles, particularly by aligning technologies such as AI/ML, digital twins, and cobots with C5.0 objectives. Although the study is based on a focus group discussion among regional participants, the findings were compared with global references, aligning with global benefits like enhanced sustainability and human-centricity reported in European and North American contexts. Future research should focus on expanding empirical studies to explore C5.0's impact on project management and organizational strategies. Specific directions include:

1. Conducting a cost-benefit evaluation of C5.0 technologies and investigating how C5.0 can reshape project management methodologies, including decision-making, stakeholder collaboration, and risk management.
2. Developing strategies to integrate C5.0-driven agility into project workflows, addressing challenges in resource allocation and timeline optimization.

REFERENCES

- Abioye, S. O., Oyedele, L. O., Akanbi, L., Ajayi, A., Davila Delgado, J. M., Bilal, M., Akinade, O. O., and Ahmed, A. (2021). Artificial intelligence in the construction industry: a review of present status, opportunities and future challenges. *Journal of Building Engineering*, 44, 103299.
- Almusaed, A., Yitmen, I., Almssad, A., and Myhren, J. A. (2024). Construction 5.0 and sustainable neuro-responsive habitats: integrating the brain-computer interface and building information modeling in smart residential spaces. *Sustainability (Switzerland)*, 16(21).
- Bello, J. O., Stephen, S., Adetoro, P., and Mogaji, I. J. (2024). Supply chain resilience in the construction industry: a bibliometric review on operations management practices from Industry 4.0 to Industry 5.0. *Benchmarking: An International Journal*, ahead-of-print.
- Biswas, H. K., Sim, T. Y., and Lau, S. L. (2024). Impact of building information modelling and advanced technologies in the AEC industry: a contemporary review and future directions. *Journal of Building Engineering*, 82, 108165.
- Brozovsky, J., Labonnote, N., and Vigren, O. (2024). Digital technologies in architecture, engineering, and construction. *Automation in Construction*, 158, 105212.
- Chen, X., Chang-Richards, A. Y., Yiu, T. W., Ling, F. Y. Y., Pelosi, A., and Yang, N. (2023). A multivariate regression analysis of barriers to digital technologies adoption in the construction industry. *Engineering, Construction and Architectural Management*, ahead-of-print.
- Chen, X., Liu, F., Ghaffari-anhoseini, A., GhaffarianHoseini, A., and Guo, B. (2024a). From Construction 4.0 to Construction 5.0: principles and enabling technologies. *International Conference of Smart and Sustainable Built Environment (SASBE 2024)*, Auckland, New Zealand.
- Chen, X., Chang-Richards, A. Y., Ling, F. Y. Y., Yiu, T. W., Pelosi, A., & Yang, N. (2024b). Digital technology-enabled AEC project management: practical use cases, deployment patterns and emerging trends. *Engineering, Construction and Architectural Management*, ahead-of-print.
- Chen, X., Chang-Richards, A., Ling, F. Y. Y., Yiu, K. T. W., Pelosi, A., & Yang, N. (2024c). Effects of digital readiness on digital competence of AEC companies: a dual-stage PLS-SEM-ANN analysis. *Building Research and Information*, 52(8), 905–922.
- Cisneros-Gonzalez, J. J., Rasool, A., and Ahmad, R. (2024). Digital technologies and robotics in mass-timber manufacturing: a systematic literature review on construction 4.0/5.0. *Construction Robotics*, 8, 29.
- Davila-Gonzalez, S., and Martin, S. (2024). Human digital twin in Industry 5.0: a holistic approach to worker safety and well-being through advanced AI and emotional analytics. *Sensors*, 24, 655.
- Ghobakhloo, M., Iranmanesh, M., Mubarak, M. F., Mubarik, M., Rejeb, A., and Nilashi, M. (2022). Identifying industry 5.0 contributions to sustainable development: a strategy roadmap for delivering sustainability values. *Sustainable Production and Consumption*, 33, 716–737.
- Hadi, A., Cheung, F., Adjei, S., and Dulaimi, A. (2023). Evaluation of lean off-site construction literature through the lens of Industry 4.0 and 5.0. *Journal of Construction Engineering and Management*, 149(12), 03123007.
- Hermann, M., Pentek, T., and Otto, B. (2016). Design principles for industrie 4.0 scenarios. *2016 49th Hawaii International Conference on System Sciences (HICSS)*, 3928–3937.
- Hu, C., Liu, P., Yang, H., Yin, S., and Ullah, K. (2023). A novel evolution model to investigate the collaborative innovation mechanism of green intelligent building materials enterprises for construction 5.0. *AIMS Mathematics*, 8(4), 8117–8143.
- Ikudayisi, A. E., Chan, A. P. C., Darko, A., and Adedeji, Y. M. D. (2023). Integrated practices in the

- architecture, engineering, and construction industry: current scope and pathway towards Industry 5.0. *Journal of Building Engineering*, 73, 106788.
- Ivanov, D. (2023). The Industry 5.0 framework: viability-based integration of the resilience, sustainability, and human-centricity perspectives. *International Journal of Production Research*, 61(5), 1683–1695.
- Jiménez Rios, A., L. Petrou, M., Ramirez, R., Plevris, V., and Nogal, M. (2024). Industry 5.0, towards an enhanced built cultural heritage conservation practice. *Journal of Building Engineering*, 96, 110542.
- Maddikunta, P. K. R., Pham, Q. V., B, P., Deepa, N., Dev, K., Gadekallu, T. R., Ruby, R., and Liyanage, M. (2022). Industry 5.0: a survey on enabling technologies and potential applications. *Journal of Industrial Information Integration*, 26, 100257.
- Marinelli, M. (2023). From Industry 4.0 to Construction 5.0: exploring the path towards human–robot collaboration in construction. *Systems*, 11, 152.
- McKinsey & Company. (2020). *The Next Normal in Construction*. Brussels, Belgium.
- Müller, J. (2020). *Enabling Technologies for Industry 5.0*. European Commission, 8-10.
- Musarat, M. A., Irfan, M., Alaloul, W. S., Maqsoom, A., and Ghufraan, M. (2023). A review on the way forward in construction through Industrial Revolution 5.0. *Sustainability (Switzerland)*, 15, 13862.
- Nahavandi, S. (2019). Industry 5.0—a human-centric solution. *Sustainability*, 11(16), 4371.
- O.Nyumba, T., Wilson, K., Derrick, C. J., and Mukherjee, N. (2018). The use of focus group discussion methodology: insights from two decades of application in conservation. *Methods in Ecology and Evolution*, 9, 20–32.
- Ohueri, C. C., Masrom, M. A. N., and Noguchi, M. (2024). Human-robot collaboration for building deconstruction in the context of construction 5.0. *Automation in Construction*, 167, 105723.
- Rane, N. (2023). ChatGPT and similar generative artificial intelligence (AI) for building and construction industry: contribution, opportunities and challenges of large language models for Industry 4.0, Industry 5.0, and Society 5.0. *SSRN Electronic Journal*.
- Saradara, S. M., Khalfan, M. M. A., Jaya, S. V., Swarnakar, V., Rauf, A., and El Fadel, M. (2024). Advancing building construction: a novel conceptual framework integrating circularity with modified lean project delivery systems. *Developments in the Built Environment*, 20, 100531.
- Sawhney, A., Riley, M., and Irizarry, J. (2020). *Construction 4.0: An Innovation Platform for the Built Environment* (A. Sawhney, M. Riley, and J. Irizarry (eds.); Vol. 10). Routledge.
- Tunji-Olayeni, P., Aigbavboa, C., Oke, A., and Chukwu, N. (2024). Research trends in Industry 5.0 and its application in the construction industry. *Technological Sustainability*, 3(1), 1–23.
- Ürge-Vorsatz, D., Khosla, R., Bernhardt, R., Chan, Y. C., Vérez, D., Hu, S., and Cabeza, L. F. (2020). Advances toward a net-zero global building sector. *Annual Review of Environment and Resources*, 45, 227–269.
- Wang, J. X., Burke, H., and Zhang, A. (2022). Overcoming barriers to circular product design. *International Journal of Production Economics*, 243, 108346.
- World Economic Forum. (2023). *Annual Report 2022/2023*. Retrieved from: <https://doi.org/10.1002/vetr.3399>
- Xu, X., Lu, Y., Vogel-Heuser, B., and Wang, L. (2021). Industry 4.0 and Industry 5.0—inception, conception and perception. *Journal of Manufacturing Systems*, 61, 530–535.
- Yitmen, I., Almusaed, A., and Alizadehsalehi, S. (2023). Investigating the causal relationships among enablers of the construction 5.0 paradigm: integration of operator 5.0 and society 5.0 with human-centricity, sustainability, and resilience. *Sustainability (Switzerland)*, 15, 9105.
- Zizic, M. C., Mladineo, M., Gjeldum, N., and Celent, L. (2022). From Industry 4.0 towards Industry 5.0: a review and analysis of paradigm shift for the people, organization and technology. *Energies*, 15, 5221.

RESEARCH ON A METHODOLOGY OF AI-DRIVEN SUBSTATION LAYOUT DESIGN BASED ON LARGE LANGUAGE MODEL

Jianyong SHI¹, Zeyu PAN², Longfei PAN³, and Junpeng Hu⁴

1)Ph.D., Assoc. Prof., Department of Civil Engineering, Shanghai Jiao Tong University, Shanghai, China. Email: shijy@sjtu.edu.cn

2)Ph.D. Candidate, Department of Civil Engineering, Shanghai Jiao Tong University, Shanghai, China. Email: panzeyu@sjtu.edu.cn

3) M. Eng., Sichuan Research Institute, Shanghai Jiao Tong University, Shanghai, China. Email:

4) M. Eng., Department of Civil Engineering, Shanghai Jiao Tong University, Shanghai, China. Email: hu-j.p@sjtu.edu.cn

Abstract: In the design stage of substation projects, there is an enduring task for designing an optimal layout for equipment under various constraints from functional, geometric, and economic aspects. In traditional practices, professionals need to recursively adjust the positions of involved objects in the specialized working spaces to meet the requirements of different projects and to comply with design codes. It is highly dependent on professional skills and understanding of regulatory documents. To streamline this process, we propose an AI-driven substation layout design approach using large language model (LLM). This approach exploits the capacity of large language model by converting the task of generating the full layout plan into generating sequences of the positions of involved equipment. We have finetuned two models based on a base model Llama3.1-7B with two auto-generated datasets under professional guidance. One dataset consists of the input requirements and output layout scheme, while the other is augmented through chain-of-thought (CoT) to elicit the underlying language model to retrieve optimal capacity with the consideration of specific design constraints. To implement the output scheme, an automated procedure is further developed to translate the scheme into the corresponding layout plan and information models.

Keywords: AI-driven Design, Substation, LLM, CoT, Layout Design

1. INTRODUCTION

The substation equipment layout task is a comprehensive and complex project. The specific contents involved include but are not limited to determining the type and quantity of equipment required for the substation and their precise placement. At the same time, the design scheme needs to be optimized to achieve the goals of maximizing layout accuracy, optimizing costs, and maximizing safety and reliability. The task requirements encompass ensuring the stable operation of the power system and adhering to design specifications, while constraints involve a myriad of factors, including technical standards and legal regulations. To complete this task, designers must not only have solid professional knowledge and rich engineering experience, but also have a deep understanding and mastery of the specific requirements of the task, various constraints in the design process, and related engineering prior knowledge to ensure the rationality and feasibility of the layout plan.

Although a variety of design software has been developed in the field of substation equipment to improve design efficiency, current design technology still relies mainly on manual operation. This results in a lot of repetitive work when dealing with many standard design tasks, which not only consumes a lot of money, but also leads to errors and deviations caused by human factors, as well as some unavoidable problems. The high degree of manual intervention in the design process has affected efficiency and accuracy to a certain extent.

In response to the above problems, this study proposed an artificial intelligence-driven substation layout design method using a large language model. This study introduced artificial intelligence technology and focused on the semantic expression of substation layout drawings, model pre-training and fine-tuning, semantic conversion and other technical aspects. The use of artificial intelligence technology to generate drawings improves the efficiency of substation design; the use of procedural and modular ideas to generate design drawings reduces human intervention and improves the accuracy of the design. For professional designers, it significantly reduces the workload and further improves the efficiency of substation layout design, which is of great significance to the development of power grid engineering.

Studies on automatic layout generation have appeared several times in literature (Agarwala et al., 2011). Recent approaches to layout generation consider both unconditional generation (Arroyo et al., 2021) and conditional generation in various setups, such as conditional inputs of category or size (Gupta et al., 2021, Jiang et al., 2022), relational constraints (Jiang et al., 2022, Kikuchi et al., 2021), element completion (Gupta et al., 2021), and refinement (Jiang et al., 2022). Some attempt at solving multiple tasks in a single model (Kikuchi et al., 2021, Kong et al., 2022).

As an important part of generative models, language models have achieved remarkable achievements in the current development. The advent of the Transformer architecture, grounded in deep learning and large-scale pre-trained models, has revolutionized natural language understanding and generation by language models. These models play a key role in natural language processing tasks such as text classification, machine translation, and

question-answering systems. From automated document writing to the initial conception of design schemes, language models have demonstrated their powerful text generation capabilities. Their application in the field of natural language processing has penetrated into multiple levels of scheme design. In scheme design tasks, these models are widely used to generate design instructions, technical specifications, and even directly participate in the preliminary layout of design schemes. However, this paper considers exploring the generation of substation equipment layout schemes from another perspective, that is, transforming the traditional task goal from directly generating a plane layout scheme diagram to generating a form of equipment metadata sequence. The core idea of this method is to use the semantic comprehension and inference capabilities of the language model to generate a series of metadata sequences representing the substation equipment layout by encoding equipment attributes and layout rules, and initially have the ability to control the design scheme generation process according to contextual information such as design requirements and design constraints. This provides designers with a new design idea and tool to improve the flexibility and automation level of the design while ensuring the rationality of the layout.

This paper uses a LLM to perform AI-driven substation layout design. By collecting substation layout drawings, key information is extracted from the drawings, a big language model fine-tuning dataset is constructed, and the big language model is fine-tuned to output semantic design drawing information. Finally, the semantic information is converted into layout drawings through a semantic conversion tool to complete the intelligent layout of the substation.

2. METHOD

Large language model (LLM) is a natural language processing model based on deep learning technology with billions to hundreds of billions of parameters. It is pre-trained with massive data and can understand and generate natural language text. It has versatility, contextual understanding and natural fluency. It can perform multiple language tasks including text generation, translation, question answering, etc. It is an important breakthrough in the current field of natural language processing. The Meta Llama 3.1 multilingual large language model (LLMs) collection is a collection of pre-trained and instruction-adjusted generative models with three sizes (text input/text output): 8B, 70B and 405B. The instruction-adjusted plain text models (8B, 70B and 405B) of Llama 3.1 are optimized for multilingual conversation use cases and outperform many existing open source and closed chat models in common industry benchmarks (meta-llama). Since 70B and 405B cannot run on consumer-grade graphics cards, this experiment uses llama3.1-8B-instruct as the base model. It has about 8 billion parameters and can complete tasks well according to user instructions. It can also run on consumer-grade graphics cards and can popularize the method in this article on a large scale.

Fine-tuning (Zhao et al., 2023) is a machine learning technique based on a pre-trained model. It aims to optimize the performance of the model and adapt it to new application scenarios by training it with a small amount of task-specific datasets. This process involves adjusting some parameters of the model, usually retaining most of the learned general features and modifying only the top layer or output layer to match the requirements of the specific task. Fine-tuning can effectively utilize the knowledge of the pre-trained model, reduce training costs, and achieve better performance on the target task. Common fine-tuning methods include full model fine-tuning (Lv et al., 2023), frozen bottom layer fine-tuning (Zheng et al., 2024), prompt tuning, and low-rank adaptation (LoRA) (Hu et al., 2021). Full model fine-tuning allows the entire model to be adjusted. Its advantage is that it can make full use of task-specific data, but its disadvantage is that it consumes a lot of computing resources and is prone to overfitting. Frozen bottom layer fine-tuning only adjusts the top-level parameters of the model, saving resources and reducing the risk of overfitting, but it may not fully capture task-specific features. Prompt fine-tuning affects the model output by modifying the input without updating the model parameters. It is computationally efficient, but its applicability may be limited. LoRA simulates parameter updates by introducing low-rank matrix decomposition, which has the advantages of high computational efficiency and low resource consumption, while keeping most of the parameters of the pre-trained model unchanged, effectively avoiding catastrophic forgetting. Comprehensive analysis shows that the LoRA fine-tuning method can efficiently adapt to new tasks while maintaining the advantages of pre-training. It is a resource-efficient and robust fine-tuning strategy.

This paper is dedicated to exploring the application of artificial intelligence in substation layout design. By utilizing the powerful capabilities of the big language model, a new substation layout design method is realized. We first systematically organize the existing substation design drawings, and use the prompt word technology of the big model to construct a design scheme dataset rich in thinking logic chain in combination with specific design conditions and specification constraints. On this basis, we use the LoRA fine-tuning technology to customize the big model to obtain LoRA weights that are more suitable for substation layout design. Then, we deploy the fine-tuned LoRA together with the big model as a local service, and develop a plug-in compatible with Revit software to provide users with an intuitive interactive interface. Users input key parameter information such as the number of equipment, and the plug-in encapsulates these parameters and constructs them into prompt words, and sends requests to the local big model service. After the server parses the returned design scheme, it presents it to the user intuitively, thus realizing the automation and intelligence of substation layout design.

There are two datasets, one of which includes input requirements and output layout plans. The input

requirements include the range, device name, and size of the required layout. The output layout plan includes the device name, size, and layout information (the distance of the device from the x-axis and y-axis, and the orientation). These two fields can be regarded as the input and output of the model during fine-tuning. System prompts are also required during fine-tuning. Construct the system prompts "You are now an experienced architectural designer. Use your design experience and your design talents to think step by step to get a design plan. Design plan 1 is expressed in standard markdown json format." The first fine-tuning dataset is obtained as shown in Figure 1.

```
"instruction": "Please provide a detailed design, use your design experience, and create a design that answers the user's question.",


```

Figure 1. Basic fine-tuning dataset

Another dataset is based on the previous dataset and integrates design specification constraints and chain of thought (COT) technology (Wei et al., 2022), aiming to improve the design accuracy of the model and reduce model hallucination phenomena.

Based on the above two data sets, this paper designed four experimental schemes. The first one is to use no fine-tuning + One-Shot case + design requirements + adding design constraints in Prompt. The second one is also in the form of no fine-tuning + Few Shot case + design requirements + design constraints. The third one is a fine-tuning model + design requirements for the combination of design requirements and design solutions. The fourth one is a fine-tuning model + design requirements for the combination of cot + sft (design requirements + intermediate design process + design solution). The experimental results show that the layout solution obtained by the cot + sft solution is the best. The specific technical method is shown in Figure 2.

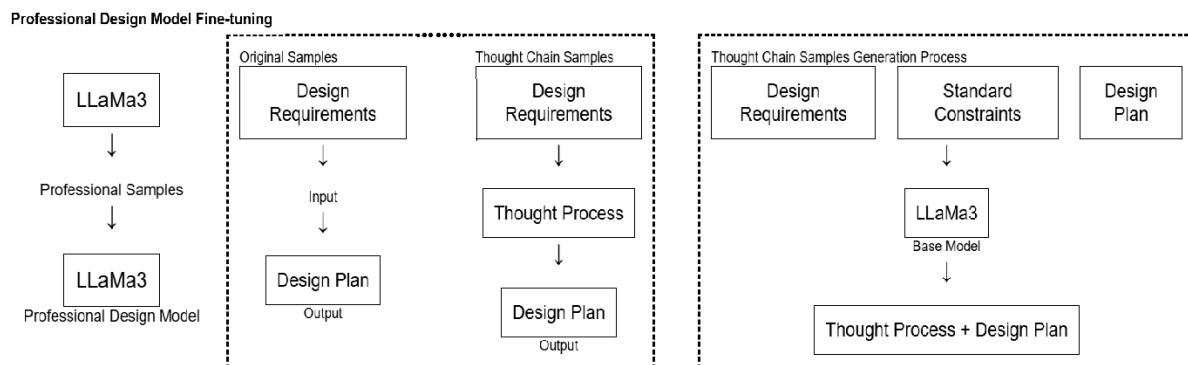


Figure 2. Schematic diagram of the cot dataset construction process

The standardized design scheme for substation layout refers to providing a set of preset layout schemes for substation equipment based on standards and specifications, including different types such as modular, compact and partitioned designs. These schemes convert the design intent into a serialized representation of the equipment layout, that is, convert the type, size, location and orientation of the equipment into a data token sequence using a large language model, thereby realizing the conversion from the traditional image generation problem to the determination of the location and orientation data of the equipment bounding box in the plane. This process not only simplifies the design complexity, but also makes it possible for automated layout.

Since the current large language model has a good understanding of data in JSON format, we consider

converting the substation into a semantic string in JSON format. Converting the substation layout drawing into a JSON file is a complex but important process, which can achieve structured storage and convenient interaction of data. Figure 3 shows the JSON file content of the drawing information interpreted by the fine-tuned large model.

```
"Design Plan Description": "Please arrange the following devices within a 9.8m x 17.1m rectangular area:
12.8m x 1.5m device 1,
11.9m x 0.8m device 2,
1.6m x 0.6m device 3,
7.2m x 0.6m device 4, two 2.4m x 0.8m devices connected to No. 2,
two 2.4m x 0.8m devices connected vertically to the above-mentioned devices,
two 0.9m x 2.0m devices connected to the above-mentioned devices, and two 2m x 2m platforms!
12.8m, 0.8m and 0.9m devices are placed horizontally,
1m x 2.8m devices are placed vertically.
Layout principle: people and devices are arranged symmetrically.
Devices face forward, people face forward.
Two represent the negative direction, three represent the positive direction."

"Device Layout":
  • "Name": "Basic platform, offset -200mm_200"
    ◦ "Length": "17.1"
    ◦ "Width": "9.8"
  • "Name": "Device 1"
    ◦ "Length": "12.8"
    ◦ "Width": "1.5"
  • "Name": "Device 2"
    ◦ "Length": "7.3"
    ◦ "Width": "3.3"
```

Figure 3. Semantic information of drawings in json format

We use the big language model to output the thought chain process, that is, given the design conditions, design constraints, and design results, let the big language model output the intermediate process of the design. The design constraints sorted out based on the existing data are as follows:

The width (clearance) of various channels in the high-voltage distribution room should not be less than the following:

When the switch cabinet is arranged in a single row, in the channel classification, the clear width of the maintenance channel behind the cabinet should not be less than 0.8m; in the operation channel, if the switch cabinet is fixed, the clear width of the operation channel should not be less than 1.5m, and if the switch cabinet is a push type, the clear width of the operation channel should not be less than the length of a single vehicle + 1.2m; and the clear width of the channel leading to the explosion-proof interval should not be less than 1.2m.

When the switch cabinet is arranged with switch cabinets on both sides and face to face, in the channel classification, the clear width of the maintenance channel behind the cabinet should not be less than 0.8m; in the operation channel, if the switch cabinet is fixed, the clear width of the operation channel should not be less than 2m, and if the switch cabinet is a push type, the clear width of the operation channel should not be less than the length of two vehicles + 0.9m; and the clear width of the channel leading to the explosion-proof interval should not be less than 1.2m.

When the switch cabinet is arranged with switch cabinets on both sides and back to back, in the channel classification, the net width of the maintenance channel behind the cabinet should not be less than 1m; in the operation channel, if the switch cabinet is fixed, the net width of the operation channel should not be less than 1.5m, and if the switch cabinet is hand-push, the net width of the operation channel should not be less than the length of the single vehicle + 1.2m; and the net width of the channel leading to the explosion-proof interval should not be less than 1.2m.

When the length of the high-voltage distribution room exceeds 7m, two doors should be opened and arranged at both ends. The height of the handling door of the GC-1A (F) type high-voltage switch cabinet is 2.5~2.8m and the width is 1.5m.

When the power supply enters from the back of the cabinet and an isolating switch and its manual operating mechanism need to be installed on the wall behind the cabinet, the net width of the channel behind the cabinet should not be less than 1.50m; when the protection level of the back of the cabinet is IP2X, it can be reduced to 1.30m.

The built prompts are as follows:

Please combine the design requirements: {requirement} and the design constraints: {limit}. Please briefly quote the design constraints, give the design steps step by step, and reason to get the final design result: {result}. Please note: only output the intermediate thinking process, do not add the final design result.

The above prompt words do not allow the big language model to add the final design results because the output of the big language model will have hallucinations. In order to avoid this phenomenon, the final design

results can be concatenated with strings after generation.

After using the local big model to build a local service, input the data set and prompt words into the big model to obtain a thinking chain data set with design logic and design constraints.

Based on the fine-tuning dataset and large model, the method for LoRA fine-tuning is as follows: First, replace the key layers in the large model with LoRA layers, which contain trainable low-rank matrices to reduce the number of fine-tuning parameters. Next, freeze the original parameters of the large model and only train the LoRA layer and classification layer. Use an appropriate optimizer (such as AdamW) and learning rate scheduler to fine-tune the model with the training dataset, while monitoring the performance of the validation set to adjust the hyperparameters. After fine-tuning, use the test set to evaluate the model performance to ensure that the LoRA fine-tuned model performs well on specific tasks. This paper uses the llama factory fine-tuning framework (Llama-Factory) to implement lora fine-tuning. The fine-tuning training loss and validation loss images are shown in Figure 4.

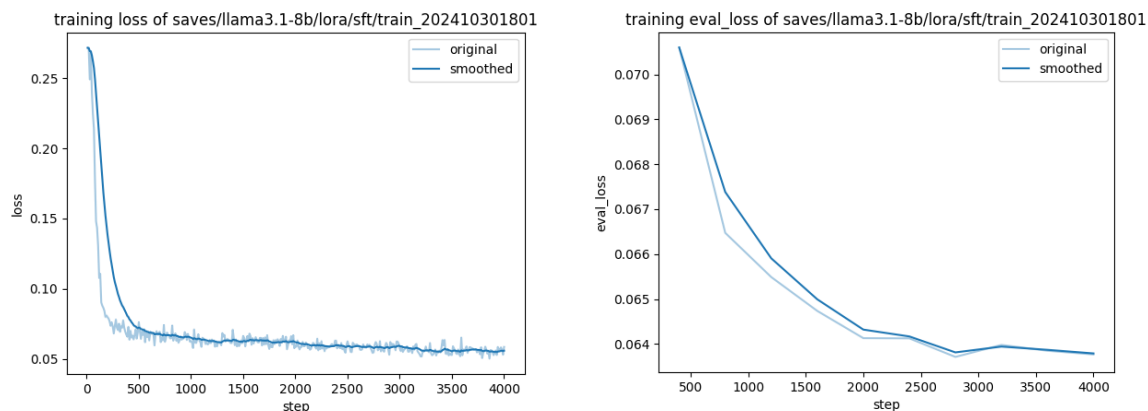


Figure 4. A fine-tuning training and validation loss image

Judging from the training loss graph, the LoRA fine-tuning process of the Llama 3.1B model showed some remarkable characteristics. The loss value dropped sharply at first, showing the model's ability to quickly learn and adapt to new tasks. As the training progressed, the rate of decline of the loss value gradually slowed down and entered a relatively stable stage, indicating that the model is gradually approaching the optimal solution. Despite this, the loss value remains at a low level, reflecting the model's ability to continuously optimize during training. Overall, the results of this LoRA fine-tuning are satisfactory, and the model has shown stable learning ability and continuous improvement potential during training.

3. RESULTS

After fine-tuning the large language model, the model and LoRa weights are deployed locally using the VLLM framework (vllm-project), and the large model is accessed using the API. The results show that the model has preliminary layout capabilities after fine-tuning.

Using the big model Agent technology, we build a drawing tool agent that can call the Revit API, which can convert JSON data into drawings in RVT format. For equipment information, we intelligently create corresponding family instances in Revit based on the data in JSON. At the same time, we process layout information, which involves creating elements such as rooms and areas, and setting their boundaries and properties. Figures 5 and 6 show the results of the case model output.

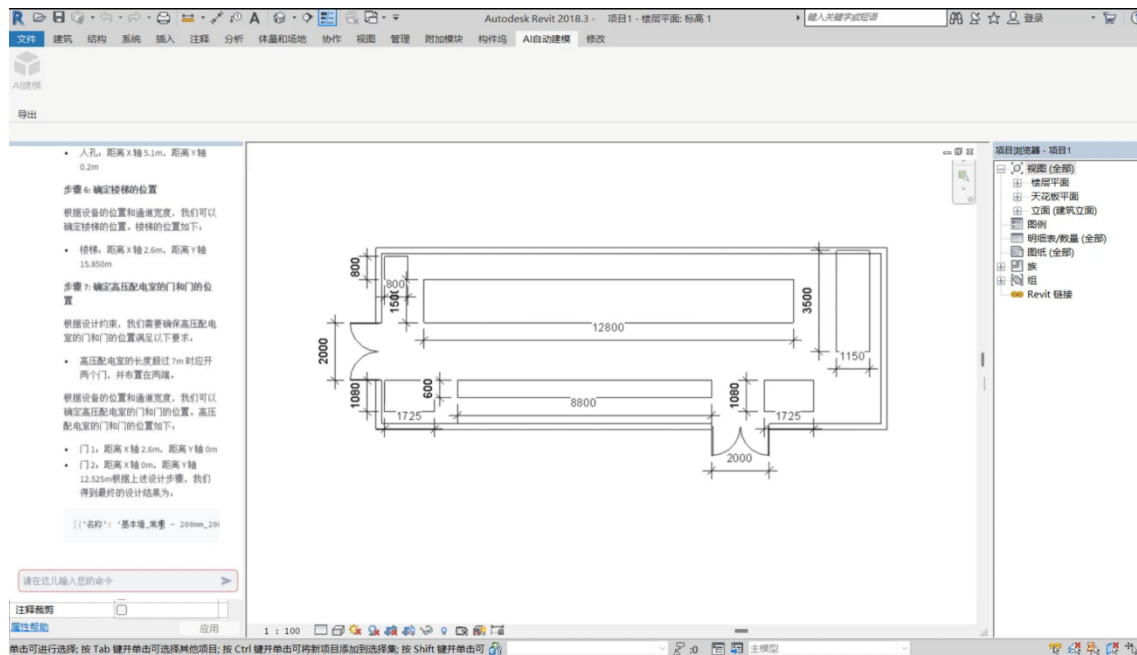


Figure 5. Layout drawing after a model is exported

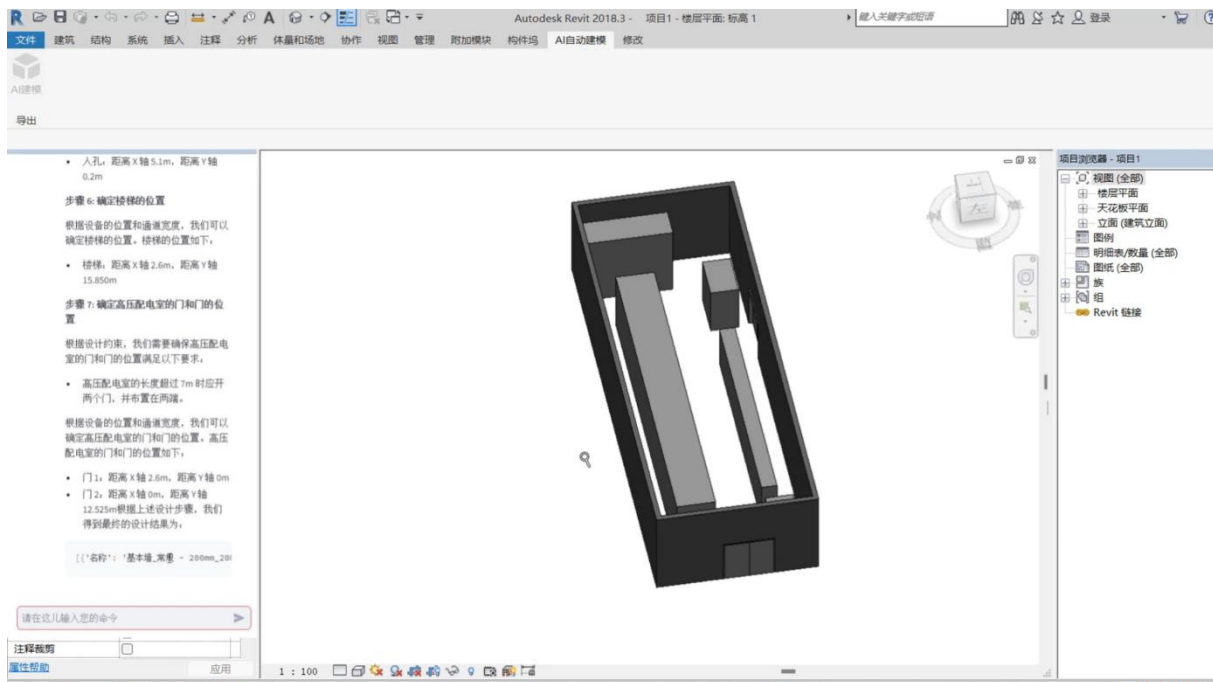


Figure 6. Layout drawing after a model is exported (3D)

Determine the evaluation indicators and display the comparison results. The four contents mentioned in the previous article are used for the comparison experiment. The first is to use no fine-tuning + One-Shot case + design requirements + adding design constraints in the prompt. The second is also in the form of no fine-tuning + Few Shot cases + design requirements + design constraints. The third is a fine-tuning model + design requirements for the combination of design requirements and design solutions. The fourth is a fine-tuning model + design requirements for the combination of (design requirements + intermediate design process + design solution). Display the generation results of the floor plan and the information model under different solutions.

Evaluation index: defines whether the format generated by the model satisfies json as J . If yes $J = true$, otherwise $J = false$.

Define the number of devices in the layout conditions as n , the number of devices output by the model as n' , and the accuracy of the number of devices generated is as Equation (1):

$$T_n = \frac{|n - n'|}{n} \quad (1)$$

Define the x -axis range of the device i to be placed as $[x_{i1}, x_{i2}]$, the y -axis range as $[y_{i1}, y_{i2}]$, the x -axis and y -axis coordinates of device i predicted by the model are x'_i and y'_i , then the device range accuracy is as Equation (2):

$$S = \frac{\sum_1^n x'_i \in [x_{i1}, x_{i2}] \& y'_i \in [y_{i1}, y_{i2}]}{n} \quad (2)$$

The following table 1 compares some of the methods used in the study:

Table 1. Comparison of methods

Methods	Indicator J	Number accuracy T_n	Device range accuracy S
One-shot+Requirements+Constraints	False		
Few-shot+Requirements+Constraints	true	85.1%	15.7%
Sft+Requirements+Constraints	true	95.3%	85.2%
Sft+Requirements+Constraints	true	98.2%	95.1%

From the evaluation results, we can see that the Sft+cot+demand+constraint solution outperforms the other solutions in all aspects, followed by the Sft+demand+constraint solution. Although the Few-shot+demand+constraint solution can complete the task, its range accuracy needs to be improved. The One-shot+demand+constraint solution seems to be the least ideal and needs further improvement or adjustment.

4. CONCLUSIONS

In the field of substation layout design, the use of large language models to achieve AI-driven scene intelligence has far-reaching significance. It can improve the level of design automation, reduce human errors, and significantly improve design efficiency. However, the existing image generation-based solution is incapable of dealing with specific design specifications and engineering constraints, and lacks the necessary flexibility and controllability. The method proposed in this paper achieves a significant improvement in design efficiency and precise control of design details by converting the design scheme into a token sequence consisting of the positioning and orientation data of the equipment bounding box. Nevertheless, this method is limited in its applicability to the diversity of substation types and plane forms. In the future, it is necessary to expand and fine-tune the data set to enhance its adaptability. At the same time, exploring more reasonable layout serialization forms and effectively converting design ideas in actual engineering projects into data that can be understood by the model are also key directions for improvement.

ACKNOWLEDGMENTS

This project is funded by the Shanghai Municipal Electric Power Company's technology project "Research on Key Technologies for AI Intelligent Design of 10kV Substations" SGTYHT/24-JS-001.

REFERENCES

- Agarwala, M., Li, W., & Berthouzoz, F. (2011). Design principles for visual communication. *Communications of the ACM*, 54(4).
- Arroyo, D. M., Postels, J., & Tombari, F. (2021). Variational transformer networks for layout generation. In *CVPR*.
- Gupta, K., Achille, A., Lazarow, J., Davis, L., Mahadevan, V., & Shrivastava, A. (2021). LayoutTransformer: Layout generation and completion with selfattention. In *ICCV*.
- Jiang, Z., Sun, S., Zhu, J., Lou, J.-G., & Zhang, D. (2022). Coarse-to-fine generative modeling for graphic layouts. In *AAAI*.
- Kikuchi, K., Simo-Serra, E., Otani, M., & Yamaguchi, K. (2021). Constrained graphic layout generation via latent optimization. In *ACM MM*.
- Kong, X., Jiang, L., Chang, H., Zhang, H., Gong, H., & Essa, I. (2022). BLT: Bidirectional layout transformer for controllable layout generation. In *ECCV*.
- Retrieved from github website: https://github.com/meta-llama/llama-models/blob/main/models/llama3_1/MODEL_CARD.md
- Zhao, W. X., et al. (2023). A survey of large language models. *arXiv preprint*, 2303.18223.
- Lv, K., et al. (2023). Full parameter fine-tuning for large language models with limited resources. *arXiv preprint*, 2306.09782.

Zheng, Y., et al. (2024). Llamafactory: Unified efficient fine-tuning of 100+ language models. arXiv preprint, 2403.13372.

Hu, E. J., et al. (2021). LoRa: Low-rank adaptation of large language models. arXiv preprint, 2106.09685.

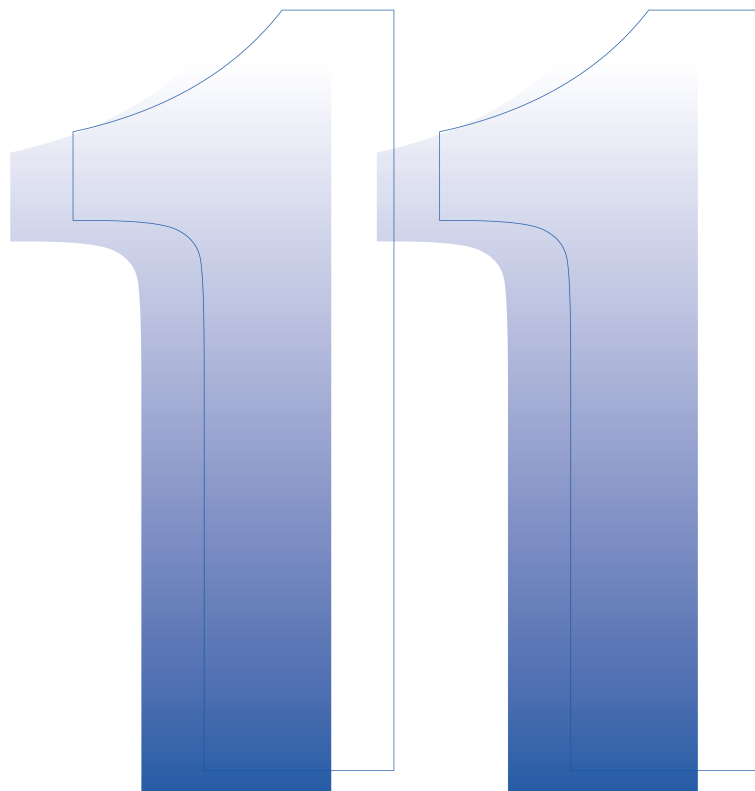
Wei, J., et al. (2022). Chain-of-thought prompting elicits reasoning in large language models. Advances in Neural Information Processing Systems, 35, 24824-24837.

Retrieved from github website:

<https://github.com/hiyouga/LLAMA-Factory>

Retrieved from github website:

<https://github.com/vllm-project/vllm>



Part 11

AI in Construction: Innovations, Risks, and Best Practices

CLOUD TYPE CLASSIFICATION THROUGH SEMANTIC SEGMENTATION FOR ANALYSIS OF EARTH'S RADIATIVE BALANCE

Yu-Wen Wu¹, Nofel Lagrosas², Sheng-Hsiang Wang³, and Albert Y. Chen⁴

1) Graduate Research Assistant, Department of Civil Engineering, National Taiwan University, Taipei, Taiwan. Email: r12521504@ntu.edu.tw

2) Assoc. Prof., Civil Engineering Department, Kyushu University, Fukuoka, Japan. Email: nofel@civil.kyushu-u.ac.jp

3) Professor, Department of Atmospheric Sciences, National Central University, Taoyuan, Taiwan. Email: carlo@g.ncu.edu.tw

4) Professor, Department of Civil Engineering, National Taiwan University, Taipei, Taiwan. Email: albertchen@ntu.edu.tw

Abstract: Classifying nighttime clouds is crucial for understanding their impact on earth's radiative balance. This study presents a semantic segmentation model using U-Net with a MobileNetV3 backbone for classification of the following cloud types: Cirrus, Nimbus, Stratus, and Cumulus from nighttime images. Despite challenges from reduced visibility at night, cloud types and coverage were effectively detected, classified and measured. The model results potentially facilitate future research on nighttime radiation analysis.

Keywords: Night clouds, Ground-based cloud images, Semantic segmentation model

1. INTRODUCTION

Clouds are visible masses of sparse liquid water droplets, ice crystals, or the mixture of both, suspended in the earth's atmosphere. Clouds form when water vapor in the air condenses or sublimates around microscopic particles, known as cloud condensation nuclei, in the atmosphere. This process occurs when air is cooled to its dew point or becomes saturated with moisture.

Clouds are classified based on their appearance, altitude, and the processes that lead to their formation. Cloud types play crucial role in the earth's weather and climate systems; clouds influence temperature, precipitation, and radiative balance by reflecting or trapping heat. Unlike daytime clouds, which are easily observed, nighttime clouds are more difficult to observe due to the absence of sunlight, making their classification challenging. Understanding nighttime clouds is essential to reasoning and quantifying, especially after sunset, how they influence surface temperatures by trapping or releasing heat, thereby affecting weather patterns. The ability to classify clouds enables for more accurate predictions of nighttime temperature variations and weather events, which can be useful in enhancing public safety and preparedness for sudden atmospheric changes.

The significance of classifying nighttime clouds extends beyond immediate weather forecasting and into the realm of long-term climate studies. Nighttime clouds play a pivotal role in the earth's radiative balance, as they can either insulate the surface by trapping longwave radiation or allow cooling by permitting heat to escape into space. Accurately identifying these clouds can lead to refining climate models, helping researchers and modelers predict shifts in temperature, extreme weather, and the broader effects and changes of climate. Since clouds are a major source of uncertainty in climate modeling, a better understanding of their nighttime behavior can improve projections of atmospheric warming and other climate phenomena. This knowledge is critical for making informed decisions regarding climate adaptation and mitigation efforts, as well as for enhancing the accuracy of both weather and climate predictions.

2. LITERATURE REVIEW

Current research on cloud observation relies primarily on both satellite and ground-based observations. Satellite imagery offers convenient access to large-scale global atmospheric motion data at cloud tops. However, the limited resolution of these images often lacks required detailed local information. In contrast, ground-based cloud images provide more detailed regional information, which are particularly crucial for applications such as air traffic control. Additionally, ground-based observations tend to have lower data collection costs (Singh & Glennen, 2005). The types and amounts of clouds also significantly influence radiation and weather changes (Chen et al., 2000). Consequently, this study aims to achieve two main objectives: estimating cloud cover and recognizing cloud types.

To estimate cloud cover, the primary methods are divided into traditional thresholding techniques and deep learning approaches. Thresholding involves analyzing the differences in RGB or grayscale values between clouds and the background in an image, followed by applying a threshold to distinguish the two. Notable examples of this approach include adaptive thresholding that utilizes R-B or R/B ratios (Li et al., 2011) and conversion to grayscale cloud images that consider both visible and infrared light (Lagrosas et al., 2021). In contrast, deep learning approaches leverage architectures such as convolutional neural networks (CNNs) and U-Net (Ronneberger et al., 2015). These models take raw images as input to generate segmented images that effectively separate clouds from the background, as demonstrated in works such as CloudSegNet (Dev et al., 2019), SegCloud (Xie et al., 2020), and CloudU-Net (Shi et al., 2020).

For cloud type recognition, recent methodologies predominantly utilize deep neural networks. CNNs excel at capturing the shape and texture features of various cloud types (Ye et al., 2017), with some studies further classifying clouds into more distinct categories (Zhang et al., 2018). Techniques such as Bagging and AdaBoost have been shown to enhance classification accuracy (Zhang et al., 2020). However, this line of research often encounters challenges in accurately marking the coverage area of clouds and struggles when multiple cloud types are present in the same image (Ye et al., 2019).

Despite the advancements in i) cloud detection through image segmentation and ii) cloud recognition through image classification, there has been considerably less research that effectively combines these two approaches. One such method involves segmenting the image into superpixels before classifying them (Ye et al., 2019). More recent efforts have included end-to-end methods using encoder and decoder architectures (Ye et al., 2022) and a U-Net approach integrated with an attention mechanism (Shi et al., 2024). Notably, all these methods have primarily focused on daytime cloud datasets.

Overall, the advancement of deep learning models has significantly propelled research in cloud segmentation and classification. Nevertheless, there remains a notable gap in studies that address both tasks concurrently, particularly in relation to nighttime cloud datasets. Nighttime clouds are essential for understanding the earth's radiation balance, influencing both short-term weather and long-term climate. Therefore, this study aims to develop a deep learning-based semantic segmentation model for nighttime clouds to address this research gap. To this end, MobileNetV3 (Howard et al., 2019) will be utilized due to its small parameter size, rapid computation speed, and ability to maintain good accuracy, making it a suitable choice for constructing the model in this study.

3. METHOD

3.1 Overall Architecture

The architecture is based on U-Net, which is widely used for image segmentation. By taking a 416x320 ground cloud image as input, the U-Net performs feature extraction and up-sampling, ultimately producing an output of the same size that predicts which type of cloud each pixel belongs to, as shown in Figure 1. This allows for the estimation of area proportions of different cloud types, which serves as a basis for comparison with the observed radiation levels. The predicted cloud types include four categories: Cirrus, Nimbus, Stratus, and Cumulus, represented by red, green, blue, and yellow, respectively, in the final prediction image (along with black for the sky).

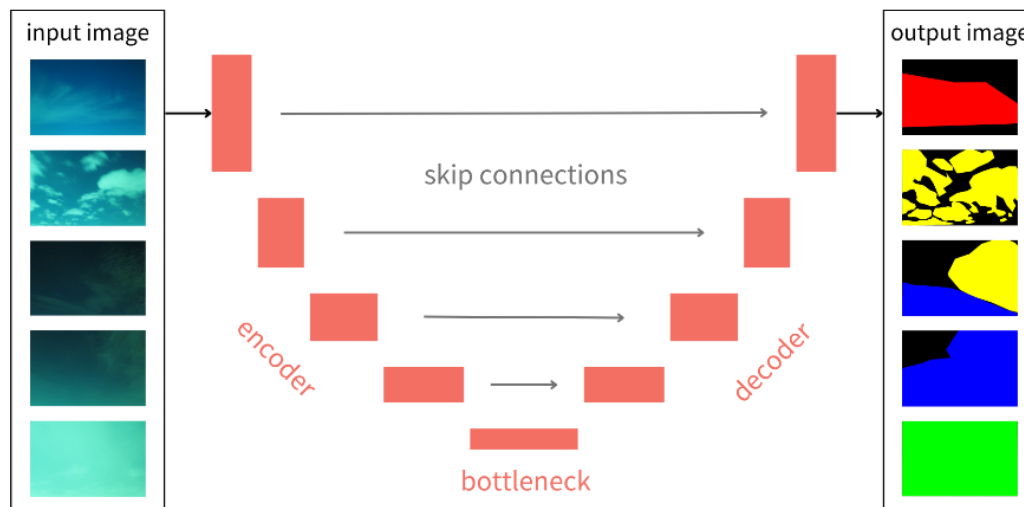


Figure 1. Framework of the cloud segmentation model

3.2 Model Configuration and Training Setup

The model employs a U-Net architecture with MobileNetV3 as the backbone. To enhance generalization, data augmentation was applied using a variety of transformations, including horizontal flip, vertical flip, grid distortion, random brightness-contrast adjustment, and Gaussian noise. The batch size is set to 4, and pretrained weights from ImageNet (Krizhevsky et al., 2012) were utilized to speed up model convergence, and fine tuning/transferring learning was conducted for 150 epochs of training. The optimization process uses the AdamW optimizer (Ilya Loshchilov & Frank Hutter, 2017) with a weight decay of $1e-4$ to mitigate overfitting. The initial learning rate is set to $1e-3$, and a cosine annealing scheduler (Ilya Loshchilov & Frank Hutter, 2016) is applied to dynamically adjust the learning rate during training, improving model performance.

Table 1. Training setup details

Parameter	Details
Model Architecture	U-Net with MobileNetV3 as the backbone
Data Augmentation	Horizontal Flip, Vertical Flip, Grid Distortion, Random Brightness-Contrast, Gaussian Noise
Batch Size	4
Pretrained Weights	ImageNet
Training Epochs	150
Optimizer	AdamW
Weight Decay	1e-4
Initial Learning Rate	1e-3
Learning Rate Scheduler	Cosine Annealing

3.3 Evaluation Metrics

The performance of the model is evaluated using several common metrics for segmentation tasks. First, cross entropy loss is employed as the loss function to measure the pixel-wise classification error between the predicted segmentation and the ground truth.

Additionally, pixel accuracy is utilized to assess the overall proportion of correctly classified pixels in the entire image, providing a general indication of the model's classification accuracy.

However, since pixel accuracy can be biased towards dominant classes in imbalanced datasets, the weighted Intersection over Union (wIoU) is also included as a complementary metric. The wIoU considers the intersection and union of the predicted and true segments for each class, using the proportion of each class in the ground truth as a weight, providing a more robust evaluation. The formula for wIoU is shown in Equation (1), where N is the total number of classes, and both i and $j = \{1, 2, 3, 4, 5\}$, corresponding to the five classes: Sky, Cirrus, Nimbus, Stratus, and Cumulus. GT refers to the number of ground truth pixels. TP , TN , FP , and FN refer to True Positive, True Negative, False Positive, and False Negative, respectively.

$$wIoU = \sum_{i=1}^N \frac{GT_i}{\sum_{j=1}^N GT_j} \cdot \frac{TP_i}{TP_i + FP_i + NP_i} \quad (1)$$

4. RESULTS

Data was collected with a Canon A2300 camera, programmed to continuously take nighttime sky images every 5 min at National Central University, Taiwan (24.97°N, 121.19°E). The infrared cut filter of the camera was manually removed so that the camera can function as an infrared camera. By doing this, a brighter image of the nighttime sky was observed. A total of 223 ground-based cloud images between 15 December 2021 to 19 February 2022 were then categorized into five types: Cirrus, Nimbus, Stratus, Cumulus, and Mix. Figure 2 depicts sample cloud images from the dataset.

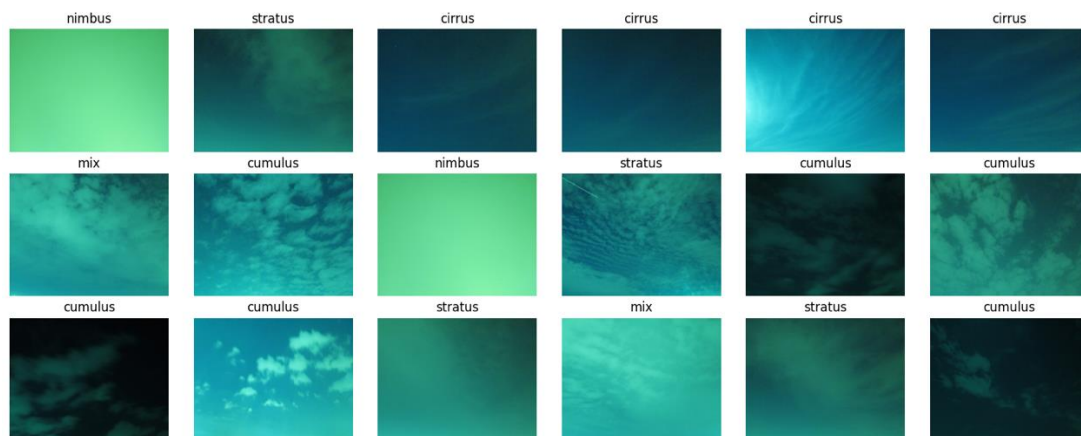


Figure 2. Sample cloud images from the dataset

To evaluate the model's performance, the dataset was split into training and validation sets with a 4:1 ratio. Additionally, to ensure the model's robustness, the proportion of different cloud types in both the training and validation sets was kept consistent. Table 2 shows the detailed information.

Table 2. Composition of training and validation datasets

Types	Training Size	Validation Size	Total
Cirrus	40	9	49
Nimbus	58	14	72
Stratus	24	5	29
Cumulus	43	10	53
Mix	16	4	20

Figure 3 illustrates the learning curves of the model throughout training, showing the progression of three key metrics: loss, average pixel accuracy, and average wIoU. All metrics exhibit a consistent improvement during the training process and gradually stabilize as they approach convergence. Notably, the validation curves for all metrics maintain a close alignment with the training curves, with the validation performance often surpassing the training results slightly. This indicates that the model not only avoids overfitting but also generalizes well, demonstrating strong robustness and the ability to perform effectively on unseen data.



Figure 3. Learning curves of the model

The model with the lowest validation loss was selected as the final model, and some prediction results on the validation dataset are shown in Figure 4. These results indicate that the model not only performs well under ideal conditions but also exhibits robustness when confronted with challenging scenarios, such as thinner cloud layers and darker images. Table 3 provides a comprehensive overview of the final model's performance across different categories.

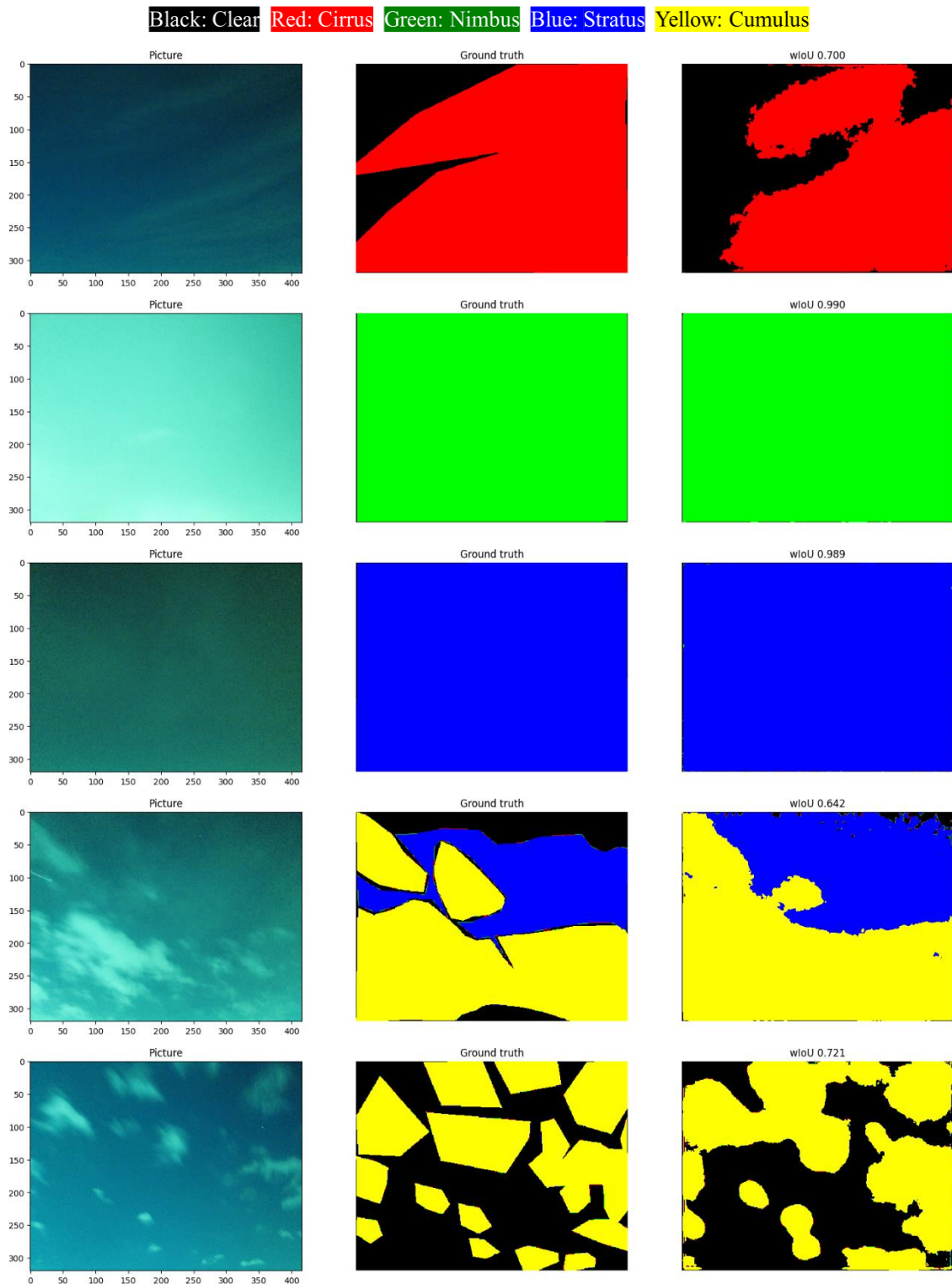


Figure 4. Some results on the validation dataset

Table 3. Final model performance on validation dataset

Types	Average Pixel Accuracy	Average wIoU
Cirrus	0.82	0.73
Nimbus	0.99	0.99
Stratus	0.97	0.96
Cumulus	0.87	0.84
Mix	0.64	0.47

The Cirrus cloud type presents a greater degree of variability, both in terms of its shape and the way it interacts with lighting conditions. This cloud formation often appears delicate and wispy, with significant differences in density and texture across different regions. Our prediction results, illustrated in Figure 5, demonstrate that the model has successfully identified and captured these nuances. The model was able to detect not only the denser, more prominent portions of the cloud but also the finer, more translucent areas. This highlights the model's ability to discern both subtle and obvious features within the cloud structure, showcasing its effectiveness in handling the complexity and variability inherent in Cirrus formation.

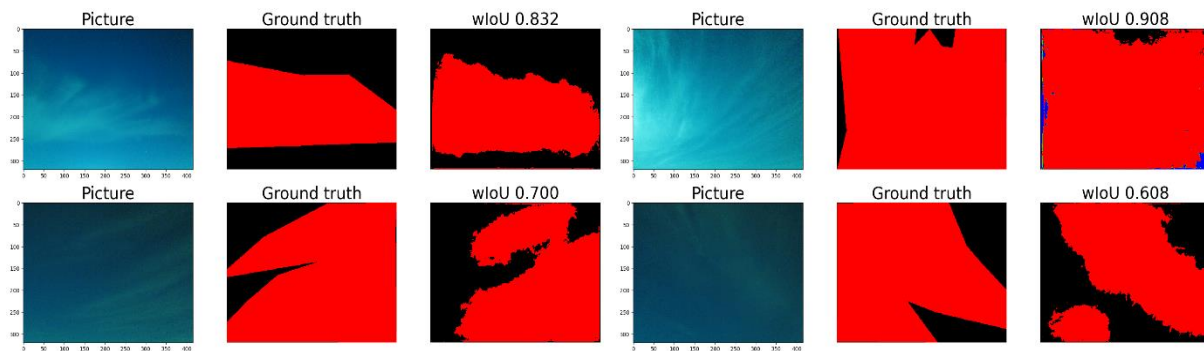


Figure 5. Cirrus cloud prediction results

The Nimbus cloud type, when present, typically cover the entire sky with their dense masses. This extensive coverage distinguishes Nimbus from other cloud types, which may appear more localized or scattered across the sky. Additionally, the color distribution in images containing Nimbus clouds is unique, as shown in Figure 6. This distinct color profile further sets them apart from other cloud formations, facilitating a more accurate classification of cloud types.

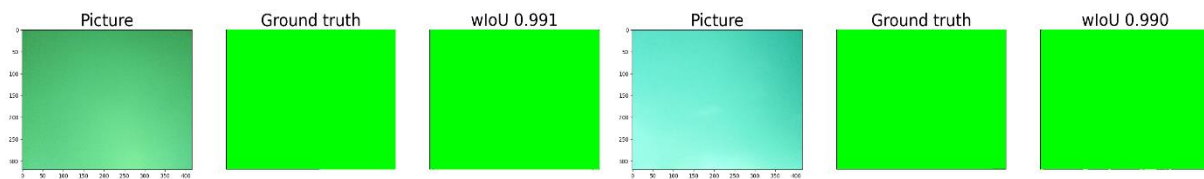
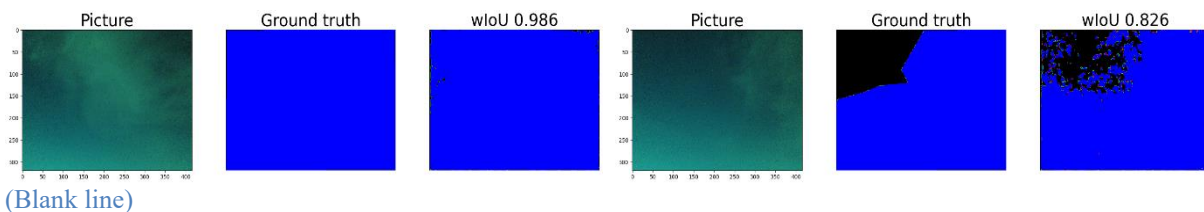


Figure 6. Nimbus cloud prediction results

Stratus clouds, when present, also typically cover the sky extensively (as shown in Figure 7), contributing to improved model accuracy. However, the model does not mistakenly classify entire images as Stratus clouds solely due to their presence. This refined behavior is particularly evident in the predictions for mixed cloud formations, as illustrated in Figure 9.



(Blank line)

Figure 7. Stratus cloud prediction results

Cumulus clouds typically exhibit distinct shapes and bright white colors, making them ideal for model recognition. The results were promising, as expected, with the model performing consistently well across datasets with varying brightness levels.

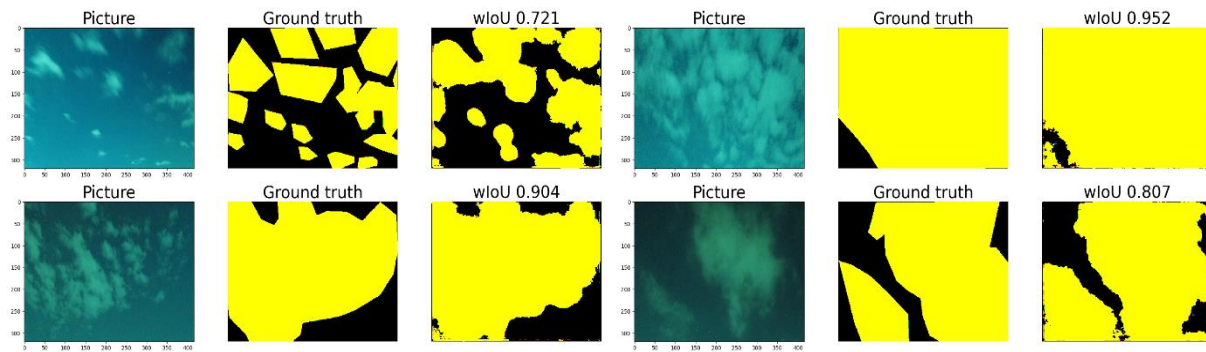


Figure 8. Cumulus cloud prediction results

In terms of predicting mixed cloud cover, we also achieved good results. This work demonstrates that it is possible to simultaneously perform cloud segmentation and classification for nighttime clouds, even when multiple types of clouds are present in the same image. However, there is still room for improvement in predicting images with very low brightness or extremely thin clouds, which makes our average pixel accuracy under the 'Mix' category look quite good, but the performance of the average wIoU is relatively poorer (as shown in Table 3). This will be one of the directions for future efforts.

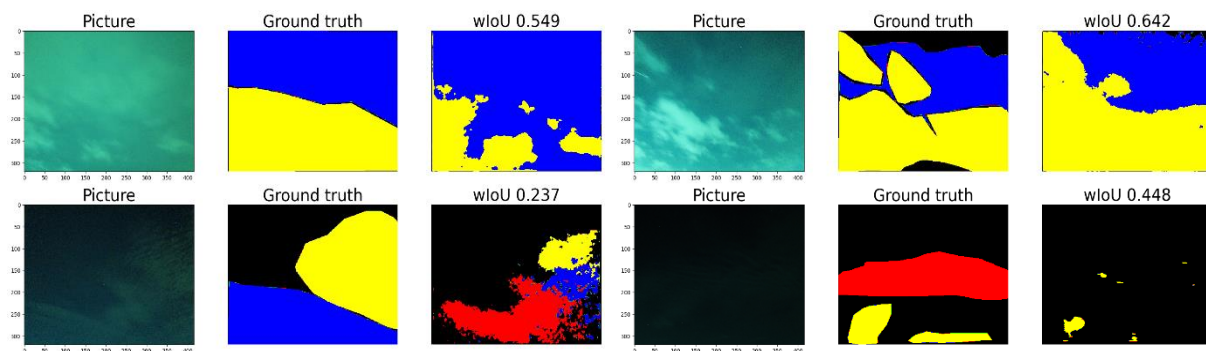


Figure 9. Mix cloud prediction results

To evaluate the effectiveness of the model, Table 4 presents a comparison of average pixel accuracy and average IoU with other studies. Although fewer categories were predicted, which will be one of the future improvement directions, the model still demonstrates good performance.

Table 4. Comparison with previous studies

Research	Predicted Categories	Acc	IoU
Ye et al., 2019	Sky, Cu, Sc, St, As, Ac, Cc, Cs, Ci	0.71	0.34
Ye et al., 2022	Sky, Cu, Sc, St, As, Ac, Cc, Cs, Ci	0.85	0.37
Shi et al., 2024	Sky, Cirrus, Stratus, Cumulus, Cumulonimbus & Laminatus	0.91	0.62
Ours	Sky, Cirrus, Nimbus, Stratus, Cumulus	0.86	0.80

5. CONCLUSION AND FUTURE WORKS

This study developed a deep learning-based semantic segmentation model to classify nighttime cloud types. The model demonstrated satisfactory accuracy and robustness across various cloud types, including Cirrus, Nimbus, Stratus, and Cumulus, showing promise for application in nighttime cloud monitoring. Notably, the model effectively managed challenges associated with low visibility at night, capturing essential cloud features and enabling further analysis of cloud type contributions to the earth's radiative balance.

Future work will address current model limitations, including detecting extremely low-brightness images in mixed clouds. Cloud types may also be further classified, such as dividing Stratus into Stratocumulus and Stratocirrus. Furthermore, the amount of net longwave radiation that can be attributed to each cloud type will be explored to understand the radiative effects of each cloud type.

ACKNOWLEDGMENTS

The authors thank the support from the National Science and Technology Council of Taiwan for grant NSTC 113-2628-E-002-026-MY3.

REFERENCES

- Chen, T., Rossow, W. B., & Zhang, Y. (2000). Radiative effects of cloud-type variations. *Journal of climate*, 13(1), 264-286.
- Dev, S., Nautiyal, A., Lee, Y. H., & Winkler, S. (2019). CloudSegNet: A deep network for nychthemeron cloud image segmentation. *IEEE Geoscience and Remote Sensing Letters*, 16(12), 1814-1818.
- Howard, A., Sandler, M., Chu, G., Chen, L. C., Chen, B., Tan, M., ... & Adam, H. (2019). Searching for mobilenetv3. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 1314-1324).
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.
- Lagrosas, N., Shiina, T., & Kuze, H. (2021). Observations of nighttime clouds over Chiba, Japan, using digital cameras and satellite images. *Journal of Geophysical Research: Atmospheres*, 126(17), e2021JD034772.
- Li, Q., Lu, W., & Yang, J. (2011). A hybrid thresholding algorithm for cloud detection on ground-based color images. *Journal of atmospheric and oceanic technology*, 28(10), 1286-1296.
- Loshchilov, I. (2017). Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.
- Loshchilov, I., & Hutter, F. (2016). Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*.
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18* (pp. 234-241). Springer International Publishing.
- Shi, C., Zhou, Y., Qiu, B., Guo, D., & Li, M. (2020). CloudU-Net: A deep convolutional neural network architecture for daytime and nighttime cloud images' segmentation. *IEEE Geoscience and Remote Sensing Letters*, 18(10), 1688-1692.
- Shi, C., Su, Z., Zhang, K., Xie, X., Zheng, X., Lu, Q., & Yang, J. (2024). CloudFU-Net: A Fine-grained Segmentation Method For Ground-based Cloud Images Based On An Improved Encoder-Decoder Structure. *IEEE Transactions on Geoscience and Remote Sensing*.
- Singh, M., & Glennen, M. (2005). Automated ground-based cloud recognition. *Pattern analysis and applications*, 8, 258-271.
- Xie, W., Liu, D., Yang, M., Chen, S., Wang, B., Wang, Z., ... & Zhang, C. (2020). SegCloud: A novel cloud image segmentation model using a deep convolutional neural network for ground-based all-sky-view camera observation. *Atmospheric Measurement Techniques*, 13(4), 1953-1961.
- Ye, L., Cao, Z., & Xiao, Y. (2017). DeepCloud: Ground-based cloud image categorization using deep convolutional features. *IEEE Transactions on Geoscience and Remote Sensing*, 55(10), 5729-5740.
- Ye, L., Cao, Z., Xiao, Y., & Yang, Z. (2019). Supervised fine-grained cloud detection and recognition in whole-sky images. *IEEE Transactions on Geoscience and Remote Sensing*, 57(10), 7972-7985.
- Ye, L., Cao, Z., Yang, Z., & Min, H. (2022). Ccad-net: A cascade cloud attribute discrimination network for cloud genera segmentation in whole-sky images. *IEEE Geoscience and Remote Sensing Letters*, 19, 1-5.
- Zhang, J., Liu, P., Zhang, F., & Song, Q. (2018). CloudNet: Ground-based cloud classification with deep convolutional neural network. *Geophysical Research Letters*, 45(16), 8665-8672.
- Zhang, J., Liu, P., Zhang, F., Iwabuchi, H., e Ayres, A. A. D. H., & De Albuquerque, V. H. C. (2020). Ensemble meteorological cloud classification meets internet of dependable and controllable things. *IEEE Internet of Things Journal*, 8(5), 3323-3330.

TRANS-DIMENSIONAL BAYESIAN INVERSION FOR ESTIMATING EARTH PRESSURES ON IN-SERVICE UNDERGROUND STRUCTURES

Zhiyao Tian¹ and Xianfei Yin²

1) Ph.D., Postdoc, Department of Architecture and Civil Engineering, City University of Hong Kong, Hong Kong, China.
Email: zhiytian@cityu.edu.hk

2) Ph.D., Asst. Prof., Department of Architecture and Civil Engineering, City University of Hong Kong, Hong Kong, China.
Email: xianfyin@cityu.edu.hk

Abstract: Identification of earth pressures acting on in-service underground structures is critical for their health monitoring and performance prediction. Given that the extensive deployment of sensors on poorly performing structures to measure pressure incurs high costs and presents technical challenges, the inversion of these pressures from easily observed deformation data has become increasingly desirable. However, traditional pressure inversion methods require subjective assumptions about the complexity of the pressure, necessitating extensive engineering judgment that may not be confidently applied in practice. To address this challenge, this paper proposes a trans-dimensional Bayesian method for pressure inversion. This method simultaneously incorporates pressure complexity and quantities into the inversion by parameterizing a set of previously unknown parameters, where the number of parameters itself is unknown. A Bayesian framework is then used to represent the posterior distribution of the parameter set. A trans-dimensional Markov chain is specifically designed to estimate the posterior distribution of the parameters, achieving statistical inversion of the parameters as well as the distributed pressures. A recorded case study is presented for illustration and verification. It is found that the proposed method yields good inversion results on the pressures on a diaphragm wall, whereas traditional methods lead to poor inversion results due to inadequate assumptions. These outcomes highlight the advancements of the proposed method. Lastly, deficiencies and future extensions are discussed in the conclusion.

Keywords: Underground structures, Structural health monitoring, Load inversion, Bayesian method, Trans-dimensional Markov chain.

1. INTRODUCTION

With the rapid development of infrastructure construction, a wide range of underground structures have been put into service. However, the in-service environment of these structures is exceedingly complex (Yin et al., 2020; Tian et al., 2022). Factors such as extensive human engineering activities and unforeseen changes in geotechnical conditions can significantly disturb these structures, often resulting in performance deterioration that far exceeds initial design expectations (Zhou et al., 2020; Gong et al., 2023). This issue is well-documented by numerous practical cases (Van Empel et al. 2016; Bowers & Moss, 2018; Zhang et al., 2019); for instance, excessive deformations and serious defects have been frequently observed in subway tunnel linings worldwide, particularly in cities characterized by soft soil conditions (Carpio et al., 2019; Tian et al., 2023a). Ensuring the operational safety of these infrastructures has emerged as a significant challenge for engineers.

Earth pressure acts as a mediator in the complex interactions between underground structures and the surrounding strata, playing a crucial role in influencing structural performance. Accurately identifying the current earth pressures is essential for the health monitoring and performance prediction of these in-service structures (Liu et al., 2016; Tian et al., 2023a). However, embedding sensors to monitor the pressures in structures that have already been cast and buried presents significant technical challenges (Zhou et al., 2024). Furthermore, the cost of equipping all newly built underground structures with sensors is prohibitively expensive (Zhou et al., 2021). Consequently, the inversion of earth pressures from easily observable structural responses, say deformations, has emerged as a more desirable approach.

Pressure inversion consists of two primary process, i) parametrization, i.e., use parameters to represent *a priori* unknown pressures; and ii) inference, i.e., deduce the values of these parameters from observed data to reconstruct the entire pressure distribution. A straightforward approach assumes earth pressures based on a predefined design mode, parameterizing the distributed pressures into specific parameters defined by the mode (e.g., $\mathbf{q}=(q_1, q_2, q_3)^T$ in Fig. 1a), and then employing optimization methods to find solutions (values of \mathbf{q}) that yield structural responses closely aligned with the observed data (Yan et al., 2019).

However, due to the complex underground environment, earth pressures acting on in-service structures may distribute unevenly, often deviating significantly from the design mode. To address this, an interpolation technique has been proposed that parameterizes the unknown pressures using a set of unknown knots (e.g., $\mathbf{q}=(q_1, \dots, q_n)^T$ in Fig. 1b) (Gioda & Jurin, 1981). This interpolation approximates the unknown pressures, transforming the inversion of pressures into the inversion of the unknown knot parameters.

While this technique abandons strict design mode assumptions, it introduces two significant issues by expanding the parameter space: ill-conditioning and non-uniqueness (Sanchez & Benaroya, 2014). Specifically,

minor errors in observed data can lead to substantial biases in inversion results (ill-conditioning), while vastly different pressures may yield similar structural responses (non-uniqueness). (Liu et al., 2019) and (Liu et al., 2021) observed these phenomena and employed regularization techniques to address ill-conditioning, achieving satisfactory inversion results. Additionally, (Tian et al. 2023b) and (Tian et al., 2024a) introduced a Bayesian approach capable of quantifying uncertainties associated with non-uniqueness, thereby informing the strategic addition of observations to mitigate non-uniqueness. Despite these advancements, a notable limitation persists: the complexity of pressure distributions must still be predefined. As indicated in Fig. 1(b), the number and locations of knots are assumed prior to inversion. For example, the unknown knots are typically assumed to be evenly distributed across the structural domain. However, for complex pressure distributions, as exemplified in Fig. 1(c), abrupt changes in pressures necessitate densely predefined unknown knots locally for improved inversion outcomes. These assumptions require considerable engineering judgement, which may not always be confidently applied in practice.

To address this problem, this paper introduces a trans-dimensional Bayesian inversion approach that jointly incorporates the complexity and specific values of *a priori* unknown earth pressures into the inversion process, eliminating the need for preconceived assumptions and engineering judgment regarding pressure distribution. Section 2 details this method, while Sections 3–4 presents a numerical case to discuss the superiority of the proposed method. Deficiencies and future extensions are discussed in the conclusion.

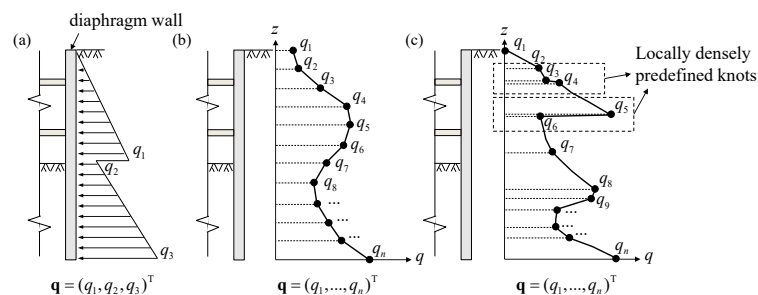


Figure 1. Schematic illustrating the parameterization of earth pressures on in-service underground structures: (a) parameterization based on a predefined design model; (b) parameterization utilizing an interpolation technique; (c) parameterization for complex pressure distributions exhibiting abrupt changes

2. THE TRANS-DIMENSIONAL BAYESIAN METHOD

2.1 Parameterization

Targeting the limitations of the methodologies previously discussed, an idea is proposed that why not integrates both pressure complexity and pressure values into the inversion process? To this end, a trans-dimensional parametrization method is introduced, as detailed below. As shown in Fig. 2, z denotes a generalized coordinate on a structure, such as depth on a diaphragm wall or a polar angle on a shield tunnel ring. $q(z)$ represents the *a priori* unknown distributed earth pressures acting on the underground structure across its domain. The parameterization of the earth pressure still employs an interpolation method using a set of knots. However, this method now incorporates the number of knots (n), the locations of the knots (\mathbf{z} , where $\mathbf{z}=(z_1, \dots, z_n)^T$), and the values at these knots (\mathbf{q} , where $\mathbf{q}=(q_1, \dots, q_n)^T$) jointly into the inversion process. This integration eliminates the need for assumptions regarding the complexity of pressure, i.e., n and \mathbf{z} . The interpolation is performed using the interpolation function presented in Equation (1):

$$q(z) = \mathbf{I}(z)\mathbf{q} \quad (1)$$

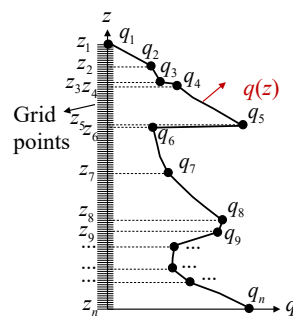


Figure 2. Schematic illustrating the trans-dimensional parameterization method

As a result, the inversion of the *a priori* unknown earth pressures $q(z)$ is transformed into the inversion of an unknown parameter set \mathbf{m} , where $\mathbf{m}=(n, \mathbf{z}, \mathbf{q})$, and $\mathbf{I}(z)$ is the interpolation operator related to n and \mathbf{z} . A detailed derivation is presented in (Press, 2007). Regarding the unknown parameters, n is a discrete random variable taking values in \mathbf{N}^+ ; q_i ($i=1, \dots, n$) is a continuous random variable defined over \mathbf{R} ; for mathematical convenience, z_i ($i=1, \dots, n$) is considered a discrete random variable taking values at grid points that are evenly distributed across the entire structural domain (Fig. 2). In practical applications, it is advisable to set the grid points as densely as possible, with subsequent case study potentially providing a reference for their specific configurations.

2.2 The Bayesian framework

The objective is the inference of the values of \mathbf{m} given the observed structural responses \mathbf{d} , say deformations. This inference is conducted within a Bayesian framework as represented by Equation (2):

$$p(\mathbf{m}|\mathbf{d}) \propto p(\mathbf{d}|\mathbf{m})p(\mathbf{m}) \quad (2)$$

where, $p(\mathbf{m}|\mathbf{d})$ is the posterior distribution, representing the statistical inference results of \mathbf{m} given \mathbf{d} ; $p(\mathbf{d}|\mathbf{m})$ and $p(\mathbf{m})$ represent the likelihood function and the prior distribution, respectively.

(1) Prior distribution

The prior distribution is further elaborated in Equation (3):

$$p(\mathbf{m}) = p(n)p(\mathbf{z}|n)p(\mathbf{q}|n, \mathbf{z}) \quad (3)$$

A Poisson distribution (with parameter λ) is employed to specify the prior for the number of parameters, i.e., $n \sim \text{Poisson}(\lambda)$. This prior effectively limits model complexity and aids in preventing overfitting (Denison et al., 1998). For clarity, let N_g denote the number of grid points across the structural domain. The prior for \mathbf{z} given n is considered as placing n knots without replacement among the N_g points, with each point being selected with equal probability; The parameter \mathbf{q} is assumed to follow a uniform distribution within the physically plausible range $[q_{\min}, q_{\max}]$, with mathematical details presented in Equation (4). These priors are intentionally designed to be weakly informative to accommodate scenarios in which the earth pressure on in-service structures may significantly exceed the values anticipated during the design stage. To prevent the imposition of overly stringent and potentially erroneous constraints from the priors on the inversion results, this strategy allows the observed data to exert a more pronounced influence on the posterior distribution.

$$\begin{cases} p(n) = \frac{\lambda^n}{n!} e^{-\lambda} \\ p(\mathbf{z}|n) = \frac{1}{C_{N_g}^n} \\ p(\mathbf{q}|n, \mathbf{z}) = \prod_{i=1}^n p(q_i|n, z_i) = \prod_{i=1}^n \frac{1}{q_{\max} - q_{\min}} \end{cases} \quad (4)$$

(2) Likelihood function

The likelihood function quantifies the fit between observed structural responses and those predicted by the forward model $g(\cdot)$ under earth pressure $q(z)$. In line with the Central Limit Theorem, the likelihood function is generally assumed to follow a zero-mean Gaussian distribution, as shown in Equation (5):

$$p(\mathbf{d}|\mathbf{m}) = \frac{1}{(2\pi\sigma_d^2)^{l/2}} \exp\left\{-\frac{[\mathbf{d} - g(q(z))]^T [\mathbf{d} - g(q(z))]}{2\sigma_d^2}\right\}, \quad (5)$$

where, σ_d is the expected standard deviation, measuring the dispersion between observed and predicted structural responses; $g(q(z))$, also referred to as the load-structure model, computes the structural responses under earth pressure $q(z)$, determined by the parameter set \mathbf{m} . Given the potential complexity of the distribution, the Finite Element Method (FEM) can be utilized to construct the forward model, detailed in Equation (6):

$$g(q(z)) = g(\mathbf{I}(z)\mathbf{q}) = \mathbf{K}^{-1}\mathbf{f}(\mathbf{I}(z)\mathbf{q}) \quad (6)$$

where, the structure is discretized into a series of elements, with \mathbf{K} representing the global stiffness matrix; \mathbf{f} being a vector-valued function where $\mathbf{f}(\mathbf{I}(z)\mathbf{q})$ denotes the equivalent global forces equivalent to the distributed pressure $\mathbf{I}(z)\mathbf{q}$, adhering to the transformation rules of virtual work. For a detailed derivation of \mathbf{K} and \mathbf{f} regarding typical underground structures such as piles and tunnels, the reader is referred to (Tian et al., 2023b) and (Tian et al., 2024a).

2.3 The Trans-dimensional Markov chain

Directly substituting Equations (3)–(6) into Equation (2) to derive an analytical solution is impractical. Instead, the Markov Chain Monte Carlo (MCMC) method provides a numerical approach for sampling the posterior distribution. However, traditional MCMC is tailored for fixed random variables and cannot accommodate the parameter set \mathbf{m} , where the number of parameters n is itself unknown. To address this, a trans-dimensional MCMC method is presented. (Green, 1995) proposed the reversible jump MCMC method, which effectively tackles trans-dimensional problems by establishing the detailed balance condition for trans-dimensional parameters, as shown in Equation (7):

$$p(\mathbf{m}^*|\mathbf{d})Q(\mathbf{m}^*|\mathbf{m}_c)\alpha(\mathbf{m}^*|\mathbf{m}_c) = p(\mathbf{m}_c|\mathbf{d})Q(\mathbf{m}_c|\mathbf{m}^*)\alpha(\mathbf{m}_c|\mathbf{m}^*)|\mathbf{J}|, \quad (7)$$

where, $Q(\mathbf{m}^*|\mathbf{m}_c)$ is the proposal function that generates candidate parameters \mathbf{m}^* from the current state \mathbf{m}_c ; $\alpha(\mathbf{m}^*|\mathbf{m}_c)$ represents the acceptance ratio for adopting \mathbf{m}^* as the next state in the Markov chain; \mathbf{J} is the Jacobian matrix for the diffeomorphism from \mathbf{m}_c to \mathbf{m}^* . This setup enables the design of a proposal function that facilitates transitions between different dimensions, leading to the derivation of the necessary acceptance ratio (Equation (8)) for constructing a trans-dimensional Markov chain:

$$\alpha(\mathbf{m}^*|\mathbf{m}_c) = \min\left\{1, \frac{p(\mathbf{m}_c|\mathbf{d})Q(\mathbf{m}_c|\mathbf{m}^*)}{p(\mathbf{m}^*|\mathbf{d})Q(\mathbf{m}^*|\mathbf{m}_c)}|\mathbf{J}|\right\}. \quad (8)$$

Equation (9) shows the expansion of the proposal function. Following the principles outlined by (Green, 1995), we have developed subsequent proposal functions for addressing the earth pressure inversion problem, based on the three components of Equation (9). Specifically, each iteration of the chain involves proposing three types of proposals—birth, death, and move—with equal probability, as depicted in Fig. 3. These steps enable transitions between different parameter dimensions, as indicated in Equation (10):

$$Q(\mathbf{m}^*|\mathbf{m}_c) = Q(n^*, \mathbf{z}^*, \mathbf{q}^*|n_c, \mathbf{z}_c, \mathbf{q}_c) = Q(n^*|n_c, \mathbf{z}_c, \mathbf{q}_c)Q(\mathbf{z}^*|n_c, \mathbf{z}_c, \mathbf{q}_c, n^*)Q(\mathbf{q}^*|n_c, \mathbf{z}_c, \mathbf{q}_c, n^*, \mathbf{z}^*) \quad (9)$$

$$Q(n^*|n_c, \mathbf{z}_c, \mathbf{q}_c) = Q(n^*|n_c) = \begin{cases} 1/3 & n^* = n_c + 1, \text{ birth} \\ 1/3 & n^* = n_c - 1, \text{ death} \\ 1/3 & n^* = n_c, \text{ move} \end{cases} \quad (10)$$

i) Birth step (Fig. 3a): increase the number of parameters by 1; add a new knot z_b randomly at a candidate point previously unoccupied; and assign a value to this new knot. The value q_b for the new knot is generated based on a Gaussian perturbation to previous interpolated state value q_p , following $q_b \sim N(q_p, \sigma_b^2)$, where σ_b^2 is the variance of the proposed increments. Accordingly, the remaining two components of the proposal function for the birth step can be written as:

$$\begin{cases} Q_{\text{Birth}}(\mathbf{z}^*|n_c, \mathbf{z}_c, \mathbf{q}_c, n^*) = Q(z_b|n, \mathbf{z}) = \frac{1}{C_{N_g-n}^1} = \frac{1}{N_g - n} \\ Q_{\text{Birth}}(\mathbf{q}^*|n_c, \mathbf{z}_c, \mathbf{q}_c, n^*, \mathbf{z}^*) = Q(q_b|q_p) = \frac{1}{(2\pi\sigma_b^2)^{1/2}} \exp\left[-\frac{(q_b - q_p)^2}{2\sigma_b^2}\right] \end{cases} \quad (11)$$

ii) Death step (Fig. 3b): decrease the number of parameters by 1; and choose one existing knot randomly to delete, e.g., delete z_d and the corresponding coefficient q_d . Similarly, the remaining two components of the proposal function for the death step can be detailed:

$$\begin{cases} Q_{\text{Death}}(\mathbf{z}^*|n_c, \mathbf{z}_c, \mathbf{q}_c, n^*) = Q(z_d|n) = \frac{1}{C_n^1} = \frac{1}{n} \\ Q_{\text{Death}}(\mathbf{q}^*|n_c, \mathbf{z}_c, \mathbf{q}_c, n^*, \mathbf{z}^*) = Q(q_b|z_d) = 1 \end{cases} \quad (12)$$

iii) Move step (Fig. 3c): keep the number and location of the existing knots unchanged; and then propose a Gaussian perturbation for the current values q_c , following $N(q_c, \mathbf{C}_m)$, where \mathbf{C}_m is the covariance matrix for the proposal. The remaining two components of the proposal function for the move step can be detailed:

$$\begin{cases} Q_{\text{Move}}(\mathbf{z}^*|n_c, \mathbf{z}_c, \mathbf{q}_c, n^*) = 1 \\ Q_{\text{Move}}(\mathbf{q}^*|n_c, \mathbf{z}_c, \mathbf{q}_c, n^*, \mathbf{z}^*) = Q(\mathbf{q}^*|\mathbf{q}_c) = \frac{1}{(2\pi)^{n/2} |\mathbf{C}_m|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{q}^* - \mathbf{q})^T \mathbf{C}_m^{-1}(\mathbf{q}^* - \mathbf{q})\right] \end{cases} \quad (13)$$

It is important to note that the birth step is the inverse of the death step. By incorporating Equations (9) –

(13) into Equation (8), the acceptance ratios for birth, move, and death steps are derived as Equations (14)–(16):

$$\alpha_{Birth}(\mathbf{m}^* | \mathbf{m}_c) = \min \left\{ 1, \frac{p(\mathbf{d} | \mathbf{m}^*)}{p(\mathbf{d} | \mathbf{m}_c)} \cdot \frac{1}{Q(q_b | q_p)} \cdot \frac{\lambda / (n+1)}{q_{\max} - q_{\min}} \right\} \quad (14)$$

$$\alpha_{Death}(\mathbf{m}^* | \mathbf{m}_c) = \min \left\{ 1, \frac{p(\mathbf{d} | \mathbf{m}^*)}{p(\mathbf{d} | \mathbf{m}_c)} \cdot Q(q_b | q_p) \cdot \frac{(q_{\max} - q_{\min})}{\lambda / n} \right\} \quad (15)$$

$$\alpha_{Move}(\mathbf{m}^* | \mathbf{m}_c) = \min \left\{ 1, \frac{p(\mathbf{d} | \mathbf{m}^*)}{p(\mathbf{d} | \mathbf{m}_c)} \right\}. \quad (16)$$

The Markov chain operates by selecting, with equal probability, to proceed with birth, death, or move processes at each step. It then generates a candidate sample using the corresponding proposal function outlined in Equations (11)–(13), and moves forward by accepting the candidate based on the acceptance ratios specified in Equations (14)–(16). Proper tuning of parameters such as σ_b and \mathbf{C}_m is crucial, as inadequate settings may impede convergence. For enhanced optimization of convergence speed, readers are directed to the adaptive reversible jump MCMC algorithm proposed by the authors (Tian et al., 2024b) which includes ready-to-use attached codes. After the convergence of the chain, a set of posterior samples $\{\mathbf{m}_s, s=1, \dots, S\}$, can be extracted to estimate the posterior distribution and reconstruct the earth pressure as shown in Equations (17) and (18), where $\delta(\cdot)$ is the Dirac delta function:

$$p(\mathbf{m} | \mathbf{d}) \approx \frac{1}{S} \sum_{s=1}^S \delta(\mathbf{m} - \mathbf{m}_s) \quad (17)$$

$$p(q(z) | \mathbf{d}) \approx \frac{1}{S} \sum_{s=1}^S \delta(q(z) - q_{\mathbf{m}_s}(z)) \quad (18)$$

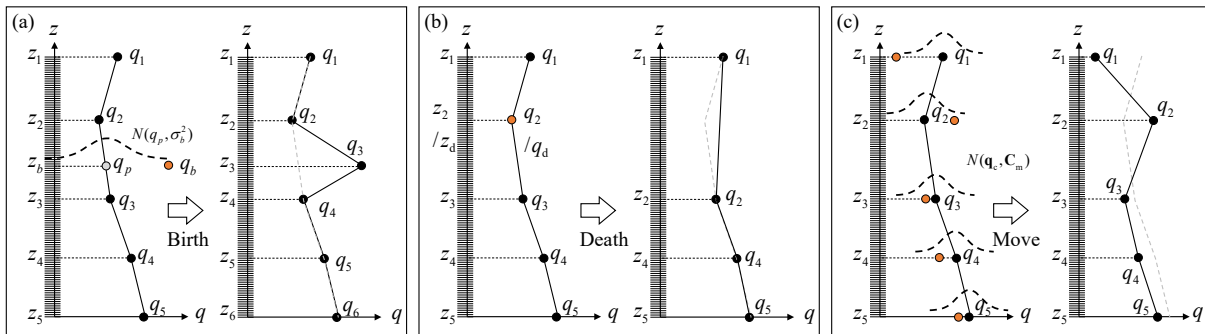


Figure 3. Designed proposal types for the trans-dimensional Markov chain: (a) birth; (b) death; (c) move

3. CASE STUDY

3.1 Preliminaries

A case study introduced by (Tian et al., 2023b) is utilized to apply and verify the proposed method. Notably, in this case, the authors pre-assumed pressure complexity, i.e., manually adding densely interpolated knots at locations prone to abrupt pressure changes. In this paper, we will demonstrate that our method is able to incorporate pressure complexity into the inversion process, eliminating the need for potentially subjective pre-assumptions. As shown in Fig. 4(a), A diaphragm wall bent towards a pit due to the active earth pressures behind it. Engineers recorded the deformations of the wall, inclusive of potential measurement errors, as shown in Fig. 4(b). Consequently, the objective herein is the inversion of earth pressures (Fig. 4c) using the deformation data.

The structural domain extends from 0 to 20 m. A dense grid of points is established across this domain, with a spacing of 10 points per meter, resulting in $N_g = 201$. The interpolation method with parameter set $\mathbf{m}, \mathbf{m}=(n, \mathbf{z}, \mathbf{q})$ is employed for parameterization. The prior for n is defined as $n \sim \text{Poisson}(21)$, where $\lambda=21$ is chosen somewhat arbitrarily, awaiting updates from observed data; The prior for \mathbf{z} is specified as $p(\mathbf{z}|n)=1/C(N_g, n)$; The prior for \mathbf{q} is set as $\mathbf{q}|n, \mathbf{z} \sim \text{Uniform}(0, 400)$. This decision is based on the assumption that the soil cannot exert any traction, and that 400 kPa is a substantially large value for the active pressures in this case. The "beam on elastic foundation" model (illustrated in Fig. 4c) is utilized to construct the forward model $g(\cdot)$, with EI of the wall being $2.5 \times 10^6 \text{ kN} \cdot \text{m}^2$ and m for the soil springs being $15 \times 10^3 \text{ kN/m}^4$. Drawing on general measurement experience,

the standard error in deformation measurement, σ_d is set as 1 mm for the likelihood function. Consequently, the prior (Equation (4)) and likelihood function (Equation (5)) have been established. The adaptive reversible jump MCMC algorithm (Tian et al., 2024b) is then employed to estimate the posterior distribution.

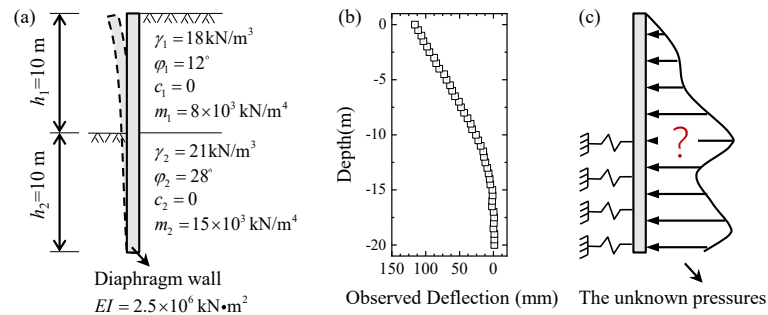


Figure 4. The case study: (a) a diaphragm wall bent towards a pit; (b) deformation data; (c) unknown pressures (Note: γ , ϕ , c , and m represent the unit weight, friction angle, cohesion, and foundation stiffness scaling factor of the soil, respectively; EI is the bending rigidity of the wall)

3.2 Results

The established Markov chain progressed through 10^6 steps. The log likelihood (LL) of the chain, together with the parameter n , is presented in Fig. 5. As shown in Fig. 5(a), LL rapidly converges from a very low initial value and stabilizes around -20. n jumps between 4 and 30 throughout the chain, indicating the chain's ability to sample parameters across various dimensions. Particularly in the second half of the chain, the behaviors of LL and n resemble white noise, suggesting convergence of the Markov chain. Consequently, samples from this second half were then used to estimate the posterior distribution of earth pressures (Equations (17)–(18)).

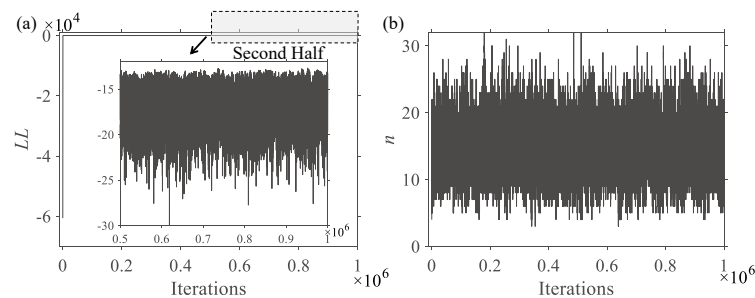


Figure 5. The established Markov chain: (a) the log likelihood (LL) along the chain; (b) n along the chain

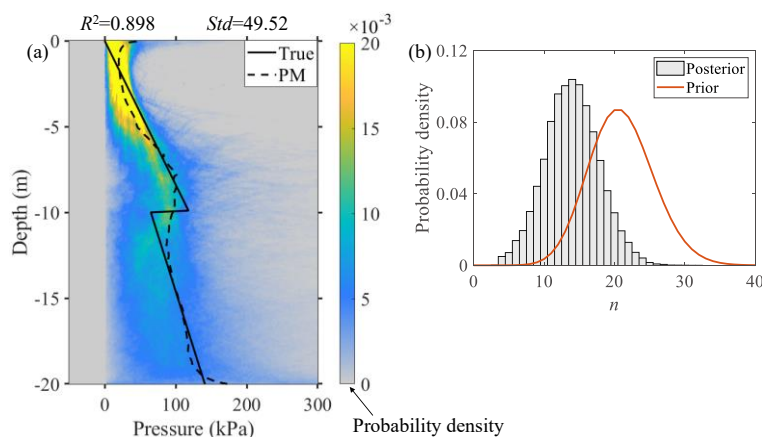


Figure 6. Inversion results: (a) the posterior distribution of earth pressures; (b) the posterior distribution of n

The inversion results are presented in Fig. 6. Fig. 6(a) illustrates the posterior distribution of earth pressures, where lighter colors signify higher probabilities, and darker colors indicate lower probabilities. The true pressures, also displayed in this figure, are predominantly confined within zones of high probability, which

underscores the effectiveness of the proposed method. In addition, the posterior mean (PM) is a representative solution of the posterior distribution. It is evident that the PM aligns closely with the actual pressures, achieving an R^2 value of 0.898, thus reaffirming the method's efficacy. Additionally, the standard deviation (Std) serves as an index to quantify the uncertainty of these results, with further details to be explored in subsequent discussions. As demonstrated in Fig. 6(b), the initial prior distribution for n centered at 21, subsequently updated to center around 15, suggests that the method can automatically infer the complexity of pressures based on observed deformation data. This advancement will be discussed in detail in the next section.

4. DISCUSSION

4.1 Advancement of the proposed method

This section will illustrate the essentiality of incorporating pressure complexity into inversion processes. It is important to reemphasize that in the discussed case study, (Tian et al. 2023b) pre-assumed pressure complexity by manually adding densely interpolated knots at locations prone to abrupt pressure changes. However, in engineering practice, pressure is unknown *a priori*, and engineers have limited information about where to insert such knots. Alternatively, one might consider a method of distributing knots evenly, and progressively densifying them as demonstrated in Figs. 7(a)–(c). With a sufficient density of knots, there may no longer be a need to presuppose pressure complexity. This strategy could potentially serve as an alternative to the method proposed herein. To validate its efficacy, comparative tests with the proposed method are conducted as follows.

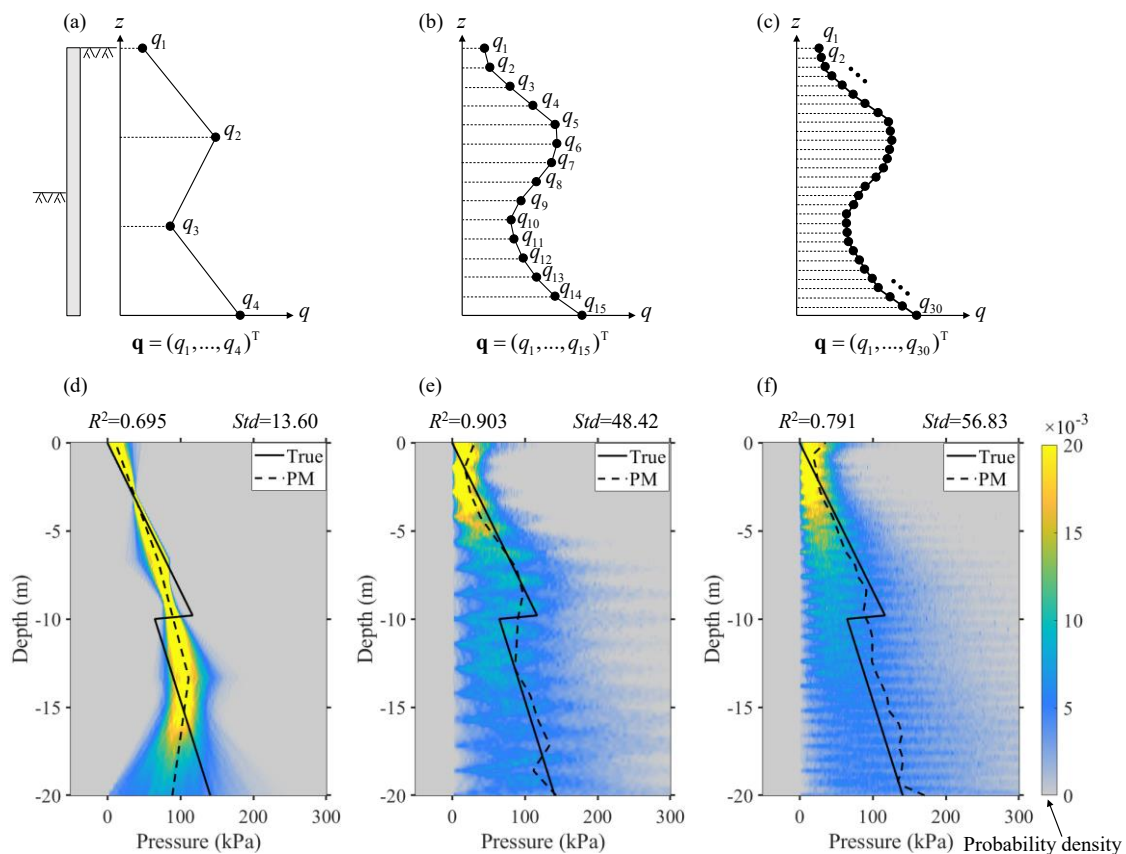


Figure 7. Parameters settings and results for the fixed-dimensional Bayesian method: (a) 4 parameters; (b) 15 parameters (c) 30 parameters; (d-f) corresponding results for 4, 15, and 30 parameters, respectively.

Additional experiments were carried out. As illustrated in Figs. 7(a)–(c), earth pressures were parameterized into a set of evenly distributed interpolated knots. The objective was the inversion of fixed parameters \mathbf{q} , where $\mathbf{q}=(q_1, \dots, q_n)^T$ with $n=4, 15$, and 30 , for the three additional cases, respectively. These experiments utilized the fixed Bayesian method as introduced by (Tian et al., 2023b). The results are presented in Figs. 7(d)–(e), and are compared with the outcomes obtained using the proposed method (Fig. 6).

As indicated in Fig. 7(d), when the number of parameters is fixed at 4, the inversion results are poor, with R^2 being 0.695; When n increases to 15, the inversion results improve with R^2 being 0.903; However, increasing n to 30 results in poorer outcomes, with R^2 dropping to 0.791. The underlying mechanism can be explained by the fact that with a small number of parameters, the pre-assumed pressure complexity is insufficient to capture the real distributed pressures. As the number of parameters increases, the method's ability to approximate actual pressure

distributions improves. However, when there are too many parameters, the available deformation data may not be informative enough to infer such a large number of unknowns, leading to increased uncertainty and deteriorated results. This is also evident in the index, Std , which grows from 13.60 to 48.42, to 56.83, illustrating an increase in the uncertainties associated with the inversion results across the three experiments. Although an R^2 value suggests that the fixed-dimensional method with 15 parameters marginally outperforms the proposed method (Fig. 6), in engineering practice, there are no "true" pressures available to fine-tune these parameters. Unlike the fixed-dimensional approach, the proposed method does not require an assumption of the pressure complexity, underscoring its advancement and applicability.

4.2 Limitation and future extensions

It is worth noting that this paper primarily illustrates the application of the proposed method to diaphragm walls. As highlighted in the Introduction, the potential extension of this method to tunnels or pipelines is of particular interest due to the pronounced structural deterioration. However, such an extension may encounter substantial technical challenges. Specifically, the inversion process necessitates numerous iterations of the trans-dimensional chain on the mechanical model, $g(\cdot)$. In this illustrative example, $g(\cdot)$ is simplified as a linear-elastic model for the diaphragm wall. However, when attempting to extend this method to tunnels or pipelines, the structural behavior may exhibit highly nonlinear characteristics, especially in the large deformation stage. Accordingly, the total number of iterations required to solve the nonlinear mechanical model via multiple Markov chain iterations may introduce an extensive computational burden. The present example provides a reference for the computational demand, with $g(\cdot)$ being linear and requiring 10^6 trans-dimensional iterations, costing about 5 hours on a desktop equipped with a Ryzen 9, 12-core, 3.8 GHz processor. As expected, when extending to nonlinear cases, the time required will multiply with the increase in iterations needed to solve the nonlinear mechanical model.

Despite these challenges, such an extension remains highly attractive. As illustrated in Fig. 8, utilizing a laser scanner to acquire deformation data (\mathbf{d}) on these structures facilitates the inversion of earth pressures ($q(z)$) with the proposed method. Upon identifying the critical factors—specifically, earth pressure—influencing structural performance, $g(\cdot)$ can be “driven” once again. This enables the execution of a real-time simulation, $\mathbf{D}=g(q(z))$, to develop a mechanical digital twin model for these in-service and poorly performing structures. Such a model aids in structural health monitoring by providing comprehensive simulation outputs, including stress and damage profiles, which warrants further research in the future.

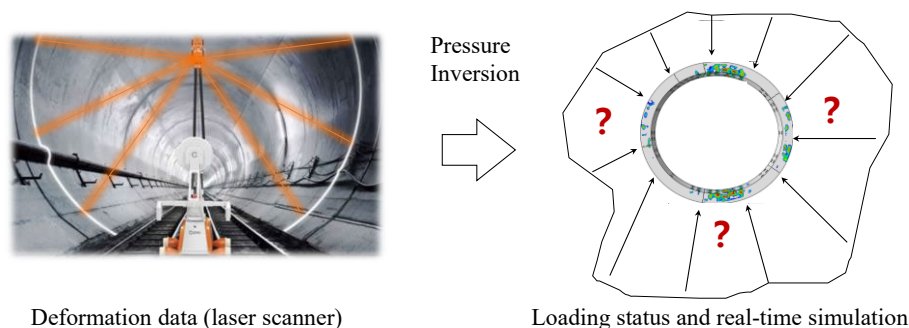


Figure 8. Expected extensions to tunnels and pipelines to assist in structural health monitoring

5. CONCLUSIONS

A trans-dimensional Bayesian load inversion method is proposed to infer earth pressures on in-service underground structures using easily observable deformation data. The distinctive feature of this method, compared to state-of-the-art approaches, is its trans-dimensional framework, which facilitates the incorporation of load complexity into the inversion process.

A case study is presented to illustrate its advancements and potential: traditional methods require a pre-assumption of pressure complexity, and inadequate assumptions can lead to poor inversion results that fail to identify true pressures. In contrast, the proposed method adaptively infers both the complexity and quantity of pressures, achieving satisfactory inversion outcomes without reliance on subjective assumptions.

This paper demonstrates the proposed method on a diaphragm wall as an example. Expanding this approach to tunnels and pipelines is of great interest and could lead to mechanical digital twins for these structures. Due to length constraints, a more detailed exploration, rigorous validation, and discussion of uncertainty will be presented in future publications.

ACKNOWLEDGMENTS

The authors acknowledge the supports from Natural Science Foundation of China (Grant No. 72404233).

REFERENCES

- Bowers, K., & Moss, N. (2018). Investigation and reconstruction of a London Underground tunnel, UK. *Proceedings of The Institution of Civil Engineers-Civil Engineering*, 171(1): 43-48.
- Carpio, F., Peña, F., & Galván, A. (2019). Recommended deformation limits for the structural design of segmental tunnels built in soft soil. *Tunnelling and Underground Space Technology*, 90, 264-276.
- Denison, D. G. T., Mallick, B. K., & Smith, A. F. M. (1998). Automatic Bayesian curve fitting. *Journal of the Royal Statistical Society*, 60(2), 333-350.
- Gioda, G., & Jurina, L. (1981). Numerical identification of soil - structure interaction pressures. *International Journal for Numerical and Analytical Methods in Geomechanics*, 5(1), 33-56.
- Gong, Q., Hui, X., & Tian, Z. (2023). Risk section classification of tunnel settlement based on land-use development simulation and uncertainty analysis. *International journal of transportation science and technology*, 12(3), 716-728.
- Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4), 711-732.
- Liu, Q., Liu, H., Huang, X., Pan, Y., Luo, C., & Sang, H. (2019). Inverse analysis approach to identify the loads on the external TBM shield surface and its application. *Rock Mechanics and Rock Engineering*, 52, 3241-3260.
- Liu, H., Liu, Q., Liu, B., Tang, X., Ma, H., Pan, Y., & Fish, J. (2021). An efficient and robust method for structural distributed load identification based on mesh superposition approach. *Mechanical Systems and Signal Processing*, 151, 107383.
- Liu, X., Bai, Y., Yuan, Y., & Mang, H. A. (2016). Experimental investigation of the ultimate bearing capacity of continuously jointed segmental tunnel linings. *Structure and Infrastructure Engineering*, 12(10), 1364-1379.
- Press, W. H. (2007). *Numerical Recipes: The Art of Scientific Computing* (3rd ed.). Cambridge University Press.
- Sanchez, J., & Benaroya, H. (2014). Review of force reconstruction techniques. *Journal of Sound and Vibration*, 333(14), 2999-3018.
- Tian, Z., Gong, Q., Di, H., Zhao, Y., & Zhou, S. (2022). What causes the excessive metro tunnel settlement in soft deposits: learned from a detailed case with factor decomposition. *Bulletin of Engineering Geology and the Environment*, 81(5), 212.
- Tian, Z., Xu, P., Gong, Q., Zhao, Y., & Zhou, S. (2023a). Health-degree model for stagger-joint-assembled shield tunnel linings based on diametral deformation in soft-soil areas. *Journal of Performance of Constructed Facilities*, 37(3), 04023019.
- Tian, Z., Zhou, S., Lee, A., Zhao, Y., & Gong, Q. (2023b). A Bayesian-based approach for inversion of earth pressures on in-service underground structures. *Acta Geotechnica*, 19(4), 1911-1928.
- Tian, Z., Zhou, S., Lee, A., Shan, Y., & Detmann, B. (2024a). How to identify earth pressures on in-service tunnel linings: Insights from Bayesian inversion to address non-uniqueness. *Transportation Geotechnics*, 48, 101344.
- Tian, Z., Lee, A., & Zhou, S. (2024b). Adaptive tempered reversible jump algorithm for Bayesian curve fitting. *Inverse Problems*, 40(4), 045024.
- Van Empel, W. H. N. C., Sip, J. W., & Haring, F. P. (2006). Design of repair measures of a damaged shield driven tunnel. *Tunnelling and Underground Space Technology*, 21(3), 338-339.
- Yan, Q., Zhang, W., Zhang, C., Chen, H., Dai, Y., & Zhou, H. (2019). Back analysis of water and earth loads on shield tunnel and structure ultimate limit state assessment: a case study. *Arabian Journal for Science and Engineering*, 44, 4839-4853.
- Yin, X., Liu, H., Chen, Y., Wang, Y., & Al-Hussein, M. (2020). A BIM-based framework for operation and maintenance of utility tunnels. *Tunnelling and Underground Space Technology*, 97, 103252.
- Zhang, D. M., Liu, Z. S., Wang, R. L., & Zhang, D. M. (2019). Influence of grouting on rehabilitation of an over-deformed operating shield tunnel lining in soft clay. *Acta Geotechnica*, 14(4), 1227-1247.
- Zhou, S., Jin, Y., Tian, Z., Zou, C., Zhao, H., & Miao, Z. (2024). Exploring the feasibility of prestressed anchor cables as an alternative to temporary support in the excavation of super-large-span tunnel. *Railway Engineering Science*, 32, 344-360.
- Zhou, S., Tian, Z., Di, H., Guo, P., & Fu, L. (2020). Investigation of a loess-mudstone landslide and the induced structural damage in a high-speed railway tunnel. *Bulletin of Engineering Geology and the Environment*, 79, 2201-2212.
- Zhou, Z., Chen, Z., He, C., & Kou, H. (2021). Investigation on the evolution characteristics and transfer mechanism of surrounding rock pressure for a hard-rock tunnel under high geo-stress: case study on the Erlang Mountain Tunnel, China. *Bulletin of Engineering Geology and the Environment*, 80(11), 8339-8361.

PROPOSAL OF THE USE OF DIFFERENTIAL RENDER IN REBAR INSPECTIONS

Yoshihiro Yasumuro¹, Youta Tanaka², Takeshi Takai³, and Satoshi Kubota⁴

- 1) D. Eng., Prof., Department of Environmental and Civil Engineering, Kansai University, Osaka, Japan. Email: yasumuro@kansai-u.ac.jp
- 2) Department of Environmental and Civil Engineering, Kansai University, Osaka, Japan.
- 3) D. Eng., Research & Development Institute, Takenaka Corporation, Chiba, Japan. Email: takai.takeshi@takenaka.co.jp
- 4) D. Eng., Prof., Department of Environmental and Civil Engineering, Kansai University, Osaka, Japan. Email: skubota@kansai-u.ac.jp

Abstract: In Japan, the number of people working in construction has been decreasing yearly since peaking in 1997 because of the declining birthrate and aging population. The Ministry of Land, Infrastructure, Transport and Tourism (MLIT) is promoting i-Construction, which aims to increase productivity by using ICT (Information Communication Technology) across the entire construction industry. In particular, the use of ICT is expected to be applied to the inspection of reinforcement rebars, which form the basis of all concrete structures, as the inspection of reinforcement bars has become stricter due to the influence of the structural calculation falsification problem that occurred in 2005. The existing methods use image triangulation to obtain a 3D point cloud of the rebar surface, so depending on the lighting conditions and the arrangement and shape of the rebars, there are some sites where it is not possible to obtain a sufficient number of point clouds, and thus the range of application is limited. This paper proposes a method for using rapidly developing differentiable rendering technology for rebar inspection. Since differential rendering makes it possible to reproduce free viewpoint images, it is possible to inspect the rebar even if the distance to the rebar surface cannot be directly measured, as long as it is possible to reproduce the appearance of the rebar as seen from a viewpoint in 3D space. Image inspection can be expected to become more versatile than before. This paper reports on the evaluation of accuracy and systematization schemes.

Keywords: Rebar Inspections, Differential Render, NeRF (Neural Radiance Fields)

1. INTRODUCTION

The influences of the declining birthrate and an aging population have been enormous in the Japanese construction industry in recent years. As shown in Figure 1, by 2023, the number of people aged 55 and over will be around 36%, and those aged 29 and under will be around 12%. Compared to all industries, the number of young people in the construction industry is decreasing, and the population is aging. The aging of the working population is inevitable, and the number of young people entering the workforce is also expected to decline, so there are further concerns about the weakening of the construction workforce. For this reason, it is necessary to promote the recruitment of young people and ensure a smooth generational shift through their retention, as well as to effectively pass on the skills, know-how, and tacit knowledge backed by the experience of skilled workers, which is an issue on construction sites, and to ensure quality. In addition, further use of ICT (information and communication technology) is attracting attention to maintain and improve productivity through improving construction technology among the limited number of workers.

The Ministry of Land, Infrastructure, Transport, and Tourism (MLIT) is promoting i-Construction, which

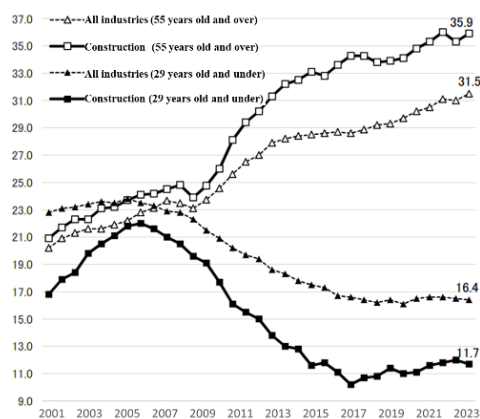


Figure 1. The aging of the construction workforce (Japan Federation of Construction Contractors (2023)) aims to improve productivity at construction sites by using ICT and other technologies in all processes, from

surveying and measurement to design, construction, and maintenance management. By consistently using 3D data, it is possible to automate the calculation of construction volumes and to understand and inspect the situation on-site using the data. Since many manual tasks, document creation, and checking are not required, labor savings and shorter construction periods can be expected. In this context, one of the items that MLIT and related companies focus on in this context is the inspection of the reinforcement of steel rebars at construction sites.

At existing construction sites, inspections of reinforcement rebars are carried out as shown in Fig. 2 (left), based on the items listed in the guidelines and other documents set out by MLIT. However, visual and manual inspections are still the norm despite the large number of inspection items. Including the time spent preparing inspection reports, the issue of a significant amount of time and effort required by many workers is raised. As a response to these issues, there are some cases where full-scale operation has been carried out since July 2023 (MLIT, 2023) to reduce the labor required for checking the reinforcement bar layout using digital data by measuring the spacing, number, diameter, and cover of the reinforcement rebars, etc., through analysis of images taken with a digital camera, and confirming the layout of the reinforcement rebars in the structure.

2. RELATED WORK

Morimoto et al. (Morimoto, 2023) developed an AI rebar placement system using images taken by a stereo camera and advanced image processing technology. They achieved implementation in the field after conducting basic measurement accuracy experiments and experiments on its application in construction sites. The service is now being provided; as of 2022, 58 units have been installed. In a technology competition held by MLIT in 2020, it received an A grade evaluation. The AI rebar inspection system can automatically measure and determine the rebar's diameter, number, and spacing simply by photographing the rebar with a stereo camera. It also has a function to display and record the automatic measurement results on the screen as shown in Fig. 2 (right).

BAIAS (Bar Arrangement Inspection AR System) is an AR rebar inspection system developed jointly by GRIFFY Corporation and Muramoto Construction Co., Ltd. It allows us to inspect the rebar in reinforced concrete structures easily. It takes photographs of the reinforcement, acquires point cloud data, and measures and identifies the steel rebars' diameter, number, and spacing. The basic principle of measuring the intervals between the reinforcement rebars is to use the LiDAR (Light Detection And Ranging) scanner function built into the iPad Pro to project laser light onto the target steel bars and then use the reflected light information to acquire the distance to the steel bars and 3D point cloud data. Like the stereo camera described above, the ranging function enables measuring the rebars' quantities and spaces. It is also possible to measure the diameter of each rebar, and it is compatible with the measurement items in MLIT's guidelines.

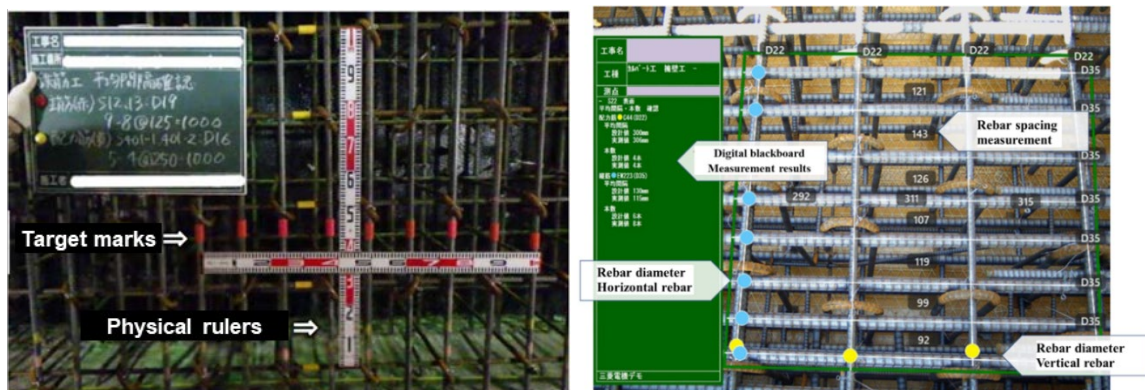


Figure 2. Conventional inspection (left) Stereo-camera based system (right) (Morimoto, 2023)

3. METHOD

3.1 Objective

As mentioned in the previous section, the construction industry, both public and private, shares the need to use ICT to reduce the labor required for rebar inspections due to the aging workforce and stricter regulations. Preceding research is being carried out into automatic rebar inspection using stereo cameras and LiDAR. However, due to the principles of these 3D data acquisition technologies, it is difficult to obtain sufficient 3D point clouds in cases where it is impossible to secure a complex structure of rebar or a shooting viewpoint, and it is challenging to meet the required data accuracy for inspection.

In this study, we will utilize a free-viewpoint image generation technology based on differential rendering called Neural Radiance Fields (NeRF) (Mildenhall, B., 2020) to enable the reproduction of the same appearance as the real thing through deep learning within NeRF, even in cases where it is challenging to acquire 3D point cloud data, such as complex structures with poor visibility or reinforcement sites where the shooting viewpoint cannot be secured. The aim is to propose an inspection method for a reproduction space that accurately reflects the

actual size of the rebar and to clarify the potential of reproduction and inspection performance.

3.2 Basic Idea

In this study, we propose introducing a completely different 3D reconstruction technology using NeRF to improve the limitations of the 3D measurement based on the principle of capturing the shape of the object surface by 3D measurement of the feature points using image triangulation. Image triangulation is not intrinsically suited to obtaining dense 3D point clouds for objects with long, thin shapes and inconspicuous colors, such as rebars. On the other hand, neural rendering, as typified by NeRF, is a technology that can reproduce the appearance of detailed scenes from arbitrary viewpoints by acquiring radiance distribution in 3D space through deep learning.

Therefore, even in cases where the scaffolding and viewpoints that can be observed on-site are limited, the shapes and arrangements of the target rebar are complex, or it is impossible to see through, it is possible to observe the rebar from a viewpoint that is virtually close, around, or inside, it is desired that the limitations of physical inspection work will be eased. The 3D volume space reconstructed using NeRF is accurate in geometry, but an unknown scale is set during the 3D reconstruction process, which does not necessarily reflect physical dimensions. Therefore, even if images are generated from a viewpoint clearly showing the target rebars, they are not directly linked to measurement inspections such as rebar diameter and intervals. In this study, we enable dimensional inspection from any viewpoint by displaying and moving the necessary measuring tape as a surface model in the space output by NeRF.

3.2 System and Process Configuration

NeRF is a technology that generates images from a 3D model from a relatively small number of images and camera positions and poses using deep learning. It generates images that allow you to view the 3D model from any viewpoint. NeRF takes as input the 5-dimensional coordinate position (x, y, z) and direction (θ, Φ) of the viewpoint in the scene, and outputs the 4-dimensional color (RGB) and transparency emitted from the viewpoint. By passing through a multilayered neural network, it learns to predict accurate output values from the input values, improving the accuracy of generating images from new viewpoints.

Hanada et al. (Hanada et al., 2023) applied NeRF to museum exhibition contents. They proposed a digital archive method and exhibition content that takes advantage of simultaneously displaying additional annotations that explain the exhibits to reflect the intentions of the museum curators within the photorealistic space rendered by NeRF. This paper proposes a method for rebar visualization by employing Nerfstudio (Tancik, M. et al., 2023), a software that can display the results of NeRF using deep learning, based on the method of Hanada et al. as shown in Figure 3. Nerfstudio uses a proprietary algorithm called nerfatto to execute NeRF, which allows NeRF to run faster than in the original paper. NeRF is executed on the backend, and the results are displayed as streaming video on the frontend web app. It is also possible to manipulate the viewpoint of the video on the web app. NeRF requires multiple images with known shooting viewpoints, prepared with the SfM (Structure from Motion) process of 3D reconstruction. Displaying a ruler model that matches the SfM's 3D coordinates' scale as additional annotation information allows a real-scale inspection of the rebars displayed by NeRF.

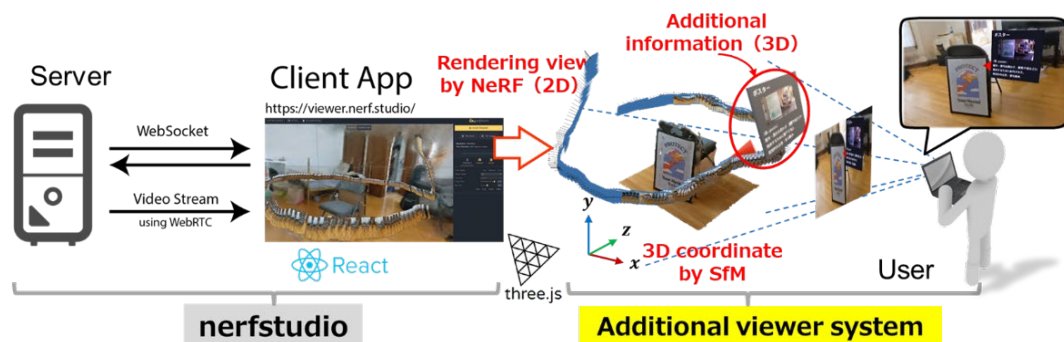


Figure 3. Process Chain of the System based on Hanada's method (Hanada et al., 2023)

3.3 Data Acquisition

The steel rebar mesh cage (800mm x 800mm x 1000mm) used in the experiment is made up of vertical and horizontal main bars and the horizontal cross rebars that support them, and the back and sides overlap the front, making it difficult to obtain a 3D point cloud from some viewpoints. A ruler was placed before the steel cage, and footage was taken using an iPhone. Repeatedly moving closer to the steel cage to capture the details and moving away to recognize the whole structure, the videos were taken slowly from the previous week to obtain overlap. The image frames were exported as 4K (2160 × 3840) resolution using Adobe Premiere Pro (Figure 4 (left)).

4. RESULTS

As shown in Figure 5, comparing the output of NeRF with the production of photogrammetry using SfM, the NeRF output could express the primary vertical rebars and horizontal rebars without any loss. In contrast, the photogrammetry output had many gaps in the horizontal thin rebars. Next, in the production results using 48 images, the NeRF output could express both vertical and horizontal rebar, as in the previous example, and the photogrammetry had a much higher number of breaks in the banded rebar than the model using 102 images (Figure 6).



Figure 4. Target sample of the rebar cage

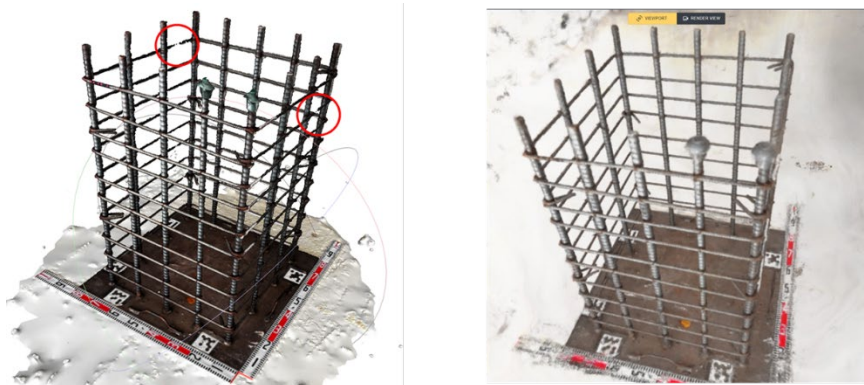


Figure 5. 3D reconstruction results from 102 images: SfM (left) and NeRF image (right)

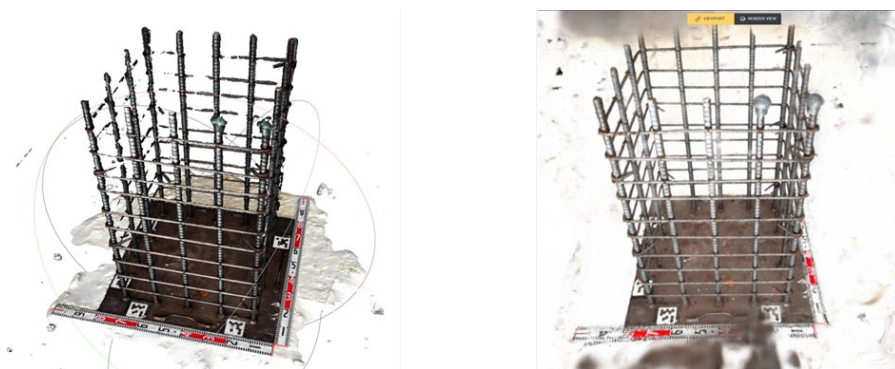


Figure 6. 3D reconstruction results from 47 images: SfM (left) and NeRF image (right)

Next, we checked the inspection accuracy of NeRF, using a set of 102 images for NeRF. The ruler model imported into the viewer was adjusted to match the actual ruler placed on the site based on the surface model output using the marching cube method (Figure 7 (left)). In addition to the 5mm grid of the actual ruler, we prepared a finer 2mm grid as a high-resolution texture to match the ruler's image generated from NeRF. The texture is automatically mapped to the surface model using UV mapping and appropriate scaling to be used universally for NeRF output. This virtual ruler made it possible to display the measurement resolution required for inspection,



Figure 7. Virtual ruler setup: scale adjustment (left) and inspection setup (right)

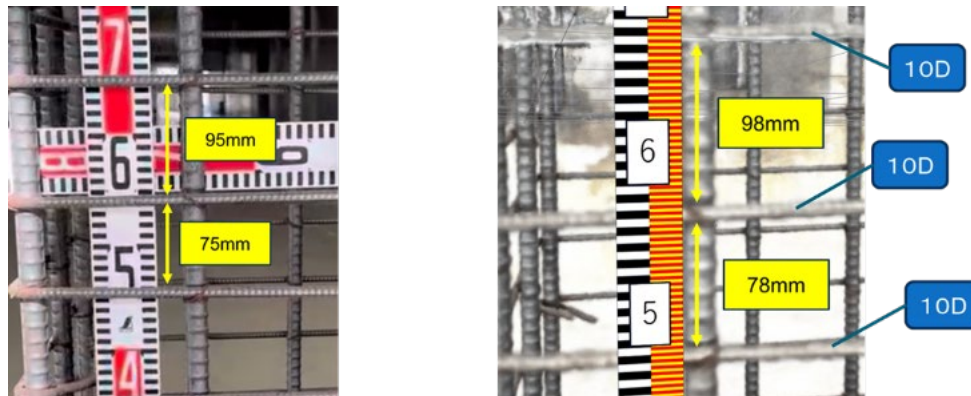


Figure 8. Images for inspection: real photo (left) and NeRF image (right)

and by simply superimposing it on the free viewpoint in the NeRF image, it became easy to read the rebar diameter and laying interval. In the same way, the scale was moved (Figure 7 (right)) as in the image where the scale was placed.

In the photo of the actual object in Figure 8 (left), the upper row was 95mm, and the lower row was 75mm, but in the viewer image, the upper row was 95mm, and the lower row was 75mm, with an error of +3mm as shown in in Figure 8 (right). The diameter of the deformed bar was also measured from the viewer image, and was 10mm, so it can be determined that it is a 10D rebar.

4. DISCUSSION

Since the construction site's structure often limits the areas that can be photographed, we applied the method of this study to a group of images photographed from only one direction, as shown in Figure 9 (left). In this case, photogrammetry added noise that was not there because the back surface was assimilated with the ground or wallpaper. In addition, the side surface needed to be sufficiently expressed and was interrupted. On the other hand, the output of NeRF can be adequately expressed not only on the front but also on the back and side surfaces. As in the previous section, the position and size of the scale were adjusted (Figure 36). This effect allows the scale to be adjusted to a position that cannot be physically placed and allows inspection to be performed from a viewpoint that cannot be photographed for structural reasons.

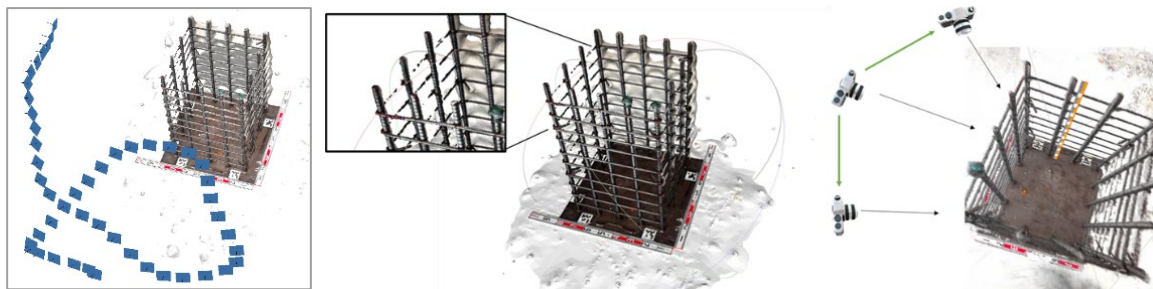


Figure 9. Limited angle of view for the input photos from 50 viewpoints (left), the result of SfM (middle) and NeRF (right)

5. CONCLUSIONS

In this paper, the output of photogrammetry and NeRF was compared by changing the number of images input for a steel-reinforced cage, and the accuracy of the rebar inspection was confirmed by adjusting the scale for the output of NeRF. In addition, assuming a limited shooting space at an actual construction site, this study was also applied to images taken from only one direction. Firstly, while NeRF could produce photorealistic representations even with a limited number of images and shooting angles, photogrammetry showed some deficiencies. From this, it is a visualization method that is relatively easy to use for on-site construction workers. In addition, in the accuracy check for the rebar inspection, the reproduction error in the rebar spacing from the actual object was within +3mm, and it was also possible to determine that the rebar diameter was the same type as the actual object. From this, it was possible to generate inspection images with the same level of accuracy as previous research that measured point clouds. In future work, the authors plan to develop a system that allows intuitive alignment of the ruler model and to reduce the time required for deep learning, which currently takes several tens of minutes.

ACKNOWLEDGMENTS

The authors would like to thank Takenaka Corporation for their budget, valuable human resources, and information, and also gratefully acknowledge the work of past and present members of our laboratory.

REFERENCES

- Japan Federation of Construction Contractors (2024), *Digital Construction Handbook* Retrieved from website: <https://www.nikkenren.com/publication/handbook/chart6-4/index.html>
- The Ministry of Land, Infrastructure, Transport, and Tourism (MLIT), (2022). *i-Construction Promotion Consortium 9th Planning Committee, Document 2p22*, December 8, 2023 (in Japanese) Retrieved from website: https://www.mlit.go.jp/tec/i-construction/pdf/02.9_kikaku_siryoku2.pdf
- Morimoto, N., Taira, Y., Hirose, M., and Hisashiba, T., (2023), Further upgrading of bar arrangement inspection system with AI technology, *Japan Society of Civil Engineers 2023 Annual Meeting*, ID:VI-120. (in Japanese)
- Mildenhall, B., Srinivasan, P., Tancik, M., Barron, J., Ramamoorthi, R., and Ng, R. (2020). NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. *Computer Vision – ECCV 2020: 16th European Conference, Proceedings, Part I*. pp. 405–421.
- Hanada, T., Yasumuro, Y., Sumida N., Muraoka, T., Suemori, K. (2023), Proposal for the Use of NeRF in Cultural Heritage Exhibitions, *The 40th Japan Society for Scientific on Cultural Property Congress*, pp.200-201. (in Japanese)
- Tancik, M., Weber, E., Ng, E., Li, R., Yi, B., Wang, T., Kristoffersen, A., Austin, L., Salahi, K., Ahuja, A., Mcallister, D., Kerr, L., and Kanazawa, A., (2023). Nerfstudio: A Modular Framework for Neural Radiance Field Development. In *ACM SIGGRAPH 2023 Conference Proceedings (SIGGRAPH '23)*. Association for Computing Machinery, Article 72, pp. 1–12.

Machine Learning-Assisted Prediction of Steam-Cured Concrete for 3D Printing Applications

Junbo Sun¹, Yufei Wang², Xianda Liu³, Jiajie Shang², Wenfu Zhang⁴ and Xiangyu Wang^{5*}

- 1) Ph.D. Senior researcher, School of Design and Built Environment, Curtin University, Perth, WA 6102, Australia. Email: tunneltc@gmail.com
- 2) Ph.D., Candidate, School of Design and Built Environment, Curtin University, Perth, WA 6102, Australia. Email: wangyf0113_suz@163.com; 1284895576@qq.com
- 3) Master, Candidate, School of Civil Engineering and Architecture, East China Jiao Tong University, Nanchang 330013, China. Email: 18879671461@163.com
- 4) Master, Candidate, School of Civil Engineering and Architecture, East China Jiao Tong University, Nanchang 330013, China. Email: name3@lmn-u.ac.jp
- 5) Ph.D., Professor, School of Architectural Engineering, Nanjing Institute of Technology, Nanjing, China. Email: zhang_wenfu@126.com
- 6) Ph.D., Professor, School of Civil Engineering and Architecture, East China Jiao Tong University, Nanchang 330013, China. Email: 8686@ecjtu.edu.cn

Abstract: Research into 3D concrete printing (3DCP) has attracted considerable interest from both academia and industry. However, comprehensive investigations into the effects of curing conditions on the mechanical properties of 3DCP remain limited. This study investigates the effects of steam curing parameters—including temperature rise rate, retention capacity, and sustained temperature—on the mechanical performance of 3D printed concrete across different curing ages. The optimal steam curing conditions for enhancing the compressive properties of 3D printed concrete in various directions were identified through the integration of macroscopic and microscopic analyses. Furthermore, experimental data were utilized to develop machine learning models, optimized with the beetle antennae search (BAS) algorithm, ensuring high predictive accuracy for the mechanical performance of steam-cured concrete. This study contributes to advancing both theoretical insights and practical applications regarding steam curing in enhancing the mechanical properties of printed components.

Keywords: 3D printing concrete; steam curing; anisotropy; compressive strength; machine learning; beetle antennae search.

1 Introduction

3D concrete printing (3DCP) is gaining global traction for its formless builds, design freedom, less waste, and quicker construction [1, 2]. Its ability to utilize recycled materials like construction waste makes it eco-friendly, reducing waste and dust compared to traditional methods [3]. By incorporating industrial by-products such as fly ash (FA), silica fume (SF), and ground granulated blast furnace slag (GGBS), 3DCP enhances sustainability, mitigates shrinkage, and improves printing quality, offsetting the high disposal costs associated with these materials [4-7].

Steam curing, a method where the concrete's surface directly contacts hot, moist air, requires heat to conduct through the mold and into the concrete for uniform heating. This process leads to higher surface temperatures in steam-cured concrete, increasing surface porosity and the presence of harmful pores compared to the interior. Surface porosity in steam-cured concrete is higher than internal porosity but can be controlled by using a geotextile cover. Furthermore, higher treatment temperatures widen the temperature gradient between the concrete's surface and interior, resulting in lower compactness and higher permeability. However, few researchers have been conducted on 3DCP.

Machine learning (ML) models, which have gained significant attention in recent years, hold considerable potential in predicting material characteristics [8]. Their ability to analyze unstructured data from datasets and generate precise, reliable predictions underpins their value [9, 10]. Among the most prominent ML models in the concrete industry are Artificial Neural Network (ANN), Support Vector Regression (SVR), and Random Forest (RF), noted for their widespread application [11]. This research introduces various ML methods to forecast the performance of 3D printed concrete under steam curing conditions.

2 Method

2.1 Materials and test setup

The materials include Portland cement (42.5R), FA (Class F), GGBS, SF, and quartz sand with particle sizes from 40 to 70 mesh. Quartz sand, preferred for its simplicity and low impurity content, enhances printability. A polycarboxylic acid type water-reducing agent is used to improve print quality. The cellulose in this study is hydroxypropyl methyl cellulose with a viscosity of 4W, and the fiber is polyethylene fiber. As illustrated in Figure 1, a three-dimensional coordinate system with perpendicular X, Y, and Z axes is established to analyze the load orientation and anisotropy of the 3D printing material, with each axis perpendicular to the central point of the printing layer cross-section, the outline surface, and the plane cross-section, respectively.

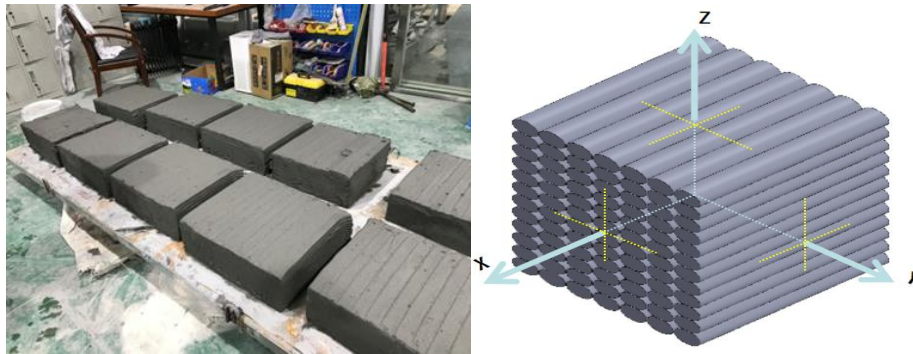


Figure 1. Coordinate system

For 3D printed concrete to meet fluidity, extrudability, and support requirements, the aggregate distribution and additive proportions are critical. This experiment utilized a 0.36 water to binder ratio for fluidity and support, and a sand to binder ratio of 1.3 for extrudability and cost efficiency. Key admixtures included fiber, cellulose, and water reducer. According to Table 1, the proportions of FA, SF, and GGBS in cementitious materials were 20wt%, 15wt%, and 10wt%, respectively, assuming a total binder content of 100wt%. PVA fiber, hydroxypropyl methylcellulose, and water reducer were added at 0.21wt%, 0.016wt%, and 0.23wt%, respectively, to improve flowability and water retention. The mix ratio was established through research. The printing nozzle had a rectangular section of 38mm by 14mm, moving at 12cm/s. Each layer was printed using 12 round-trip routes to create a 20-layer vertical sedimentary structure, with a printing path interval of 40mm and a layer height of 13mm. Cube samples with 100mm sides were mechanically cut for testing after 8 hours of curing.

Table 1. Control mix proportion of this test.

Components (wt.%)								
Cement	FA	SF	GGBS	Sand	Water	Water reducer	PVA fibre	Hydroxypropyl methylcellulose
56	9	15	20	130	36	0.23	0.21	0.016

The goal of the orthogonal test was to analyze the impact of steam curing conditions on the mechanical strength of 3D printed concrete, adhering to Chinese standards [12] and previous research [13], Table 2 shows an orthogonal experiment with variables including temperature rise rate (e), sustained temperature time (E), and sustained temperature (f) across three levels. Additionally, 18 experimental groups were incorporated for the training of machine learning (ML) models. The study utilized range analysis for its simplicity and clarity. This method identifies the dominant factors affecting the outcome by calculating the range values from the averages of the highest and lowest levels. The chosen resting time of 8 hours aligns with the initial setting phase of conventional concrete, which acquires enough structural strength in 4-7 hours to withstand internal phase expansion. A cooling rate fixed at 10°C/h and post-curing conditions maintained at 20°C ± 5 with 70% ± 10 humidity were selected to prevent damage to the print transition zone.

Table 2. The curing conditions of samples in this test.

Test group	Curing conditions		
	e (°C/h)	E (h)	f (°C)
1	10	6	50
2	10	8	70
3	10	10	60
4	15	6	70
5	15	8	60
6	15	10	50
7	20	6	60
8	20	8	50
9	20	10	70
C _c	Curing in natural condition		
C _p			

2.2 Machine learning

The Beetle Antennae Search (BAS) method is utilized for hyperparameter optimization in machine learning models like back propagation neural network (BPNN) and Random Forest (RF) to fine-tune settings automatically for peak performance across various problems [14, 15]. Inspired by the behavior of the longhorn beetle, which moves in the direction of its longer antenna to find food, adjusting its course based on the sensory input, the BAS algorithm mimics this by exploring in any given dimension. In this context, x_l and x_r represent different orientations of the left and right antennas respectively, and the following equation represents the position of the i^{th} time:

$$\begin{aligned} x_l^i &= x^i + d^i b \\ x_r^i &= x^i - d^i b \end{aligned} \quad (1)$$

where b is a random vector, which is determined by the equation Eq (2).

$$b = \frac{\text{rand}(k, 1)}{\|\text{rand}(k, 1)\|} \quad (2)$$

where $\text{rand}(k, 1)$ denotes a random function that generates random vectors in a specified dimension, and k is the dimensionality of the search space. The next three equations depict the position vector of the beetle and the step size after the antenna length is updated.

$$x^i = x^{i-1} + \delta^i b \text{sign}(f(x_r^i) - f(x_l^i)) \quad (3)$$

$$d^i = 0.95d^{i-1} + 0.01 \quad (4)$$

$$\delta^i = 0.95\delta^{i-1} \quad (5)$$

where $\text{sign}()$ is the sign function, δ is the step length, $f(x)$ is the fitness function.

To evaluate ML model performance, the Root Mean Square Error (RMSE) and correlation coefficient (R) are used.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i^* - y_i)^2} \quad (6)$$

$$R = \frac{\sum_{i=1}^n (y_i^* - \bar{y}^*)(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (y_i^* - \bar{y}^*)^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (7)$$

where n is the number of observations, y_i^* is the anticipated output, and y_i is the final output, and the average of the predicted and the average of the actual results are expressed in \bar{y}^* and \bar{y} respectively.

3 Results and analysis

Figure 2 shows the variation in 28-day compressive strengths across different directions, indicating the end of the hydration phase for the specimen. The impact of varying steam curing conditions on compressive properties becomes evident under consistent material and printing parameters [16]. Orthogonal experiments and subsequent calculations provide average values for compressive strengths in the X (CS_x), Y (CS_y), and Z (CS_z) directions. The influence of steam curing conditions on CS_x is ranked as: temperature rise rate (6 MPa) > sustained temperature (4 MPa) > sustained temperature time (2 MPa). For CS_y , the order of influence is: sustained temperature time (5 MPa) > temperature rise rate (4 MPa) > sustained temperature (2 MPa). The impact on CS_z follows a similar trend: sustained temperature time (5 MPa) > temperature rise rate (3 MPa) > sustained temperature (2 MPa).

Figure 2(a) shows CS_x 's negative correlation with temperature rise rate, positive correlation with sustained temperature time, and a parabolic correlation with sustained temperature. Figure 2(b) illustrates CS_y 's negative correlation with temperature rise rate and parabolic relationships with both sustained temperature time and sustained temperature. Figure 2(c) indicates that CS_z has a parabolic correlation with all three curing conditions. The rapid heating and high sustained temperatures lead to volume expansion of free water and air within the specimen, causing micro-cracks and affecting mechanical properties post-steaming. Optimal conditions for late-stage curing in the X and Y directions include a 10°C/h temperature rise rate, 10h of sustained temperature time, and a sustained temperature of 70°C. For CS_z , the best parameters are a 20°C/h temperature rise rate, 8h sustained temperature time, and 60°C sustained temperature. The negative correlation of CS_x and CS_y with the temperature rise rate may be due to rapid curing creating a nonuniform pore distribution and a coarser pore structure, as rapid hydration does not allow even spread of hydration products [17]. With increasing maintenance temperature, the hydration rate of cement and the calcium-silica ratio of C-S-H, its main hydration product, increase, leading to C-S-H crystallization, reduced specific surface area, decreased bonding among hydration products, and increased brittleness. This results in a parabolic correlation between CS_z and the temperature rise rate [18, 19].

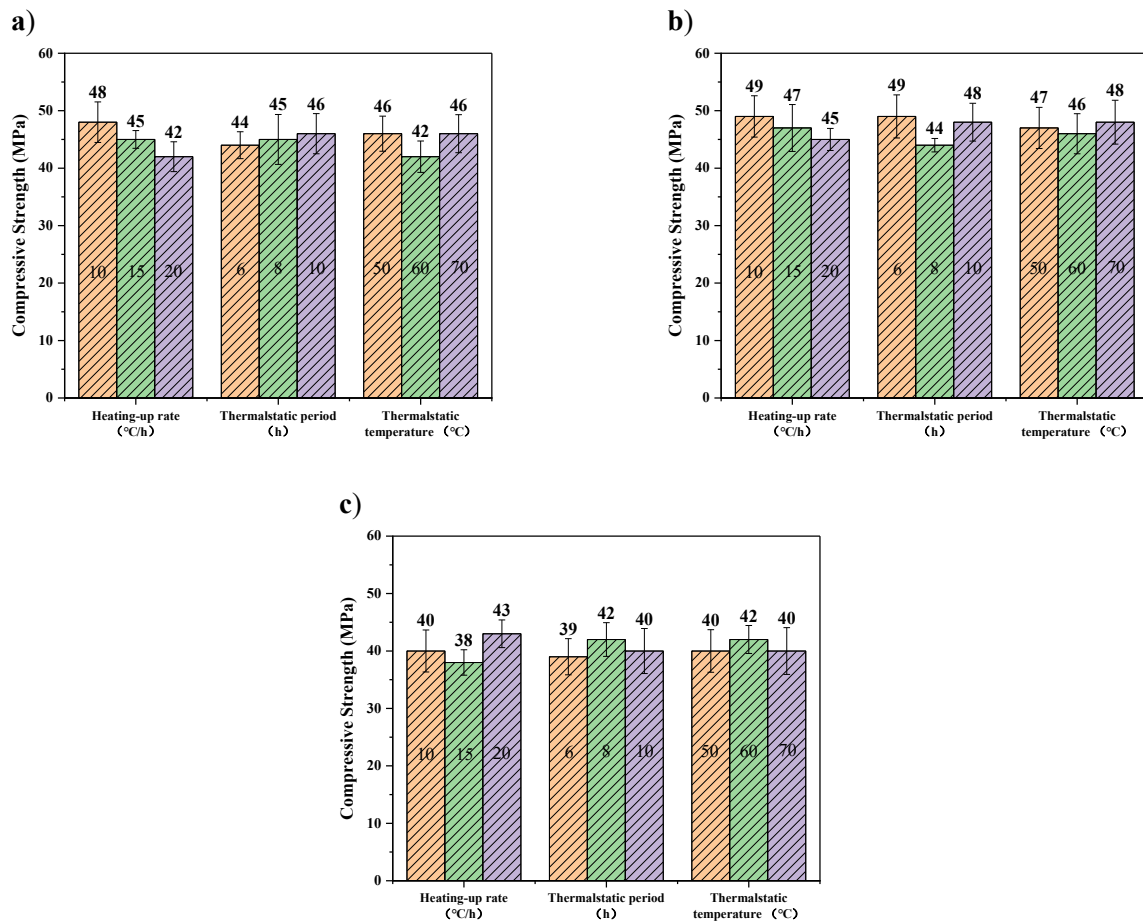


Figure 2. Effect of curing conditions on CS_x , CS_y , CS_z (sustained temperatures at 28 days).

A 30% test set was used to evaluate the performance of the RF model, which was optimized using the BAS method. Figures 3a and 3b depict the actual vs. predicted values of CS and strain, respectively, showcasing the model's accuracy on the dataset. The minimal discrepancy between the predicted and actual values demonstrates the precision of the BPNN in estimating UCS and strain. Further, a scatterplot illustrating both actual and estimated CS values across the training and test sets, along with their correlation and RMSE, is presented in Figure 4. The data for CS prediction showed a prediction value of 0.9565, a dataset correlation coefficient of 0.9108, and RMSE values of 2.0621 and 2.85. This indicates that the RF model's prediction of CS is both accurate and reliable, evidenced by a high correlation coefficient and low RMSE. Moreover, the consistency between the training and test set evaluations suggests that the model fits the data well, meeting the necessary accuracy requirements.

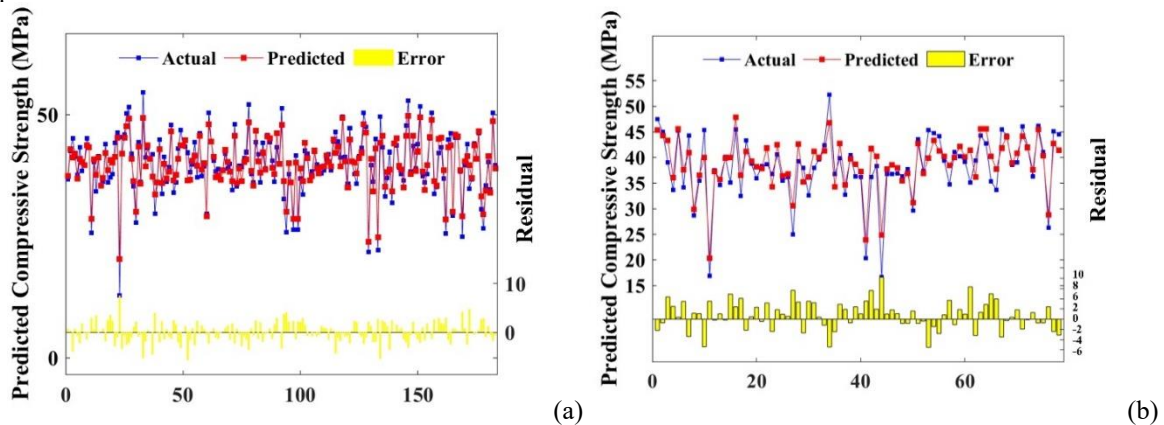


Figure 3. Scatter plot of simulated and actual CS for (a) training and (b) test sets

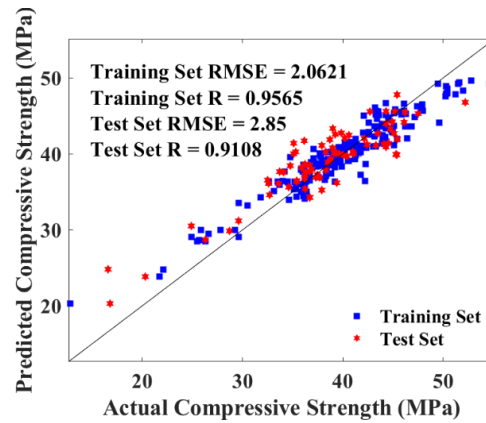


Figure 4. Scatter plot of simulated and actual CS for training and test sets

The K-fold cross-validation (with $k=5$) was performed to ensure that the model's performance was not overly dependent on a specific data split. The results demonstrated consistent RMSE and R-values across folds, highlighting the model's ability to generalize well to unseen data. In addition to cross-validation, key performance metrics, specifically RMSE and R-values, were analyzed on both the training and testing datasets. The minimal discrepancy between training and testing metrics suggests that the models were neither underfit nor overfit, further supporting their reliability.

Figure 5(a) displays a box plot comparing the discrepancy between predicted and actual values of CS by the BPNN and RF models. The RF model's interquartile range is narrower (between the upper and lower black lines), and it has a lower median (indicated by the red line), signifying a smaller prediction error. In terms of CS prediction, the maximum error limit for RF is lower than that for BPNN, showcasing the superior accuracy of the RF model over the BPNN. In addition, Figure 5(b) utilizes a Taylor chart to consolidate three critical model evaluation metrics—maximum correlation coefficient, minimum RMSE, and minimum standard deviation—into polar coordinates. For both CS prediction and overall performance, the RF model plots closer to the "ideal" position when compared to the BPNN model, further evidencing the enhanced performance and accuracy of RF in predicting CS.

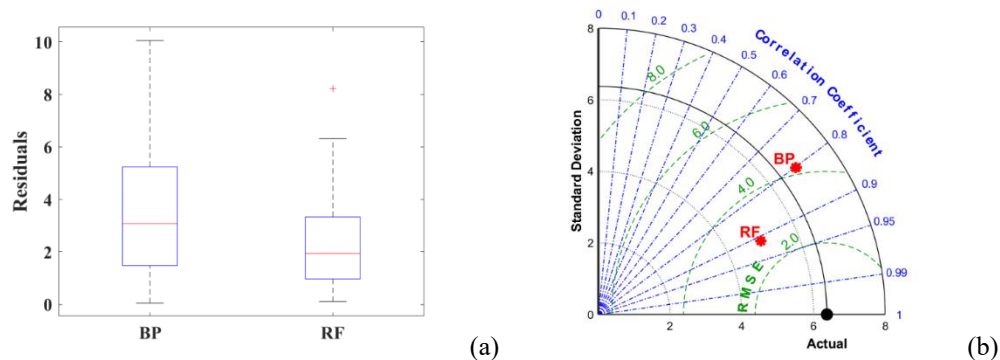


Figure 5. Performance evaluation of four models by CS prediction (a) boxplot and (b) Taylor

4 Conclusion

The study highlighted how different curing parameters influence the mechanical anisotropy of printed concrete, as evidenced by uniaxial compression test data. An accurate BAS-RF prediction model was developed to optimize maintenance parameters for various engineering demands. Key findings include:

1. Steam curing markedly improves the early strength of concrete, necessitating a balance between early strength growth and ultimate strength in the selection of curing conditions.
 2. A temperature rise rate of 10°C/h and a sustained temperature of 50°C are advised for late strength growth. The duration of thermal-static conditions appears to have a minimal impact on strength development.
 3. The BAS method effectively identified the hyperparameters for the ML model. With an R value of 0.9108 for CS, the BAS-RF model proved to be highly accurate.
 4. Box and Taylor plot comparisons underscore the RF model's superiority over BPNN models, evidenced by higher R values and lower RMSE, showcasing its better prediction capabilities and generalizability.
- Based on this research, the research limitations and future research are identified:
1. This research focuses on the impact of curing conditions on 3D-printed concrete using a single material mix proportion. The effect of various material mix designs under curing environments is not considered.

Therefore, comprehensive further research is required to evaluate the mechanical performance of 3D-printed concrete with diverse material mix proportions across different curing conditions.

2. This study primarily focuses on the compressive strength anisotropy of 3D-printed concrete under varying curing conditions. Other critical mechanical properties, such as flexural strength, remain unexplored. Therefore, future research should aim to provide a comprehensive assessment of the mechanical performance of 3D-printed concrete across different curing conditions.

References

- Sun, J., et al., (2021). Multi-objective optimization of a graphite-slag conductive composite applying a BAS-SVR based model. *Journal of Building Engineering*. **44**: p. 103223.
- Zhao, T., et al., (2020). Soil moisture experiment in the Luan River supporting new satellite mission opportunities. **240**: p. 111680. doi: 10.1016/j.rse.2020.111680.
- Sun, J., et al., (2021). Electromagnetic absorption of copper fiber oriented composite using 3D printing. *Construction and Building Materials*. **300**: p. 124026.
- Zhang, J., et al., (2020). Machine-learning-assisted shear strength prediction of reinforced concrete beams with and without stirrups. *Engineering with Computers*: p. 1-15.
- Sun, J., et al., (2022). Mechanical performance prediction for sustainable high-strength concrete using bio-inspired neural network. *Buildings*. **12**(1): p. 65.
- Asprone, D., et al., (2018). 3D printing of reinforced concrete elements: Technology and design approach. *Construction and Building Materials*. **165**: p. 218-231.
- Sun, J., et al., (2022). Molecular interfacial properties and engineering performance of conductive fillers in cementitious composites. *Journal of Materials Research and Technology*.
- Qi, B., P. Xu, and C.J.W. Wu, (2023). Analysis of the Infiltration and Water Storage Performance of Recycled Brick Mix Aggregates in Sponge City Construction. **15**(2): p. 363. doi: 10.3390/w15020363.
- Han, L., L. Wang, and W. Zhang. *Quantification of statistical uncertainties of rock strength parameters using Bayesian-based Markov Chain Monte Carlo method*. in *IOP Conference Series: Earth and Environmental Science*. 2020. IOP Publishing.
- Zhang, H., et al., (2022). Numerical study on welding residual stress distribution of corrugated steel webs. **12**(11): p. 1831. doi: 10.3390/met12111831.
- Dong, Z., et al., (2023). Asymptotic homogenization of effective thermal-elastic properties of concrete considering its three-dimensional mesostructure. **279**: p. 106970. doi: 10.1016/j.compstruc.2022.106970.
- GB/T 51231-2016 Technical standard for assembled buildings with concrete structure.
- Shi, J., et al., (2020). Properties evolution of high-early-strength cement paste and interfacial transition zone during steam curing process. *Construction and Building Materials*. **252**: p. 119095.
- Schapire, R.E., *The Boosting Approach to Machine Learning: An Overview*, in *Nonlinear Estimation and Classification*. 2003, Springer New York: New York, NY. p. 149-171.
- Wang, J. and H. Chen, (2018). BSAS: Beetle swarm antennae search algorithm for optimization problems. *arXiv preprint arXiv:1807.10470*.
- Odler, I. and M. Gasser, (1988). Mechanism of Sulfate Expansion in Hydrated Portland Cement. *Journal of the American Ceramic Society*. **71**(11): p. 1015-1020.
- Igarashi, S.-i., A. Watanabe, and M. Kawamura, (2005). Evaluation of capillary pore size characteristics in high-strength concrete at early ages. *Cement and Concrete Research*. **35**(3): p. 513-519.
- Girão, A.V., et al., (2007). Composition, morphology and nanostructure of C–S–H in white Portland cement pastes hydrated at 55 °C. *Cement and Concrete Research*. **37**(12): p. 1571-1582.
- Lachowski, E.E. *CRYSTALLINITY IN CSH GELS: INFLUENCE OF PREPARATION AND CURE CONDITIONS* EE LACHOWSKI, SY HONG AND FP GLASSER. in *PRO 13: 2nd International RILEM Symposium on Hydration and Setting-Why does cement set? An interdisciplinary approach*. 2000. RILEM Publications.

THE GREAT MISALIGNMENT IN CONSTRUCTION CONTRACTING: BEST PRACTICES AND SOFTWARE

Matt Stevens PhD¹, Wei Zhou², and Babatunde Fatai Ogunbayo³

- 1) Ph.D. School of Engineering, Design and Built Environment, Western Sydney University. Email: matt.stevens@westernsydney.edu.au
- 2) Ph.D. School of Engineering, Design and Built Environment, Western Sydney University. Email: w.zhou@westernsydney.edu.au
- 3) Ph.D. Faculty of Engineering and the Built Environment, University of Johannesburg. Email: babatundeo@uj.ac.za

Abstract: Construction software's promise of efficacy has been significantly unfulfilled. Although Information and Communication Technology (ICT) has been commonly used for over thirty years, many problems have existed relatively unchanged. There are several reasons for this, but one is overlooked: the absence of contractor best practices embedded in the industry's software. The primary cause appears to be developers' commercial interest. This paper used a literature review and an industry survey to study possible contributing factors. This paper quantifies this lack of software capability to support valuable processes. Currently, surveyed professionals perceive over 40% difference between the value of the practices tested and the enabling ability of the respondent's software. The industry is in a new age with the emergence of Artificial Intelligence (AI) and Quantum Computing. Will the same problems exist after these advanced technologies are widely adopted? This research adds to the body of knowledge by articulating a gap between industry-accepted practices' value and commercial software support that enables their execution. The paper asserts that government, associations, researchers and construction organisations should facilitate reengineering to existing ICT to reflect contractor processes. This will make the coming transformation of technology more valuable to all stakeholders.

Keywords: Construction Contracting, Technology Misalignment, Construction Software, ICT Efficacy, Practice Alignment

1. INTRODUCTION

Although Information and Communication Technology (ICT) has been increasingly used for over thirty years, many construction industry problems have continuously existed during that time, e.g., unsatisfactory safety performance, poor quality outcomes, cost blowouts, and schedule delays. Researchers have distilled several reasons for this, but one is significantly overlooked: the absence of contractor best practice-enabling modules in the industry's commercial software.

The construction industry globally has been relatively slow to adopt Information and Communication Technology (ICT). Stevens and Smolders (2023) have pointed out more than a dozen barriers to innovation, including unique business characteristics with clients and vendors. Andújar-Montoya et al. (2020) echo this: construction is more complicated than any other sector due to one-of-a-kind production. The industry has many ad-hoc production control methods, most of which are informal due to approximately 98% of organisations being small and medium enterprises with 20 employees or less (ABS, 2022). This lack of structured thinking and installation process creates uncertainty that prevents smooth production flow (Dave et al., 2016).

ICT is intended to streamline contract processes, enhance compliance, and foster better alignment across project stakeholders (Bamgbose et al., 2024). However, the extent to which these tools effectively address the root causes of misalignment remains largely underexplored. Moreover, academics and software vendors have provided many reasons for the low implementation of technology solutions in the industry, such as high acquisition cost, ongoing investment, and mastery time it requires. However, this paper asserts one more reason: construction software does not robustly support contractors' business and project practices. One indirect effect is that software firms do not promote and educate best practices via their customer training and sales presentations.

There may be a perverse incentive and, thus, a misaligned relationship. Contractors suffer from the highest bankruptcy rate in most countries, and software developers cannot afford to lose 1/3rd of their income stream. The authors' industry experience is that construction software firms have three common income sources: 1) Licensing, 2) Training, and 3) Custom Programming. This third revenue source would be reduced if their product supported best practices, and thus special modules would need to be created. The result appears to be reduced perceived value (total cost of ICT versus profit enhancement over spreadsheets) for contractors.

2. LITERATURE REVIEW

Despite its critical role in global economic activity, the construction industry faces significant management challenges (Őri & Szabó, 2024). Andújar-Montoya et al. (2020) state that projects produce significant information that must be processed and stored with efficacy, creating a ready market for software developers. The

core challenges include the pervasive misunderstanding between contractual terms, objectives, and stakeholder expectations at the project level. This phenomenon exacerbates disputes, delays, cost overruns, and inefficiencies in delivery (Solli-Sæther et al., 2015). Best practices in construction contracting have been proposed to mitigate misalignment, emphasising transparent communication, equitable risk allocation, and the adoption of standard contract forms (Ogunbayo et al., 2024). However, despite these efforts, misalignment persists, often rooted in inconsistencies in contract interpretation, inadequate stakeholder engagement, and insufficient flexibility to adapt to changing project conditions (Sande & Haugland, 2015). Digital tools and contract management software advancements have introduced promising solutions to bridge these gaps (Ye et al., 2022).

Construction is characterised by factors that affect schedule, resource allocation and quality, including single-digit percentage margins, 2% or less market share of the industry leaders, fluid workforce dynamics (highest “lost last job” of all Australian sectors) and a high percentage bankruptcy. It is considered a complicated and information-based industry primarily based on the accuracy of interpreting varied issues based on a project's unique inputs (Wu et al., 2012). Prior studies have identified the construction sector's various working methods as the main barrier to implementing an industry improvement (Dubois & Gadde, 2002). Hartmann et al. (2012) noted that existing project management best practices guide understanding and supporting BIM implementations at the operational level of an organisation. This group of researchers have advocated for tailoring software to support these best practices for over a decade.

Table 1. The Business Characteristics and Risks of Good-Producing Industries in Australia

Industry	Does Not Own Product Until Sold	Not Fixed Asset Based	Product Design Created by Each Client	Highest Bankruptcy Rate	Most Significant Insolvency Deficiency \$ Average	Highest % “Lost Last Job” Employee Rate	Highest % of Contract Employees
Agriculture, Forestry and Fishing	No	No	No	No	No	No	No
Construction	Yes	Yes	Yes	Yes*	Yes*	Yes*	Yes*
Manufacturing	No	No	No	No	No	No	No
Mining, Oil & Gas	No	No	No	No	No	No	No

*ABS 2023 comparison to all (16) Australian Industries

Constructing a project is a process dependent on hundreds of ordered tasks. The products installed are mostly non-exclusive to competitors. Systematising operations is critical to lowering costs, elevating quality, consistently meeting schedule dates, and lessening employee frustration (Kalsaas and Sacks 2011).

First, some perspective is needed to frame the dynamics properly. Construction is considered a goods-producing industry, not a service, as defined by the U.S. Bureau of Labor Statistics (n.d), along with agriculture, forestry, fishing, hunting, mining, oil, gas, and manufacturing. Interestingly, architecture and engineering are considered service industries. Table 2 shows the process innovation metrics and activity by the Australian Industry for two years ending June 2021. The construction industry ranks second of the four industries in the value of intellectual property, which includes software.

Table 2. Goods-Producing Industries’ Measures in Australia FYE 2020 & 2021 - Average. (Source: ABS 2022)

Industry	Revenue	Value of Intellectual Property*	Intellectual Property/ Revenue	Value Added**	Value Added / Revenue	IP/Value Added
Agriculture, Forestry and Fishing	\$120,542	\$345	0.3%	\$35,585	29.5%	0.9%
Construction	\$259,030	\$2,090	0.8%	\$135,451	28.4%	3.1%
Manufacturing	\$447,124	\$4,811	1.1%	\$116,004	25.9%	4.1%
Mining, Oil & Gas	\$392,966	\$1,009	0.3%	\$250,948	63.9%	0.4%

Average utilised to minimise COVID disruption effects. *Includes ICT Software **Valuation of products less cost to produce - expressed in current prices in millions AUD

Rapid standardisation and automation are possible with software tools and digitalised business processes (Matt & Rauch, 2014). The transformational potential is apparent to many researchers in the efficacious re-engineering of workflows in both the planning and execution stages. Moreover, industry leaders, including software developers, have tried to improve the current situation (Faghihi et al., 2015). It is important to note that some of this effort is manifesting. The Australian Bureau of Statistics (2024) shows a decrease in the Administrative and Support Services sectors by 18.5% from FYE 2020 to 2024—the most of any industry in Australia. The investment community concluded that the reduction is real and the opportunity significant. McKinsey and Company (2023) reported that an estimated \$50 billion was invested in AEC tech between 2020 and 2022, 85 per cent higher than the previous three years. During the same period, the number of deals in the industry increased 30 per cent to 1,229.

Construction managers must interpret and process voluminous data for proper, up-to-date decision-making in effectively running a project. Therefore, the project's success depends on the increased reliance on technology. However, individuals' readiness to adopt technology has four dimensions: optimism, innovativeness, discomfort and insecurity. The four are independent; optimism and innovativeness encourage people to use and hold a positive attitude toward technological products and services, while a lack of comfort and security prevents their adoption (McNamara et al., 2022).

Since 2002, articles reflecting ICT have been trending downward, resulting in fewer journal articles in 2015. (Adwan and Al-Soufi 2016). Additionally, minimal research has been conducted to capture the issues barring technological adoption among small firms (Clermont et al. 2020) to help them identify their real needs and wants, including their challenges. This identification can help companies make changes and better address pressure from competitors (Lasni & Boton, 2022).

Today, software development is conducted in a chaotic environment. For example, Holmstrom et al. (2012) found disordered and dynamic markets, complicated and uncertain customer demands, shorter development cycle time pressures, and Moore's Law effects in most software development projects. Notably, twelve barriers in the industry dampen construction innovation. This results in suboptimal investment in new ideas, including software development (Stevens & Smolders 2023).

There is no standard software development process due to the differing characteristics of the organisations, products, and projects. The process is contingent on the knowledge and experience of the software research and design staff and the organisational guidelines, including the economic ones they must work within (Choi et al., 2017). In many software development firms, product managers struggle to get well-defined customer perceptions, needs and wants. Often, software value and utility validation occur after deployment. Furthermore, learning from customers is neither typically formalised nor continuous. As a result, the selection and prioritisation of utilities become suboptimal, and products are misaligned from what the customers need or want (Fabijan, 2015). Literature has established that the lack of programming management commitment has been one of the top reasons for the failure of Software Process Improvement (SPI) (Abrahamson 2000a). Their framework defines the levels of success achieved in SPI initiatives.

Agile software development is well-known for its focus on speeding project implementation by considering user requirements at stages in an iterative process (U.S. Digital Service, 2024). However, while it has succeeded in efficient programming, there is an urgent need to understand customer use, perceived value, and shortcomings. Continuous deployment delivers functional software consistently to customers while learning extemporaneously about customer usage. However, the transition towards short-cycle deployment involves several barriers (Holmstrom et al., 2012). Abrahamson (2000b) suggested five dimensions to gauge the success achieved in SPI: (1) project efficiency, (2) impact on the process user, (3) business success, (4) direct operational success and (5) process improvement fit. These were adapted from the project management literature and support Lean Construction principles.

Some implementation theories in construction management advocate a "push" strategy during which current practices must be radically changed to align with software functionality. Others advocate using well-accepted construction management planning, execution and measuring processes (Hartmann et al. 2012). Mobile Computing Apps appear to be part of the march toward better solutions for practitioners of all industries. Singularly focused individuals can author them and do not have to be integrated with other software modules (Weichbroth, 2020).

User-centred design (UCD) is a viable approach to usable software applications through comprehensive user studies (Mayhew, 1999). Compared with the agile approach, UCD outweighs the user analysis more in effectively eliciting user requirements for software engineering (Cockton et al. 2018). Given the complex and knowledge-based nature of the AECO industry, UCD is applicable for capturing essential user requirements to satisfy ICT solutions. Although cost is higher in UCD, its cost-benefit is obvious from a long-run perspective to ensure usability (Park et al. 2022). It has been proposed that integrating UCD in agile development to help efficient software development satisfy users' requirements (Cockton et al. 2018).

Despite the cited literature, most software programs have been deemed insufficient to meet user requirements. Several comprehensive studies have been conducted to determine usability. Usability engineering

lifecycle (Mayhew, 1999) provides practical guides for software engineering professionals to adopt into their ICT projects, including BIM. Moreover, the Software Usability Measurement Inventory (SUMI) survey is specific to answering value-oriented questions. It uses 50 questions, making use of five defined subscales for a) Efficiency, b) Affect, c) Helpfulness, d) Control, and e) Learnability to query users' attitudes toward the software. The work on SUMI began in 1986 with Kirakowski, who was entrusted with a project with two objectives: examine the competence scale of the Computer User Satisfaction Inventory and create an international standardisation for a new questionnaire.

3. METHOD

The researchers created an online anonymous survey for industry professionals. Their university's Human Ethics Committee approved the instrument in all respects. Online questionnaires are administratively efficient, and anonymity is correlated with a higher percentage of honest disclosure. Our demographic questions centered on qualifications, experience, and current intensity regarding the use of ICT. This approach lent itself to better analysis and sensitivity to any incoherence in the data (Stantcheva 2023).

Links were shared on social media such as LinkedIn and The Construction Network of Building Researchers (CNBR) and internally to the researcher's student cohort, many of whom work in the industry while studying. The survey asked:

1. Nine demographics, such as position and years of experience, and software use, such as weekly hours and discipline focus.
2. Presents the practice statements and queries for each item below in none, low, medium and high
 - a. The value of the practice
 - b. The performance of each practice by their firm
 - c. How strongly is it embedded in the firm's software
3. What practices are not in the software you are familiar with? (text response)
4. What mobile computing applications (Apps) are needed? (text response)
5. Do you have any other thoughts you would like to share? (text response)

Stevens and Smolders (2023) listed the practices below as efficacious for construction contracting operations. Stevens (2012) discovered valuable processes as perceived by the industry and their relative effect on Overhead/Project Cost efficiency.

1. *Dual Overhead Rate Application* is a methodology that precisely assigns overhead (Office G&A) costs to site labour, equipment, material, and subcontractors. This methodology is used in both project estimating and job cost reporting. Accurate costing improves Tender success and flows through to job cost and project return on investment. This reduces wasted work acquisition efforts and gives a project's construction cost.
2. *Job Sizing Adjustment*—tailoring overhead cost percentages due to the project size variation—the difference between the company's average job size and the tendered job—based on banking and industry data. Accurate costing improves Tender success and flows through to job cost and project return on investment. This reduces wasted work acquisition efforts.
3. *Predictive Tender Modelling* - a competitive practice that determines a competitor's price based on history. It utilises all the factors a constructor uses to adjust their price and systematises the process so the company does not grossly underbid. In other words, using it helps contractors leave less margin between their price and the competitors' price, predicting competitors' price and "leaving less money on the table." This improves profitability and reduces wasted work acquisition efforts.
4. *Forecasting Project Resource Demand* – limited and shared inputs such as cash, craftsperson, managers, and equipment must be placed where they produce the most benefit across the projects a constructor will build simultaneously. Forecasting from 6 weeks to 6 months ahead allows executives to ensure shared resources are available when needed. In addition, contractors manage multiple projects at a time with limited resources. Therefore, getting more done with the same inputs positively impacts the company and its clients. This reduces wasted resources from quick decisions since planning starts six weeks in advance.
5. *Unit-Based Project Billing and Internal Reporting* – utilises a count-based number for all products installed in a building. Units include each for doors or toilets, square meters of concrete forming, cubic meters of concrete or excavation, and linear meters of handrail or coping. This allows precise progress determinations and billing calculations while encouraging quality completion of each unit. A method to estimate, cost and administrate projects more precisely. There is less conflict, especially with billing. This reduces arguments (wasted time) about physical progress and, thus, provides monthly payment justification.
6. *Task Completion Monitoring and Measuring* - all tasks to be completed are listed electronically and

assigned to the employee responsible, such as planning or budgeting tasks. Construction firms may allocate many functions to the project manager for the job. Monitoring and measuring completion timeliness increases adherence. Teams build projects. Individual members complete critical tasks such as planning or procuring, which are best done in a pre-determined order. Accomplishing these tasks ultimately and timely increases multifactor productivity. This reduces wasted time and effort by keeping employees focused on critical tasks.

7. *Project Return on Investment* - a complex equation that determines the return on investment for a project using 22 factors, such as rent, equipment, payroll, client and supplier payment terms, profit, and allocated capital. These are measured against profit earned and annualised,
8. *Staff Load Balancing* – using a dozen or more factors, such as the number of duties, new clients, meetings, and project distance, to determine the relative utilisation of each staff member to ensure a relatively equal workload. People are the enablers of safety, quality, and productivity. Overloaded staff make mistakes; thus, rework negatively affects critical outputs. This keeps wasted time – the underutilisation of some employees, resulting in overburdening others – to a minimum.
9. *Calculating Organisational Multifactor Productivity Analysis*—one such as the KLEMS Model used by the Australian Bureau of Statistics. KLEMS is the same methodology standard used to determine the industry's resource efficiency. KLEMS stands for K-cash, L-labour, E-energy, M-material, and S-services. It can be a benchmark for contractors to compare against the industry in current and past years.
10. *Project Site Material Laydown Planning and Logistics* – since approximately 70% of lost time is due to material logistics factors such as delivery timing, counts, product quality and handling, this is a critical practice to improve productivity. Since approximately 70% of lost time is due to material logistics, i.e., timely delivery, counts, quality, and handling, it is critical to pre-plan onsite material storage, handling, and flow. This reduces wasted worker time and effort handling material on the job site.

The researchers utilised a Likert Scale of 1 to 5 (5 = highest) on three dimensions:

- a) *Practice Value (PV)*-the process articulated is perceived to provide above-average results in four outputs: safety, quality, cost, and schedule.
- b) *Firm Execution (FE)*-the firm's consistency and strength in performing the surveyed practice.
- c) *Software Support (SS)*-the firm's ICT capability to organise and execute the surveyed practice.

Additionally, the survey asked respondents to indicate if they wish to receive a quarterly summary of results, which tends to boost response rates (Stantcheva 2023).

4. RESULTS

This data provided the researchers with evidence of current construction contractor software's general value and utility in work acquisition, project operations, financial management, and building information modelling.

The results show that none of the ten practices was supported significantly in the surveyed programming modules, as evidenced by our industry survey. However, for confidentiality and legal considerations, we did not disclose the names of the software packages. Most are popular with contractors and have been in the industry for five years or more.

Comparisons were made, and statistical tests were conducted. The tested practices' value (PV) is perceived higher than the firm's execution (FE) consistency and the enabling support (SS) from the respondent's software by its firm provides. See Table 3 (n=23).

Table 3. Mean Ratings of Each Practice on Three Dimensions and % Comparisons

Practice	PV	FE	SS	PV/FE %	PV/SS%
1	3.09	2.74	2.22	12.70%	39.22%
2	3.52	2.91	2.57	20.90%	37.29%
3	3.35	2.39	2.13	40.00%	57.14%
4	4.22	3.35	2.83	25.97%	49.23%
5	3.78	2.87	2.74	31.82%	38.10%
6	3.35	2.91	2.83	14.93%	18.46%
7	3.61	3.00	2.48	20.29%	45.61%
8	2.91	1.70	1.78	71.79%	63.41%
9	2.57	1.96	1.70	31.11%	51.28%
10	3.43	2.48	2.13	38.60%	61.22%

PV ranged between 2.57 and 4.22. This indicate moderate efficacy however, the average of FE was from 1.7 to 3.35 and SS between 1.70 and 2.83. The averages of the three dimensions ratings shows a relative disparity indicating relative value. Furthermore, the range of each, 1.65, 1.65 and 1.13, shows a statistically small variation and significant stability of perceptions.

The cumulative practices' rating on each dimension, i.e., Practices Value, Firm Execution, and Software Support, shows disparities. These appear to point to two dynamics: a) Firms are not executing practices they deem valuable. b) the firm's software does not support the practices at its assessed level of value. See Table 4.

Table 4. Cumulative Rating on Each Dimension – All Surveys

Dimension	Totals	% Disparity To PV
Practice Value	33.83	NA
Firm Execution	26.30	-28.60%
Software Support	23.39	-44.61%

Best practices are assessed more critically by experienced professionals regardless of industry. The researchers deem these survey participants to be more aware of practice value. The ratings from those who self-disclosed their experience of 20 years or more in the construction industry were significantly lower (27.33 versus 33.83). See Table 5.

Table 5. Respondents with 20 Years or More Industry Experience

Dimension	Totals	% Disparity To PV
Practice Value (PV)	27.33	NA
Firm Execution	21.33	-28.15%
Software Support	18.50	-47.75%

Those with high software familiarity found the disparity between software and its practice support even greater than the study average. These respondents use their firm's software for an average of 25 hours or more a week, rated enablement lower than the average (21.73 versus 23.39). This cohort is one that confirms the gap hypothesised. See Table 6.

Table 6. Respondents who average 25 Hours of More Working with the Firm's Software

Dimension	Totals	% Disparity To PV
Practice Value	32.53	NA
Firm Execution	25.25	-28.84%
Software Support	21.73	-49.69%

5. DISCUSSION

This paper's selection of effective practices represents critical contractor operational functions such as work acquisition, project operations, and financial management. Thus, the software brands investigated were more than one type. Since the perceived differences appeared consistently across this specialised software, there is credible evidence of the gap hypothesised. The researchers selected these practices based on their project manager and consultant experience.

Software businesses, in general, are economically oriented, like most for-profit organisations. They appear to be falling short of providing efficacy and, thus, value. Their construction clients seem to have found "workarounds" to perceived gaps. In the short term, this paper asserts that reprogramming is needed for the construction industry's ICT to support proven practices. In the long term, a robust research process appears required in order to document contractor best practices for consumption by construction professionals, including software developers.

Focusing software programming on construction contracting firms' overall operational practices is a slight departure from the project orientation that has been a consistent trend in academia and ICT. This paper asserts that the construction organisation is an enabler of the project outcomes. Critically, the contracting firm supports the project team, i.e., with office personnel assistance, their experience, coordination of company assets, and enforcing contract terms and conditions. Project teams can be likened to residing on an island where the limits of resources are constrained to the area they occupy. The corporation's members can connect the project team to capable resources, making task completion safer, faster, and higher quality. So, systematising company-wide practices with the support of ICT can minimise or eliminate stubborn problems such as efficient resource allocation. Software developers should be incentivised to fill the hypothesised gaps with increased sales.

Utilising low-code spreadsheets can achieve profitable standardisation and automation. Some SME contractors rely on customising computer spreadsheets to calculate the supporting information needed to execute some practices. In contrast, others are unaware of the methods or have not taken this additional step. Software

developers can improve this apparent inconsistent application and execution by creating interfaces that raise sufficient practice use levels. However, some need to know more about best practices before programming.

Lastly, this previously documented gap in most software (all industries) raises the question, "What else has been overlooked in the programming of construction software?" The researchers sense that an association-sponsored panel for each construction segment is needed. It might formally report on best practices to enlighten the industry and software developers.

6. CONCLUSIONS AND RECOMMENDATIONS

This paper delved into a widespread misalignment problem in construction contracting with computer software, examining its underlying causes, consequences, and potential remedies. The study established that misalignment in construction contracting arises from conflicting stakeholder goals, varying interpretations of contractual terms, and limited flexibility to accommodate changing project dynamics. Such misalignments frequently result in disputes, operational inefficiencies, and chaotic project completions, emphasising the urgent need for strategic intervention to address these challenges and enhance contract management practices. The industry survey data highlighted best practices focused on fostering transparent communication, promoting fair distribution of risks, and utilising standardised frameworks. Although these approaches effectively mitigate certain aspects of misalignment, their implementation is often constrained by organisational dynamics and cultural challenges.

Moreover, digital tools and advancements in constructor-centric software offer considerable potential to address these challenges by streamlining contract processes, enhancing stakeholder cooperation, and improving compliance oversight. Contracting software has the potential to solve the industry's stubborn problems, such as disproportionately high bankruptcy rates, low profitability, and stagnant multifactor productivity. There are many sources of these issues; however, this paper asserts that tailoring construction software to enable effective practices could lessen these problems.

The study hypothesised that a substantial gap exists between most construction software modules and well-accepted practices. This assertion appears credible. The survey analysis found a significant blind spot in the industry. Furthermore, academics studying the construction software industry have found it chaotic, and its firms choose commercial priorities, such as revenue generation. Finally, the practices articulated in the survey have existed for decades and are considered valuable by many industry professionals, showing a blind spot for programmers they should realise.

Adopting supporting modules for the ten practices list would be a modest start to improvement. Subsequently, intense conversations and other information-gathering approaches would discover and confirm processes that contractors believe are best. Developers should create APIs to add to their offerings. ICT salespeople and trainers should be educated in contractor practices, their effects, and how their software enhances execution.

Professional associations and governmental authorities should pursue "nudge" policies that add to the knowledge systems, intellectual property and technological capability to the construction industry and its software partners. Engaging these two industries can only result in fewer blindspots, iterative innovation and, at times, breakthroughs benefitting both parties. This ultimately would improve the built environment and the world's quality of life.

By their nature, researchers are interested in future possibilities. However, some backfilling is necessary in construction ICT to increase safety, reduce waste, and produce more value. Of course, realising the software's full potential now helps future stakeholders maximise new technologies such as AI and Quantum Computing. Also, developers that capture and systematise contractor practices best can grow their revenue. The researchers' future investigation should include a) collecting and articulating better construction processes, b) a deeper exploration of the software's indirect effect on best practice adoption, and c) life cycle cost against business benefit.

7. REFERENCES

- Abrahamsson, P. (2000a) 'Is management commitment a necessity after all in software process improvement?' Proceedings of the 26th Euromicro Conference. EUROMICRO 2000. Informatics: Inventing the Future, pp. 246-253 vol.2, <https://doi.org/10.1109/EURMIC.2000.874425>.
- Abrahamsson, P. (2000b). Measuring The Success of Software Process Improvement: The Dimensions. Eurospi'00, Copenhagen, Denmark. <https://doi.org/10.48550/arXiv.1309.4645>
- Adwan, E. & Al-Soufi, A. (2016) A Review of ICT Technology in Construction (November 22, 2016). International Journal of Managing Information Technology (IJMIT) Vol.8, No.3/4, Available at SSRN: <https://doi.org/ssrn.com/abstract=3237181>
- Ahuja, H. N., Dozzi, S.P. and AbouRizk, S.M. (1994). Project management: Techniques in planning and controlling construction projects. New York: Wiley
- Andújar-Montoya, M., Galiano-Garrigos, A., Echarrri-Iribarren, V., & Rizo-Maestre, C. BIM-LEAN as a methodology to save execution costs in building construction-An experience under the Spanish

- framework, *Appl. Sci. (Switz.)* 10 (6) (2020) 1–21, <https://doi.org/10.3390/app10061913>.
- Australian Bureau of Statistics (ABS) (2024) Job vacancies, change from February 2020 to August 2024, industry, original.
- Australian Bureau of Statistics (AB3) (2022) TABLE 02. Value of Construction Work Done, by Sector, Current Prices, Australia. 8755.0 Construction Work Done, Australia, Preliminary.
- Australian Bureau of Statistics (ABS) (2023) Table 2 Process innovation, industry by employment size – 2 years ended 30 June 2021. Data Table 81580DO004_202021 Characteristics of Australian Business, 2020-21, Businesses that introduced any new or significantly improved (methods).
- Australian Securities and Investment Commission (ASIC) (2023) ASIC Insolvency Statistics Series 3.1. Accessed July 13, 2023.
- Australian Bureau of Statistics (ABS) (2022), '8165.0 Counts of Australian Businesses, including Entries and Exits, June 2017 to June 2022'
- Bamgbose, O. A., Ogunbayo, B. F., & Aigbavboa, C. O. (2024). Barriers to Building Information Modelling Adoption in Small and Medium Enterprises: Nigerian Construction Industry Perspectives. *Buildings*, 14(2), 538. <https://doi.org/10.3390/buildings14020538>
- Choi, S., Kim, S., & Kim, J. (2017). A Technique Identifying Reusable Software R&D Process Assets for Software Process Tailoring. *International Information Institute (Tokyo). Information*, 20(10B), 7857-7864.
- Clermont, S., Lefebvre, G. and Boton. C. (2020). "Understanding the diffusion of building information modeling among contractor SMEs in the Quebec construction industry." In *Proc. of the 38th Int. Con. of CIB W78*, 762–772. Luxembourg: Luxembourg Institute of Science and Technology.
- Cockton, G., Lárusdóttir, M., Gregory, P., Cajander, Å. (2018). *Integrating User-Centred Design in Agile Development*. Springer International Publishing. ISBN:9783319812113, 3319812114
- Dave, B., Kubler, S., Främling, K., & Koskela, L., (2016). Opportunities for enhanced construction management using Internet of Things standards. *Automation in Construction*, Volume 61, Pages 86-97, ISSN 0926 5805, <https://doi.org/10.1016/j.autcon.2015.10.009>.
- Dubois, A., & Gadde, L. (2002). The construction industry as a loosely coupled system: Implications for productivity and innovation. *Construction Management and Economics*, 20(7), 621-631. <https://doi.org/10.1080/01446190210163543>
- Fabijan, A., Olsson, H.H., Bosch, J. (2015). Customer Feedback and Data Collection Techniques in Software R&D: A Literature Review. In: Fernandes, J., Machado, R., Wnuk, K. (eds) *Software Business. ICSOB 2015. Lecture Notes in Business Information Processing*, vol 210. Springer, Cham. https://doi.org/10.1007/978-3-319-19593-3_12
- Faghihi, V., Nejat, A., Reinschmidt, K. & Kang, J. (2015). Automation in construction scheduling: a review of the literature, *Int. J. Adv. Manuf. Technol.* 81 (9–12)1845–1856, <https://doi.org/10.1007/s00170-015-7339-0>.
- Hartmann, T., Van Meerveld, H., Vossebeld, N., & Adriaanse, A. (2012). Aligning building information model tools and construction management methods. *Automation in construction*, 22, 605-613.
- Holmström Olsson, H., Alahyari, H., Bosch, J. (2012). Chapter 1 Climbing the Stairway to Heaven. In: Bosch, J., Carlson, J., Holmström Olsson, H., Sandahl, K., Staron, M. (eds) *Accelerating Digital Transformation*. Springer, Cham. https://doi.org/10.1007/978-3-031-10873-0_2
- Kalsaas, B. T., & Sacks, R. (2011). Conceptualization of interdependency and coordination between construction tasks. *Proceedings of the IGLC-19*. Lima, Peru.
- Lasni, A., & Boton, C. (2022). Implementing construction planning and control software: A specialised contractor perspective. *Journal of Construction Engineering and Management*, 148(9) [http://doi.org/10.1061/\(ASCE\)CO.1943-7862.0002330](http://doi.org/10.1061/(ASCE)CO.1943-7862.0002330)
- Matt, D.T., and Rauch, E. (2014). Implementing lean in engineer-to-order manufacturing, in V. Modrak, P. Semančco (Eds.), *Handbook of Research on Design and Management of Lean Production Systems*, IGI Global, 2014, pp. 148–172, <https://doi.org/10.4018/978-1-4666-5039-8.ch008>
- Mayhew, D., J. (1999). *The Usability Engineering Lifecycle: A Practitioner's Handbook for User Interface Design (Interactive Technologies)*, 1st Edition. Morgan Kaufmann. ISBN-13 978-1558605619
- McKinsey and Company (2023) From start-up to scale-up: Accelerating growth in construction technology. Accessed November November 28, 2024. <https://www.mckinsey.com/industries/private-capital/our-insights/from-start-up-to-scale-up-accelerating-growth-in-construction-technology>
- McNamara, AJ, Shirowzhan, S & M.E. Sepasgozar, S (2022). Investigating the deterrents of intelligent construction contract adoption: a refinement of the technology readiness index to inform an integrated technology acceptance model, *Construction Innovation*, DOI 10.1108/ci-10-2021-0191.
- Ogunbayo, B. F. & Aigbavboa, C.O. Ahmed S. and Stevens M., "Assessing Monitoring and Evaluation Effectiveness for Projects in the Construction Industry," 2024 International Conference on Science, Engineering and Business for Driving Sustainable Development Goals (SEB4SDG), Omu-Aran, Nigeria, 2024, pp. 1-10, <https://doi.org/10.1109/SEB4SDG60871.2024.10630084>

- Őri, D., & Szabó, Z. (2024). A systematic literature review on business-IT misalignment research. *Information Systems and e-Business Management*, 1-31. <https://doi.org/10.1007/s10257-023-00664-w>
- Sande, J. B., & Haugland, S. A. (2015). Strategic performance effects of misaligned formal contracting: The mediating role of relational contracting. *International Journal of Research in Marketing*, 32(2), 187-194. <https://doi.org/10.1016/j.ijresmar.2015.02.002>
- SCOPUS (2024). 'SCOPUS: Empowering discovery since 2004' Accessed November 24, 2024. <https://www-scopus-com.ezproxy.uws.edu.au/search/form.uri?display=basic#basic>
- Solli-Sæther, H., Karlsen, J. T., & van Oorschot, K. (2015). Strategic and cultural misalignment: Knowledge sharing barriers in project networks. *Project Management Journal*, 46(3), 49-60. <https://doi.org/10.1002/pmj.21501>
- Stantcheva, S. (2023). How to run surveys: A guide to creating your own identifying variation and revealing the invisible. *Annual Review of Economics*, 15(1), 205-234.
- Stevens, M. (2012) The Impact Of 154 Selected Practices On Overhead Efficiency For Specialty Construction Contractors. Doctoral Dissertation. University of Florida, Gainesville.
- Stevens, M. & Smolders J. (2023). Understanding Australian Construction Contractors: A guide for emerging professionals. Routledge. London.
- Weichbroth, P. (2020). "Usability of Mobile Applications: A Systematic Literature Study," *IEEE Access*, vol. 8, pp. 55563-55577, 2020, doi: 10.1109/ACCESS.2020.2981892 .
- Wu, J. W., Tseng, J.C. Yu, W.D., Yang, J.B., Lee, S.M., and Tsai, W.N. (2012). "An integrated, proactive knowledge management model for enhancing engineering services." *Autom. Constr.* 24 (Jul): 81–88. <https://doi.org.ezproxy.uws.edu.au/10.1016/j.autcon.2012.02.006>.
- Ye, X., Zeng, N., & König, M. (2022). Systematic literature review on smart contracts in the construction industry: Potentials, benefits, and challenges. *Frontiers of engineering management*, 9(2), 196-213. <https://doi.org/10.1007/s42524-022-0188-2>
- U.S. Bureau of Labor Statistics (USBLS) (n.d.) Industries at a Glance. Accessed July 9, 2023. <https://www.bls.gov/iag/tgs/iag06.htm>
- U.S. Digital Service. (2024). What is Agile Software Development | TechFAR Hub Handbook | USDS.gov. [online] Available at: <https://techfarhub.usds.gov/pre-solicitation/agile-overview/>. Accessed 27/Nov/2024

EXPLORING THE RISKS OF INTEGRATING GENERATIVE ARTIFICIAL INTELLIGENCE INTO CONSTRUCTION RISK MANAGEMENT: INSIGHTS FROM A SYSTEMATIC LITERATURE REVIEW

Mohamed Mohamed¹; M.K.S Al-Mhdawi^{1,2}; Farzad Rahimian¹; Udechukwu Ojiako^{3,4,5}; Alan O'Connor²; and Charf Mahammed¹

¹School of Computing, Engineering & Digital Technologies, Teesside University, United Kingdom.

²Department of Civil, Structural and Environmental Engineering, Trinity College Dublin, Ireland

³Design, Manufacturing & Engineering Management, University of Strathclyde, United Kingdom

⁴Centre for Systems Studies, University of Hull, United Kingdom.

⁵Johannesburg Business School, University of Johannesburg, South Africa.

ABSTRACT

The rapid advancements in Generative Artificial Intelligence (GenAI) have unlocked transformative potential across various industries, including construction. With its ability to generate content, automate processes, and enhance decision-making, GenAI offers significant opportunities to improve the efficiency and accuracy of Construction Risk Management (CRM). However, its integration into CRM also brings a new set of risks and uncertainties that are unprecedented in traditional risk management frameworks. To this end, the purpose of this research is to identify and classify the key risks associated with integrating GenAI into CRM. To achieve this, a three-step systematic literature review was conducted, analysing 48 scholarly articles on GenAI for CRM from Scopus-indexed academic journals published between 2014 and 2024. A total of 25 risk factors associated with GenAI integration in CRM were identified and classified under seven key categories: financial risks, technological adaptability risks, information integrity risks, input quality risks, and ethical and governance risks. This study enhances the understanding of risk factors in GenAI integration by presenting a structured framework that categorises the associated risks of GenAI integration into CRM while highlighting their interconnectedness. It also lays the foundation for interdisciplinary approaches and future empirical research to validate and expand these insights across diverse construction contexts.

Keywords: Generative AI, GenAI risks, Construction Risk Management, Construction industry.

1. INTRODUCTION

The rapid advancement of Generative Artificial Intelligence (GenAI) has brought transformative potential to various industries, including construction (Al-Mhdawi et al., 2023; Chenya et al., 2022). With its ability to analyse complex data, enhance decision-making, and optimise workflows, GenAI has emerged as a promising tool for addressing the complex challenges of Construction Risk Management (CRM) (Nyqvist et al., 2024). Utilising predictive analytics, real-time data insights, and automation of repetitive tasks, GenAI has the potential to significantly improve project outcomes, streamline operations, and reduce inefficiencies (Mohamed et al., 2024; Jallow et al., 2023). These applications position GenAI as a game-changer in a sector characterised by high costs, strict timelines, and the need to balance diverse stakeholder requirements (Al-Mhdawi et al., 2023).

Despite these promises, the integration of GenAI into CRM also introduces significant risks that demand careful evaluation. Data-related risks such as unavailability, poor quality, and complexity can undermine the accuracy and reliability of AI-driven predictions (Adekunle et al., 2022; Holzmann and Lechiara, 2022). Ethical concerns, including data biases, lack of transparency, and privacy violations, further complicate its adoption. Moreover, operational risks—such as the disruption or inefficiencies arising from integrating GenAI with legacy systems, insufficient workforce training, and cultural resistance within organisations—can amplify the risks (Nabawy and Gouda, 2024; Regona et al., 2022). These risks not only threaten the successful deployment of GenAI but can also result in project delays, cost overruns and compromised stakeholder satisfaction (Mohamed et al., 2024b). While existing research has largely emphasised the benefits and applications of GenAI in construction (see., e.g., Al-Mhdawi et al., 2023; Aladag, 2023; Hofert, 2023; Nyqvist et al., 2024), studies addressing the associated risks remain limited. Prior work has primarily focused on general AI applications, without systematically examining the specific risks unique to the integration of generative technologies into CRM. Additionally, there is a lack of structured frameworks to guide decision-makers in identifying, classifying, and mitigating these risks across different stages of the project lifecycle.

To address these gaps, this research aims to identify and categorise the key risks associated with implementing GenAI into CRM. The findings contribute to the growing body of knowledge on risk management in technologically advanced construction settings and offer actionable insights for decision-makers to navigate the complexities of GenAI adoption effectively.

2. RESEARCH METHODOLOGY

This research adopts a structured, three-step approach to literature collection and analysis, aimed at thoroughly examining existing studies to identify the key risks associated with GenAI into CRM. The first step involves identifying relevant databases and journals to establish a solid foundation for the literature search. The second step focuses on strategically selecting articles through targeted keywords to ensure the inclusion of the most relevant studies. Finally, the third step involves conducting a systematic content analysis to extract meaningful insights. This methodology is informed by the frameworks proposed in several risk management studies, notably those by Al-Mhdawi et al. (2024a), Al-Mhdawi et al. (2024b), and Mohamed et al., (2024), as illustrated in Figure 1.

Step One: Database and Journal Identification

In this research scopus database was chosen due to its comprehensive coverage of relevant research disciplines and their established use in comparable literature-based studies within construction management research. The selection of target journals for this study was based on the following criteria: (1) the journals must be published in English, (2) they must have a minimum impact factor of 1.0, and (3) they must be ranked in the top quartile of the Scopus database, recognised for their significant influence in shaping construction management research.

Step Two: Keyword Identification and Article Selection

In this step, a comprehensive search was conducted using the title/abstract/keyword (T/A/K) fields in the Scopus search engine. The search keywords included GenAI risks, GenAI, GenAI in CRM, and GenAI in construction project management. Papers containing these specific terms in the title, abstract, or keywords were deemed to have met the initial criteria for further analysis. These keywords were carefully selected to capture a broad range of studies addressing the risks and applications of GenAI in CRM and related domains. Papers containing these specific terms in any of the T/A/K fields were considered to meet the preliminary inclusion criteria. The results were further filtered to eliminate duplicate entries, irrelevant studies, and papers that lacked substantive focus on the intersection of GenAI and CRM.

Step Three: Content Analysis

Hsieh and Barman et al. (2021) identify three approaches to content analysis: conventional, directed, and summative. This study employed conventional content analysis, an open-ended, data-driven method that allows categories to emerge naturally without relying on predefined frameworks (Blomkvist, 2015). This approach, suitable for both qualitative and quantitative analysis, is ideal for exploring the emerging topic of integrating GenAI into CRM, as it captures detailed themes directly from the data (Kibiswa, 2019). Unlike directed analysis, it avoids constraints from existing theories, enabling a rich, context-specific understanding (Hsieh and Shannon, 2005; Krippendorff, 2018). Through this method, the study systematically narrowed 183 initial papers to 48, identifying key risks and categories associated with GenAI in CRM.

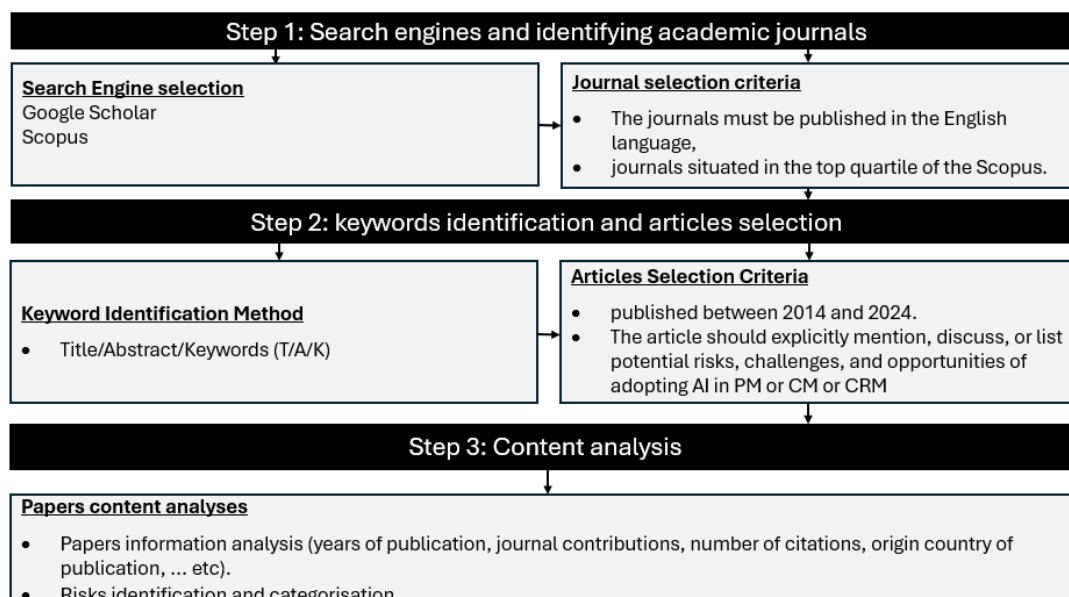


Figure 1. Adopted research methodology

3. RESULTS AND DISCUSSION

Based on the adopted research methodology, a total of 48 scholarly articles were analysed to identify the key risks of GenAI integration into CRM. This analysis yielded 25 key risk factors classified under five categories, namely: financial risks, input quality risks, technological adaptability risks, information integrity risks, and ethical and governance risks. Figure 2 illustrates the identified risks and their respective categories.

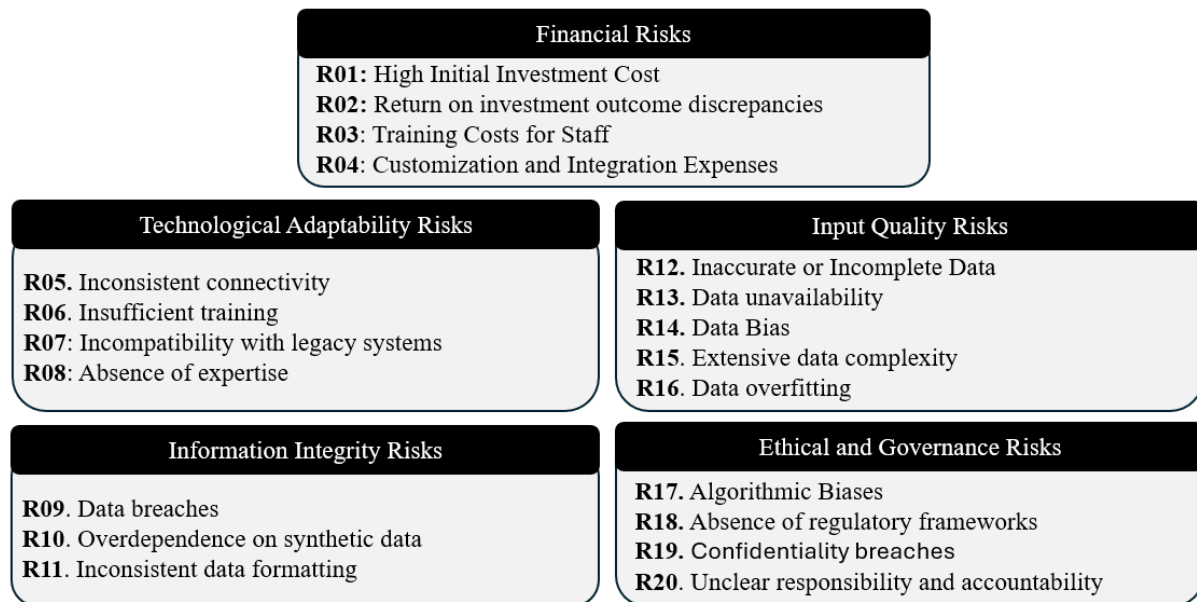


Figure 2. Identified Risks of Integrating GenAI into CRM

The findings reveal a complex interplay of technological, financial, operational, and ethical factors influencing the adoption of GenAI into CRM. Among these, financial risks stand out as a significant concern. Research by Chenya et al. (2022) and Nabawy and Gouda Mohamed (2024) highlights the necessity of conducting CRM-specific cost-benefit analyses (Al-Mhdawi et al. 2024c). These analyses not only clarify the economic feasibility of GenAI adoption but also help address stakeholder concerns by demonstrating potential long-term savings and operational efficiencies. For example, innovative financial models that account for both initial investment costs and operational gains could significantly bolster stakeholder confidence. This aligns with prior studies (e.g., Zhou et al., 2019), which highlight the critical role of financial viability in technology adoption.

Another vital dimension is information integrity risks, which are amplified by the interconnected nature of construction projects. Díaz-Curbelo et al. (2020) emphasise the prevalence of cybersecurity threats, such as data breaches, that can compromise project outcomes and stakeholder trust. Tackling these risks requires robust cybersecurity frameworks. Recent advancements, such as adversarial machine learning proposed by Mohamed et al. (2024), offer innovative solutions to protect sensitive project data. These findings resonate with broader calls for construction-specific cybersecurity measures, as highlighted by Smith et al. (2021). Input quality also emerges as a pivotal factor. The dynamic and complex nature of construction projects demands high-quality, contextually relevant data for GenAI systems to deliver accurate and actionable insights. Research by Ahuvia (2001) and Poh et al. (2018) highlights the importance of data relevance and completeness for effective decision-making. Aligning with these insights, Meng et al. (2022) advocate for integrating real-time data validation systems to enhance the reliability of AI outputs in construction applications.

Technological adaptability risks pose an additional layer of risks, particularly in resource-constrained settings. Pan and Zhang (2021) and Jallow et al. (2023) stress the need for resilient infrastructure and offline AI capabilities to maintain operational continuity under suboptimal conditions. The co-design of localised AI tools with industry professionals is especially critical for addressing site-specific challenges and ensuring the practical application of GenAI in diverse construction scenarios. This approach aligns with Marra and Kearney (2020), who emphasise co-creation strategies to bridge gaps between innovation and field-specific needs. Lastly, ethical and governance risks further complicate the integration of GenAI into CRM. Risks such as algorithmic bias, lack of transparency, and inadequate regulatory frameworks threaten equitable and effective adoption. Holzmann and Lechiara (2022) highlight the need for transparent governance mechanisms and evolving regulations to promote responsible

innovation. Similarly, Pillai and Matus (2020) argue for stakeholder-inclusive policymaking to foster ethical AI practices. These perspectives align with global discussions, such as Floridi and Cowls (2019), which advocate for accountability and equity in AI governance to build trust and ensure sustainable adoption.

4. CONCLUSIONS

This research aimed to identify the key risks associated with integrating GenAI into CRM. Through an analysis of 48 scholarly articles, 25 key risk factors were identified and categorised into five primary areas: financial risks, technological adaptability risks, information integrity risks, input quality risks, and ethical and governance risks. The findings highlight that integrating GenAI into CRM presents a complex interplay of risks spanning financial, operational, technological, and ethical dimensions. This study offers both practical and theoretical implications. Practically, it highlights the need for targeted risk mitigation strategies, such as cost-benefit analyses and innovative financial models to address financial risks, robust cybersecurity frameworks to safeguard information integrity, and real-time data validation systems to enhance input quality. Tailored technology solutions and co-created AI tools are recommended to tackle technological adaptability risks, while transparent governance mechanisms and stakeholder-inclusive policymaking can promote ethical AI practices. Organisations are also encouraged to invest in capacity building to upskill their workforce for GenAI adoption. Theoretically, the study contributes to the understanding of risk factors in GenAI integration by providing a structured framework categorising financial, operational, technological, and governance risks, and emphasising their interconnectedness. Additionally, the findings lay the groundwork for interdisciplinary approaches and future empirical research to validate and expand these insights in diverse construction contexts.

Advancing the integration of GenAI into CRM requires a comprehensive and well-coordinated approach. A critical step involves investing in return on investment (ROI) models specifically designed to highlight the unique benefits of CRM applications. These models can effectively demonstrate long-term savings and operational efficiencies, offering a clear justification for stakeholders to support adoption by emphasising value beyond the initial costs (Regona et al., 2022). Alongside financial considerations, robust cybersecurity measures are essential to safeguard sensitive project data. Real-time threat detection systems powered by advanced algorithms provide proactive defences, while continuous monitoring within interconnected project ecosystems minimises the risk of data breaches (Díaz-Curbelo et al., 2020). To enhance reliability further, implementing data standardisation practices ensures consistent and accurate inputs for AI systems. Industry-wide protocols can improve interoperability, reduce errors, and facilitate seamless integration across diverse platforms (Anysz et al., 2021). Additionally, addressing the variability of construction site conditions necessitates the development of adaptable AI tools with offline capabilities. Such tools offer scalable solutions suited to resource-constrained environments (Pan and Zhang, 2021). Equally important is building trust, which can be achieved through transparent governance structures that mitigate biases and ensure accountability. At the same time, regulatory frameworks should evolve to balance fostering innovation with upholding ethical standards (Holzmann and Lechiara, 2022).

REFERENCES

- Adedokun, O.A., Aje, I.O. and Agboola, O.J., 2019. Investigation into the severity of factors predisposing construction projects to risks in Nigeria. *Journal of Building Performance ISSN, 10*(1), p.2019.
- Adekunle, P., Aigabvboa, C., Thwala, W., Akinradewo, O., & Oke, A. (2022). Challenges confronting construction information management. *Frontiers in Built Environment*, 8, 1075674.
- Ahuvia, A. (2001). Traditional, interpretive, and reception-based content analyses: Improving the ability of content analysis to address issues of pragmatic and theoretical concern. *Social Indicators Research*, 54, 139–172.
- Aladağ, H. and Işık, Z., 2018. The effect of stakeholder-associated risks in mega-engineering projects: A case study of a PPP airport project. *IEEE Transactions on Engineering Management*, 67(1), pp.174-186.
- Aladag, H., 2023. Assessing the accuracy of ChatGPT use for risk management in construction projects. *Sustainability*, 15(22), p.16071.
- Al-Mhdawi, M.K.S., Qazi, A., Karakhan, A.A., Rahimian, F.P., Abualqumboz, M., Al Mhdawi, A.K. and Al-Raweshidy, H., 2024c. Implementation of a Combined Fuzzy Controller Model to Enhance Risk Assessment in Oil and Gas Construction Projects. *IEEE Access*.
- Al-Mhdawi, M.K.S., O'Connor, A., Qazi, A., Rahimian, F. and Dacre, N., 2024a. Review of studies on risk factors in critical infrastructure projects from 2011 to 2023. *Smart and Sustainable Built Environment*.
- Al-Mhdawi, M.K.S., O'Connor, A. and Qazi, A., 2024b. Structural equation modeling and Fuzzy set theory: Advancing risk assessment in oil and gas construction projects. *Environmental Impact Assessment Review*, 109, p.107622.
- Mohamed, M., Al-Mhdawi, M.K.S., Ojiako, U., Dacre, N., Qazi, A., and Rahimian, F., 2024A. Generative AI in Construction Risk Management: A Bibliometric Analysis of the Associated Benefits and Risks.

- Urbanization, Sustainability and Society (In Press).*
- Mohamed, M., Al-Mhdawi, M.K.S., Dacre, N., Ojiako, U., Qazi, A. and Rahimian, F., 2024B. Theoretical and practical instantiations of generative AI in construction risk management: An analytical exposition of its latent benefits and inherent risks. *Available at SSRN 5007208*.
- Chenya, L., Aminudin, E., Mohd, S. and Yap, L.S., 2022. Intelligent risk management in construction projects: Systematic Literature Review. *IEEE Access*, 10, pp.72936-72954.
- Eber, W., 2020. Potentials of artificial intelligence in construction management. *Organization, technology & management in construction: an international journal*, 12(1), pp.2053-2063.
- Hofert, M., 2023. Assessing ChatGPT's proficiency in quantitative risk management. *Risks*, 11(9), p.166.
- Holzmann, V. and Lechiara, M., 2022. Artificial intelligence in construction projects: An explorative study of professionals' expectations. *European Journal of Business and Management Research*, 7(3), pp.151-162.
- Anysz, H., Apollo, M. and Grzyl, B., 2021. Quantitative risk assessment in construction disputes based on machine learning tools. *Symmetry*, 13(5), p.744.
- Hu, Y. and Castro-Lacouture, D., 2019. Clash relevance prediction based on machine learning. *Journal of computing in civil engineering*, 33(2), p.04018060.
- Jallow, H., Renukappa, S., Suresh, S. and Rahimian, F., 2023. Artificial intelligence and the UK construction industry—empirical study. *Engineering Management Journal*, 35(4), pp.420-433.
- Lachhab, M., Béler, C. and Coudert, T., 2018. A risk-based approach applied to system engineering projects: A new learning based multi-criteria decision support tool based on an Ant Colony Algorithm. *Engineering Applications of Artificial Intelligence*, 72, pp.310-326.
- Louis, J. and Dunston, P.S., 2018. Integrating IoT into operational workflows for real-time and automated decision-making in repetitive construction operations. *Automation in Construction*, 94, pp.317-327.
- Nabawy, M. and Gouda Mohamed, A., 2024. Risks assessment in the construction of infrastructure projects using artificial neural networks. *International Journal of Construction Management*, 24(4), pp.361-373.
- Nyqvist, R., Peltokorpi, A. and Seppänen, O., 2024. Can ChatGPT exceed humans in construction project risk management?. *Engineering, Construction and Architectural Management*, 31(13), pp.223-243.
- Obiuto, N.C., Adebayo, R.A., Olajiga, O.K. and Festus-Ikhuoria, I.C., 2024. Integrating artificial intelligence in construction management: Improving project efficiency and cost-effectiveness. *Int. J. Adv. Multidisc. Res. Stud*, 4(2), pp.639-647.
- Pan, Y. and Zhang, L., 2021. Roles of artificial intelligence in construction engineering and management: A critical review and future trends. *Automation in Construction*, 122, p.103517.
- Parveen, R., 2018. Artificial intelligence in construction industry: Legal issues and regulatory challenges. *International Journal of Civil Engineering and Technology*, 9(13), pp.957-962.
- Pillai, V.S. and Matus, K.J., 2020. Towards a responsible integration of artificial intelligence technology in the construction sector. *Science and Public Policy*, 47(5), pp.689-704.
- Poh, C.Q., Ubeynarayana, C.U. and Goh, Y.M., 2018. Safety leading indicators for construction sites: A machine learning approach. *Automation in construction*, 93, pp.375-386.
- Prasad Agrawal, K., 2024. Towards adoption of generative AI in organizational settings. *Journal of Computer Information Systems*, 64(5), pp.636-651.
- Regona, M., Yigitcanlar, T., Xia, B. and Li, R.Y.M., 2022. Opportunities and adoption challenges of AI in the construction industry: A PRISMA review. *Journal of open innovation: technology, market, and complexity*, 8(1), p.45.
- Schwarzkopf, D.L., 2006. Stakeholder perspectives and business risk perception. *Journal of Business Ethics*, 64, pp.327-342.



Part 12

**Digital Twins for Building
and Infrastructure Maintenance**

Optimizing Hydropower Operations: A Tailored Digital Twin Framework For Hydropower Plant

Weiwei Chen¹, Yang Su^{2*}

1) Ph.D. Lect/Assit. Prof, Bartlett School of Sustainable Construction, University College London, UK. Email: weiwei.chen@ucl.ac.uk

2) Ph.D., Postdoc., Bartlett School of Sustainable Construction, University College London, UK. Email: suyang0627@163.com

Abstract: Digital twins (DT) represent a critical methodology for digitising infrastructures, such as hydropower plants (HPP), enhancing their efficiency, sustainability, and competitive edge. However, despite the potential benefits, there is yet to be a consensus within industrial and academic spheres on the implementation process of digital twins for hydropower plants, a relatively uncommon type of infrastructure. This study explores the feasibility of using DT to provide digital solutions that enhance the flexibility and sustainability of hydropower plants. The paper examines the unique challenges associated with implementing digital twins in this context, including the high demand for water flow real-time updates, the complex logistical geographical dispersion, and environmental considerations. Subsequently, it reviews mainstream digital twin frameworks from other domains to establish foundational references. Building on this analysis, the study proposes a tailored digital twin development framework specifically tailored for hydropower plants. This proposed framework advances the theoretical understanding by integrating interdisciplinary research and domain-specific insights, offering a robust theoretical foundation for digital twin applications in hydropower settings. Practically, it delivers actionable guidelines and detailed strategies designed to facilitate the construction of digital twins in hydropower plants. By addressing domain-specific challenges and incorporating established best practices, the framework equips engineers and project managers with the tools necessary to enhance operational efficiency and sustainability. This comprehensive framework not only aids in navigating the complexities of hydropower projects but also sets a benchmark for future digital twin implementations.

Keywords: Digital twin, Hydropower plant, Implementation framework

1. INTRODUCTION

Digital twins (DT) have emerged as a pivotal technology for modernizing hydropower plants (HPP), bolstering their efficiency, sustainability, and competitive edge. This cutting-edge approach entails creating virtual models of physical hydropower facilities (Saif et al., 2024), allowing operators to simulate operations, anticipate potential issues, and implement solutions in a virtual space before applying them in the real world. The successful application of digital twins excelling in the building and construction sector where it aids in the lifecycle management of structures from design to demolition, ensuring optimal performance and maintenance.

However, as hydropower plants are a relatively rare type of infrastructure, the implementation of digital twins has not yet achieved widespread consensus within industry and academia. When developing digital twin frameworks, the HPP context presents unique challenges that set it markedly apart from conventional applications such as those found in building and construction (Cai et al., 2024). These distinctions arise primarily from the dynamic nature of water resources, the extensive regulatory environment, and the geographical dispersion of infrastructure—all intrinsic to the operation of hydropower facilities (Kumar et al., 2022). Unlike digital twins applied in more static or predictable settings, HPP digital twins must adeptly handle fluctuating natural inputs, adhere to stringent

environmental standards (Couto & Olden, 2018), and manage complex, dispersed assets effectively (Azimov & Avezova, 2022). This fundamental differentiation underscores the necessity of a specialized approach tailored specifically for hydropower applications, as general digital twin models fall short of addressing the nuanced requirements of this sector.

This paper proposes a digital twin-based framework specifically tailored for HPP infrastructure by summarizing the experience of other existing types of digital twin systems and the unique domain characteristics of hydropower plants. From a theoretical perspective, this study introduces a novel Event-Driven Digital Twin (EDDT) framework specifically designed for hydropower plants, addressing the limitations of traditional layered DT architectures in dynamic, real-time environments. By leveraging event-based processing and decoupling system components, the framework enhances operational flexibility, scalability, and responsiveness in managing complex HPP operations. Practically, this research offers actionable strategies and a detailed framework that can be directly applied by engineers and project managers to enhance the efficiency, sustainability, and operational flexibility of hydropower plants. The proposed framework equips stakeholders with the tools necessary to overcome domain-specific challenges, thereby setting a benchmark for future implementations of digital twin technologies in similar infrastructure projects.

2. METHODOLOGY

The core of this research mainly includes three parts: review and summarise the architecture and experiments of existing DT frameworks in other fields, identify unique characteristics that may be encountered in DT implementation specifically under HPP scenarios, and propose a tailored DT framework combining the above experiences for HPP scenarios.

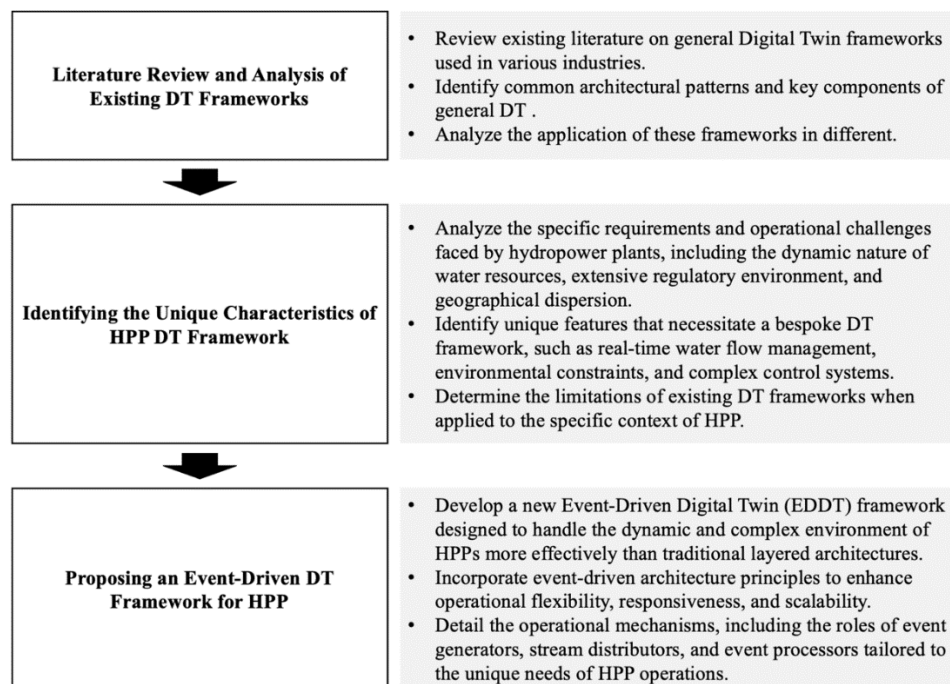


Figure 1. Research Methodology

As shown in Figure 1, the first step in the methodology involved a comprehensive review of existing literature on general DT frameworks, particularly those applied in industries such as manufacturing, construction, and energy management. The study focused on identifying common architectural patterns and key components of general DT systems, including data flow, modelling, simulation, and real-time control capabilities.

After identifying the limitations of general DT frameworks, the next step was to analyze the specific requirements and operational challenges faced by hydropower plants. HPPs have unique characteristics that distinguish them from other infrastructure types, such as dynamic water resource management, geographical dispersion, environmental and regulatory constraints (Zhang et al., 2014), complex operational requirements, integration with energy markets, and long asset lifecycles. This analysis highlighted the need for a more flexible and responsive DT framework that can handle these challenges, particularly in dynamic and highly regulated environments.

Based on the insights from the literature review and the specific requirements of HPPs, this study proposes an Event-Driven Digital Twin framework (EDDT) tailored to the unique needs of hydropower plants. The event-driven framework is particularly suited for hydropower plants because their operations are influenced by highly dynamic, real-time factors such as water flow and environmental conditions, which require immediate responses that traditional, sequential data processing frameworks cannot deliver with the same level of flexibility and speed (Cai et al., 2024).

3. RESULT

3.1 Existing DT framework

The mainstream general DT framework architecture in current research consists of the following layers: Physical Layer, Data Layer, Modeling Layer, Communication Layer, Analysis Layer, Application Layer, Security Layer, and Feedback & Control Layer (Zheng et al., 2019; Tuhaise et al., 2023; Aheleroff et al., 2021), as shown in Figure 2.

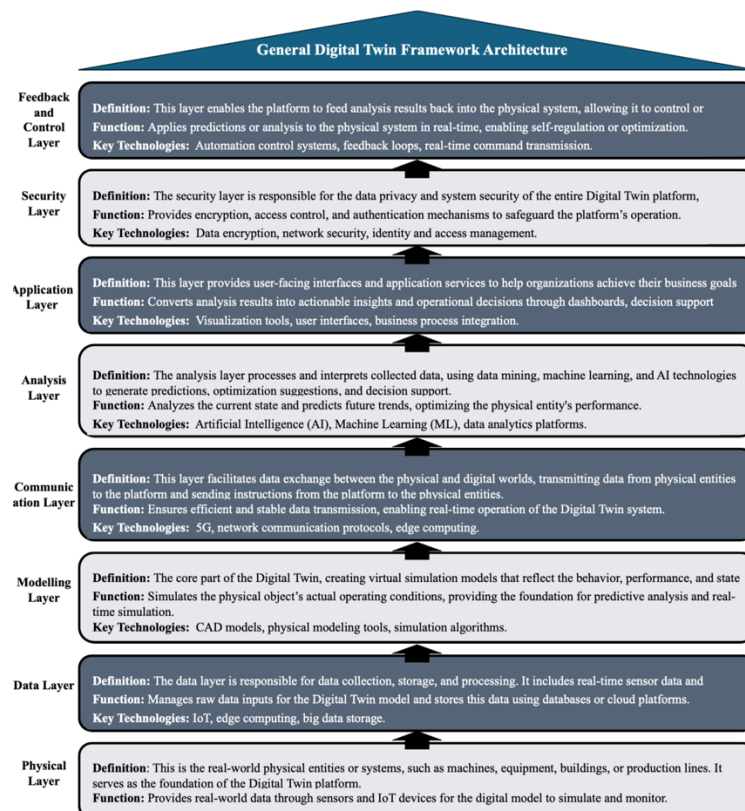


Figure 2. General digital twin framework architecture

The general digital twin framework architecture consists of multiple interconnected layers, each building upon the other to create a cohesive system that mirrors real-world physical entities in the digital space (Menon et al., 2023). The relationships between these layers are crucial to ensure the successful

functioning of the digital twin, from data collection to real-time control and optimization.

Physical Layer forms the foundation of the system. It consists of the real-world physical assets, such as machines, infrastructure, or equipment, that are mirrored digitally. This layer is critical as it supplies the real-time data needed to accurately represent the physical world. The physical layer interacts directly with the **Data Layer** through sensors and IoT devices.

Data Layer collects and processes raw data from the physical assets. This layer ensures that the information coming from the physical world is gathered, structured, and stored in a way that can be further utilized by other layers. The data layer relies on **communication technologies**, such as the Internet of Things (IoT) and edge computing, to ensure that data flows efficiently from the physical world into the system.

Once the data is collected, it moves into the **Modeling Layer**, which uses this data to build virtual representations (models) of the physical entities. These models reflect real-time conditions and behaviours, allowing for simulation and analysis (Tuhaise et al., 2023). The accuracy of the models directly depends on the quality of the data received from the previous layer, making this layer deeply interconnected with both the **Data** and **Analysis Layers**.

The **Communication Layer** is responsible for maintaining the data flow between the physical and digital worlds. This layer ensures that data moves from the physical assets to the modelling and analysis systems in real-time, and conversely, enables the digital twin system to send instructions or feedback to the physical assets. Reliable communication is essential for the synchronization of the real and digital environments, making it a critical layer for real-time operations and analysis.

The **Analysis Layer** plays a vital role in processing the collected data and running simulations on the models created in the modelling layer. This layer utilises advanced technologies such as Artificial Intelligence (AI) and Machine Learning (ML) (Haung et al., 2021) to generate insights, predictions, and optimisation suggestions. These insights are vital for making informed decisions about how to optimize or adjust physical operations.

These insights and analysis results are presented to users through the **Application Layer**. This layer serves as the interface between the digital twin system and the end-users, providing them with actionable insights and decision-support tools. Through dashboards, visualizations, and user-friendly applications, this layer ensures that decision-makers can easily access and act on the data and insights generated by the digital twin.

The **Security Layer** is fundamental in ensuring that the entire system remains secure and protected from potential threats. It is responsible for safeguarding data privacy, controlling access to the system, and ensuring the integrity of the platform. The security layer spans across all other layers, ensuring that data transmission, storage, and user access are protected throughout the digital twin's operation.

Finally, the **Feedback and Control Layer** closes the loop by allowing the digital twin system to feed the analysis results and predictions back into the physical system. This layer plays a crucial role in enabling the system to autonomously adjust or optimise physical operations in real time based on the insights generated by the analysis layer.

3.2 Unique domain characteristics of HPP digital twin framework

In the HPP digital twin framework, several unique domain characteristics set it different from general digital twin applications. These distinct features arise due to the specific operational, environmental, and logistical challenges associated with hydropower plants, as shown in Figure 3.

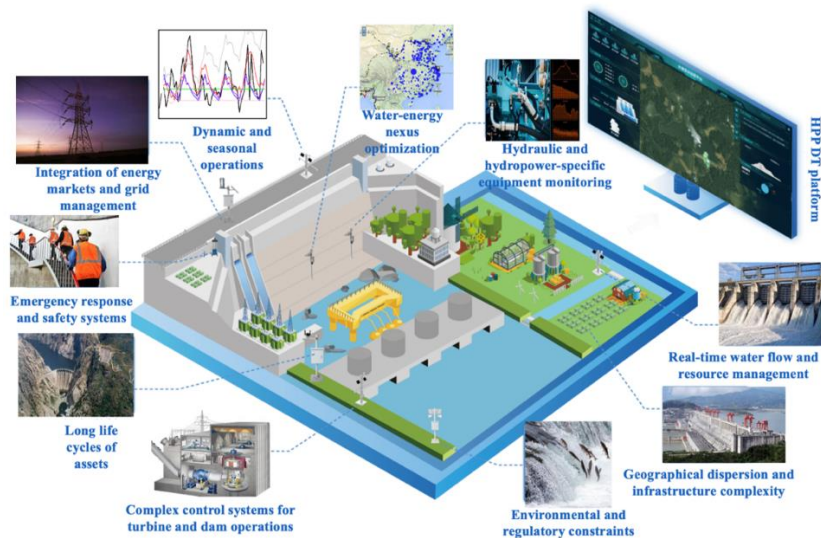


Figure 3. Unique domain characteristics of HPP digital twin framework

Real-time water flow and resource management. HPPs depend on natural water flow for energy generation, meaning the real-time monitoring and prediction of water levels, flow rates, and reservoir conditions are essential. Unlike other DT systems, the variability of natural resources (water availability, seasonal changes, etc.) directly influences HPP operations. The DT must be capable of integrating and forecasting hydrological data to optimise energy production while managing ecological impacts like fish migration and flood risks.

Geographical dispersion and infrastructure complexity. Hydropower plants are often located in remote areas, and the infrastructure (such as dams, turbines, and reservoirs) is typically spread across large geographical regions. The HPP DT must be able to handle geographically dispersed data sources, making it necessary to integrate with IoT and edge computing for real-time remote monitoring. This contrasts with DTs used in more localized or factory-based environments.

Environmental and regulatory constraints. Hydropower operations are subject to strict environmental regulations, particularly concerning water usage, wildlife conservation, and emissions reductions (Khoiyanbham, 2021). The HPP DT needs to incorporate environmental monitoring, compliance reporting, and impact assessments as part of its core functionality, making it more complex than DTs in industries without such stringent environmental constraints.

Complex control systems for turbine and dam operations. HPPs rely on complex control systems to regulate water flow through turbines, manage reservoir levels, and operate gates and spillways (Bernardes et al., 2022). The HPP DT needs to simulate and control these systems in real time, ensuring operational efficiency and safety. The high level of physical system interaction and safety-critical operations are more intense compared to other industries where such real-time, high-stakes control may not be as central.

Long life cycles of assets. Hydropower infrastructure typically has very long operational life cycles, often extending to several decades. The DT must be designed to monitor ageing infrastructure, predict maintenance needs over long periods, and ensure long-term operational efficiency. This requires robust predictive maintenance capabilities to prevent costly downtime or failures, a feature that is not as critical in industries with shorter asset life cycles.

Emergency response and safety systems. HPPs often operate under potentially hazardous conditions, such as dam safety risks (Furtado & Ravena, 2019), flood scenarios (Nguyen-Tien et al., 2018), or

mechanical failures. The HPP Digital Twin must integrate sophisticated emergency response simulations and safety protocols to minimize risks. It needs to run scenarios to predict failures in dam infrastructure or turbine breakdowns, ensuring that safety measures are deployed swiftly. This feature is more critical compared to industries with less severe consequences of system failures.

Integration of energy markets and grid management. HPPs often need to balance energy production with demand (Algarvio et al., 2020), particularly in integrated energy markets where hydropower is a key component of renewable energy portfolios. The HPP DT must integrate with energy grids and market platforms to ensure efficient energy dispatch and grid stability.

Dynamic and seasonal operations. The performance of hydropower plants is highly influenced by dynamic, seasonal factors such as rainfall, snowmelt, and droughts (Carpentier et al., 2017), which can vary significantly over time. The HPP DT must have sophisticated predictive capabilities to model these seasonal changes and optimize operations. This differs from DTs in industries with more predictable operating environments, like manufacturing.

Water-energy nexus optimization. Hydropower plants sit at the intersection of water management and energy production, making the efficient use of water critical not just for power generation but also for flood control (Nguyen-Tien et al., 2018), irrigation, and drinking water supplies (Kotulla et al., 2022). The HPP DT must balance these competing demands, optimizing water use for power while ensuring that other water resource needs are met. This adds an additional layer of complexity that general DTs, focused only on energy production, do not need to address.

Hydraulic and hydropower-specific equipment monitoring. HPPs involve specialized equipment such as turbines, penstocks, spillways, and transformers that are unique to hydropower generation (Kumar & Saini, 2022). The Digital Twin must be tailored to monitor, simulate, and predict the performance and maintenance needs of these hydraulic systems and power generation equipment.

3.3 Tailored digital twin framework for HPP

3.3.1 Event-driven architecture

The operational environment of hydropower plants is highly dynamic and complex. Traditional layered architectures, while well-structured, may lack the flexibility and responsiveness needed for real-time operations. The EDDT architecture enhances HPP efficiency by rapidly processing real-time events, such as changes in water flow or system failures, without delays in data transmission. It allows for immediate monitoring and adaptation of operations, optimizing energy output and ensuring system safety. It also promotes modularity and maintainability by reducing dependencies between components, enabling independent updates without affecting system stability.

Specifically, four key reasons for choosing event-driven architecture include **1) Real-time Data Processing and Responsiveness.** Hydropower stations rely on real-time monitoring of parameters like water flow and levels. Event-driven architecture allows rapid responses to changes, automatically adjusting operations for optimal energy output and resource management. It also handles events and alerts from sensors, deploying immediate measures such as adjusting water levels during emergencies. **2) Adaptability to Dynamic Operating Environments.** The hydropower station's operation is influenced by seasonal changes and environmental factors. Event-driven architecture adjusts predictions and operations in real-time, optimizing maintenance and resource strategies. **3) System Decoupling.** This architecture minimizes dependencies between system components, allowing independent updates and maintenance, and enhancing overall reliability and flexibility. **4) Facilitating Innovation and Integration.** It supports integration with other renewable energy sources and systems, enabling innovation in applications such as demand response and market operations.

3.3.2 Operational mechanisms

The implementation of the Event-Driven Architecture (EDA) in the Hydropower Plant Digital Twin framework encompasses several critical components and processes, all working in harmony to ensure efficient, flexible operations and real-time management. The following describes the core elements and their interactions within the updated architecture:

Event Producer: These are the sensors deployed throughout the physical HPP infrastructure. They monitor various physical parameters such as water level, soil moisture, flow rate, rainfall, and environmental conditions like temperature and wind speed. This network of advanced sensors, including flow meters, water level gauges, osmotic pressure sensors, flood monitors, and emergency detection devices, continuously collects real-time data and generates events when predefined thresholds or anomalies are detected. Internet of Things (IoT) devices form the backbone of this network, ensuring continuous data generation and transmission.

Stream Distributor: The Stream Distributor is a central element in the framework, responsible for receiving the raw events generated by the sensors and distributing them to the appropriate consumers (event processors). It functions as a message broker or event bus, facilitating efficient event flow between sensors (event producers) and consumers (event processors). This component ensures that the right data reaches the correct processor based on predefined criteria.

Event Processors (Consumers): Each event processor (consumer) specializes in processing specific types of events. In the updated framework, the key event processors include: 1) **Power Grade Processor:** Monitors power output, optimizing generation based on real-time hydrological data. 2) **Environmental Processor:** Analyzes environmental data (e.g., rainfall, water quality) and ensures regulatory compliance. 3) **Cyber Security Processor:** Identifies and mitigates cyber threats to the HPP infrastructure. 4) **Water Resource Processor:** Balances the water resource needs of the plant with external demands like irrigation and drinking water. 5) **Hydraulic Equipment Processor:** Monitors and ensures the health of turbines, gates, and other critical mechanical components. 6) **Weather Prediction Processor:** Analyzes weather data to forecast conditions that may affect HPP operations. 7) **Market Integration Processor:** Facilitates real-time integration with energy markets to adjust power output based on demand and pricing. 8) **Operational Safety Processor:** Ensures that the system remains in safe operating conditions by monitoring infrastructure risks like dam stability or flood hazards.

These processors leverage processing software and other microservices to classify, filter, and analyze events, ensuring timely and appropriate responses. 1) **Decision Support System:** Based on the output from the event processors, this system generates operational commands and suggestions for adjustments, such as regulating sluice gates, adjusting power generation, or initiating emergency response plans. The Decision Support System integrates Artificial Intelligence (AI) and Machine Learning (ML) techniques to optimize decision-making and enhance predictive maintenance capabilities. It communicates its findings and suggestions with both the HPP DT platform and key stakeholders/managers for further decisions. 2) **Feedback and Control:** This component executes the operational commands generated by the Decision Support System. Automated control systems adjust physical equipment, such as turbines and sluice gates, based on real-time data and AI-driven insights. This ensures fast, accurate execution of necessary operations and system adjustments, maintaining optimal performance and safety standards.

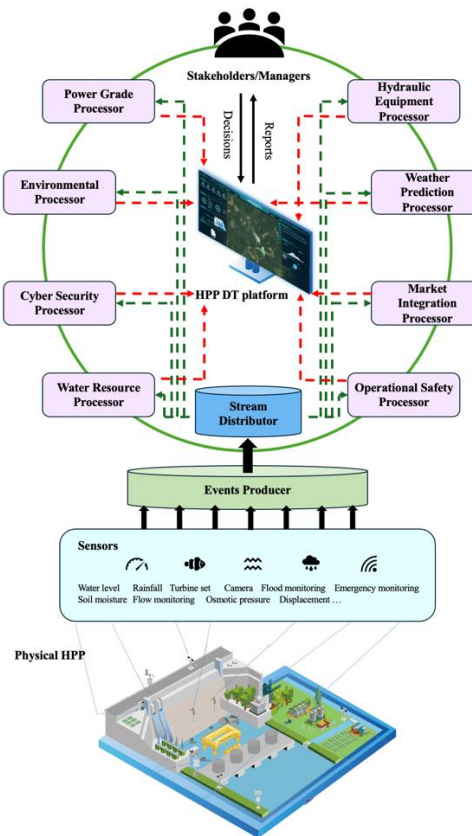


Figure 4. HPP digital twin framework architecture

3.3.3 Core processors

Power Grade Processor. This processor is dedicated to managing and optimizing the power generation aspects of the HPP DT. It processes real-time data concerning hydrological conditions and power output requirements. It evaluates the efficiency of power generation processes, predicts future power supply based on real-time environmental data, and adjusts operational parameters to maximize energy output while ensuring the longevity of physical assets.

Environmental Processor. This processor focuses on processing and responding to environmental data collected from various sensors monitoring weather, water quality, and ecological conditions. This processor uses data to ensure compliance with environmental regulations and to adjust operations to minimize environmental impact. It also predicts potential environmental risks like flooding and adjusts operational strategies accordingly. Utilizes AI-driven models to simulate and predict environmental impacts, incorporating real-time data from IoT devices to facilitate rapid response to environmental changes and alerts.

Cyber Security Processor. It safeguards the integrity and security of the HPP DT system by monitoring, detecting, and responding to cybersecurity threats and vulnerabilities. Manages data security protocols, oversees network security operations, and implements protective measures against potential cyber-attacks and breaches. It also ensures compliance with data protection laws and industry standards. Employs advanced cybersecurity technologies including intrusion detection systems, firewall management, and encryption protocols to protect data transmissions and system operations.

These core processors play a vital role in ensuring the HPP DT system operates efficiently, safely, and sustainably. Each processor is equipped with specific functionalities that address different aspects of the digital twin's operations, from power management to environmental monitoring and cybersecurity,

forming a robust foundation that supports the overall integrity and effectiveness of the event-driven architecture.

4. DISCUSSION

4.1 Expected outcomes

The implementation of the EDDT framework for hydropower plants is expected to deliver several key outcomes, improving both operational efficiency and adaptability compare with traditional layer structured DT frameworks: 1) Increased Flexibility: Enables real-time responses to environmental and operational changes, optimizing water management and power generation. 2) Real-Time Responsiveness: Enhance real-time monitoring and control, allowing immediate reactions to key operational events, preventing failures, and optimizing processes. 3) Scalability: Support a growing number of sensors and data inputs without increasing complexity, ensuring robust system growth. 4) Enhanced Predictive Capabilities: Incorporate AI/ML to improve predictive maintenance, reducing downtime and extending equipment lifespan. 5) Better Resource Allocation: Dynamically prioritize resources, ensuring swift responses to high-priority events while efficiently managing routine tasks. 6) Environmental Compliance: Ensure adherence to environmental standards through real-time monitoring of water management and conservation. 7) Improved Safety: Enhance safety through real-time monitoring of potential hazards, enabling faster response to emergencies such as dam safety or equipment failure. In summary, the EDDT framework will boost the efficiency, sustainability, safety, and scalability of hydropower plant operations.

4.2 Potential Challenges

While the EDDT framework offers significant benefits, it also presents challenges in complex hydropower environments: 1) Increased System Complexity: Managing asynchronous event flows, ensuring consistency, and avoiding duplication or loss becomes harder as the system scales. This requires advanced event routing, logging, and monitoring to maintain integrity. 2) Event Overload: High-frequency events, such as sudden weather changes, can overwhelm the system. Effective filtering and prioritization, including complex event processing (CEP), are necessary to prevent delays in handling critical issues. 3) Difficult Debugging and Tracing: Debugging in event-driven systems is more complex than in traditional layered architectures due to concurrent and independent event processing. Advanced logging and monitoring tools are needed to trace and resolve issues, increasing maintenance demands.

4.3 Case scenarios

The proposed EDDT framework will be tested and applied in the Kremasta and Ilarion Hydropower Plants in Greece. The Kremasta HPP is the largest Hellenic hydroelectric power plant. It is located on Acheloos river/basin in West Continental Greece, and it is used for hydropower production and flood control. It is an earth fill power plant with total installed capacity 437 MW produced by four Francis turbines with a nominal power of 109 MW each. The HPP's mean annual production is 848 GWH and the height of waterfall is 132 m. And the Ilarion is a relatively new (operated commercially for the first time in 2014) hydro power plant located on Aliakmon River in Western Macedonia. It is used for hydropower production, irrigation and water supply and it has a total installed capacity 153 MW that is produced by two Francis turbines with a nominal power of 76 MW each. The HPP's mean annual production is 320 GWH and the height of waterfall is 104 m.

5. CONCLUSIONS

This study proposed an EDDT framework specifically tailored for hydropower plants, aiming to enhance operational flexibility and responsiveness. The framework effectively leverages event-driven architecture to allow real-time system adjustments in response to environmental and operational changes.

This adaptation is crucial for managing the dynamic and complex nature of water resources inherent in hydropower systems. While the EDDT framework promises significant improvements in efficiency and sustainability, its implementation is challenging due to the increased complexity and management demands. Future research should focus on refining the integration and operational processes to mitigate these challenges. As part of the D-HYDROFLEX project, the proposed EDDT framework will be tested and applied in real-world conditions, specifically in operational the Kremasta and Ilarion Hydropower Plants in Greece (D-HYDROFLEX, 2024). This will allow us to validate the expected outcomes and address potential challenges in a practical setting, contributing valuable insights to future digital twin implementations in the hydropower sector. Overall, the proposed digital twin framework sets a foundational approach for future advancements in the digital management of renewable energy systems.

ACKNOWLEDGMENTS

The research is supported and funded by the Horizon Europe project (Project 101122357 — D-HYDROFLEX).

REFERENCES

- Aheleroff, S., Xu, X., Zhong, R. Y., & Lu, Y. (2021). Digital twin as a service (DTaaS) in industry 4.0: an architecture reference model. *Advanced Engineering Informatics*, 47, 101225.
- Algarvio, H., Lopes, F., & Santana, J. (2020). Strategic operation of hydroelectric power plants in energy markets: A model and a study on the hydro-wind balance. *Fluids*, 5(4), 209.
- Azimov, U., & Avezova, N. (2022). Sustainable small-scale hydropower solutions in Central Asian countries for local and cross-border energy/water supply. *Renewable and Sustainable Energy Reviews*, 167, 112726.
- Bernardes Jr, J., Santos, M., Abreu, T., Prado Jr, L., Miranda, D., Julio, R., ... & Bastos, G. S. (2022). Hydropower operation optimization using machine learning: A systematic review. *AI*, 3(1), 78-99.
- Cai, Z., Wang, Y., Zhang, D., Wen, L., Liu, H., Xiong, Z., ... & Feng, R. (2024). Digital Twin Modeling for Hydropower System Based on Radio Frequency Identification Data Collection. *Electronics*, 13(13), 2576.
- Carpentier, D., Haas, J., Olivares, M., & De la Fuente, A. (2017). Modeling the multi-seasonal link between the hydrodynamics of a reservoir and its hydropower plant operation. *Water*, 9(6), 367.
- Couto, T. B., & Olden, J. D. (2018). Global proliferation of small hydropower plants—science and policy. *Frontiers in Ecology and the Environment*, 16(2), 91-100.
- D-HYDROFLEX: Digital solutions for hydroelectric plants. (2024). <https://d-hydroflex.eu/>.
- Furtado Louzada, A., & Ravena, N. (2019). Dam safety and risk governance for hydroelectric power plants in the Amazon. *Journal of Risk Research*, 22(12), 1571-1585.
- Huang, Z., Shen, Y., Li, J., Fey, M., & Brecher, C. (2021). A survey on AI-driven digital twins in industry 4.0: Smart manufacturing and advanced robotics. *Sensors*, 21(19), 6340.
- Khoiyangbam, R. S. (2021). Wetlands in Loktak: Issues and challenges of merging Wildlife conservation and Hydropower generation—An Overview. *International Journal of Lakes and Rivers*, 14(2), 223-236.
- Kotulla, M., Goño, M., Goño, R., Vrzala, M., Leonowicz, Z., Kłosok-Bazan, I., & Boguniewicz-Zabłocka, J. (2022). Renewable energy sources as backup for a water treatment plant. *Energies*, 15(17), 6288.
- Kumar, K., & Saini, R. P. (2022). A review on operation and maintenance of hydropower plants. *Sustainable Energy Technologies and Assessments*, 49, 101704.

- Menon, D., Anand, B., & Chowdhary, C. L. (2023). Digital twin: exploring the intersection of virtual and physical worlds. *IEEE Access*, 75152 - 75172.
- Nguyen-Tien, V., Elliott, R. J., & Strobl, E. A. (2018). Hydropower generation, flood control and dam cascades: A national assessment for Vietnam. *Journal of hydrology*, 560, 109-126.
- Saif, W., RazaviAlavi, S., & Kassem, M. (2024). Construction digital twin: a taxonomy and analysis of the application-technology-data triad. *Automation in Construction*, 167, 105715.
- Tuhaise, V. V., Tah, J. H. M., & Abanda, F. H. (2023). Technologies for digital twin applications in construction. *Automation in Construction*, 152, 104931.
- Zhang, J., Xu, L., Yu, B., & Li, X. (2014). Environmentally feasible potential for hydropower development regarding environmental constraints. *Energy Policy*, 73, 552-562.
- Zheng, Y., Yang, S., & Cheng, H. (2019). An application framework of digital twin and its case study. *Journal of ambient intelligence and humanized computing*, 10, 1141-1153.

Proposal for a Simple Inspection Tool using HoloLens2

Tomohiro Mizoguchi¹

1) Ph.D., Professor, Department of Informatics and Data Science, Faculty of Engineering, Sanyo-Onoda City University, Yamaguchi, Japan. Email: mizoguchi@rs.socu.ac.jp

Abstract: The importance of maintenance is growing as civil engineering structures age in Japan. For proper maintenance, it is important to carry out regular inspections at high frequencies and to accumulate the data in a way that makes it easy to use later. In recent years, with the spread of 3D scanning technology, it has become an effective method to use 3D models obtained by 3D scanning of the target structure as a base, and to associate electronic inspection information such as photos with the relevant locations and manage them for a long term. By providing location information on the 3D model to the inspection information in this way, it has the advantage that it becomes easier to grasp the spatial location of deterioration and damage, and later searches can be made intuitive and easy. In this paper, we propose a simple inspection tool using the head-mounted device HoloLens 2. The developed tool can perform 3D scanning while moving around the structure based on SLAM. And when deterioration or damage is found, images can be captured with the device's camera. In addition, the corresponding location on the 3D scanned model can be specified by simply pointing to the relevant location in the real structure with a hand gesture, and inspection information such as images and documents acquired can be registered here. Experiments are conducted on an actual bridge with this tool to demonstrate that entire inspection process, including 3D scanning, taking photos, and registering the photos on a 3D model, can be performed with very simple and intuitive operations at the inspection site. In addition, the accuracy of the constructed 3D model was evaluated by local plane fitting and comparing it with the TLS point cloud using ICP alignment. Due to the nature of the device, the scanning distance of the developed tool is short, so it is limited to relatively small structures, but it is suitable for simple inspections.

Keywords: Infrastructure inspection, 3D scanning, HoloLens, BIM/CIM, Hand Gesture.

1. INTRODUCTION

The importance of maintaining civil engineering structures that were built during the period of high economic growth is increasing as they deteriorate in Japan. For example, there are about 730,000 road bridges, of which about 65% are managed by relatively small local government. Although the number of structures is huge, unfortunately there is a shortage of engineers involved in management, and the budget is limited, so regular and daily inspections are sometimes not carried out sufficiently. For maintenance, it is important to carry out regular inspections at high frequency, record information such as images that capture damage and deterioration, and store this information for the long term in a way that is easy to use later. However, photos that record deterioration and damage generally only show a small part of the structure and do not include location information. Therefore, in recent years, it is becoming common to associate related image and other information with the corresponding parts of 3D models such as point clouds and BIM/CIM, and manage them centrally (Sacks et al., 2018; Shim et al., 2019). By providing location information on the 3D model to inspection information, it is possible to easily grasp deterioration and damage visually and spatially. However, since there are almost no 3D models for old infrastructures, it is necessary to scan the actual infrastructures in 3D and directly use the obtained point cloud data, or to use BIM/CIM created from the point cloud as the base for management. Many previous studies have been reported on the use of 3D models for the maintenance management of infrastructures, but most of them have targeted large structures (Mizoguchi et al., 2023). As a result, there are few examples of the development of inexpensive and efficient inspection tools for relatively small structures.

1.1 3D Scanning for Infrastructure Maintenance

With the rapid spread of 3D scanning technology, it has also become widely used for infrastructure maintenance and management. The well-known terrestrial laser scanner (TLS) has the advantage of being able to obtain high-precision, high-quality point clouds. However, since it is not possible to scan the entire object from a single location, scans are usually taken from multiple locations and then processed to combine the data. This requires movement and the installation of a tripod for each scanning, which has the disadvantages of being inefficient in scanning and not being able to scan in inaccessible places (Rashidi et al., 2020). On the other hand, with the spread of SLAM-based 3D mobile laser scanning (MLS) technology, it has become possible to efficiently obtain 3D point clouds for civil infrastructures of tens to hundreds of meters in size (Yan and Hajjar, 2021; Blaskow and Maas, 2024). Although SLAM-based methods are inferior to TLS in terms of accuracy, they have the advantage of being highly efficient in scanning and being able to obtain point clouds with few deficiencies. In addition, if they are installed on a UAV, scanning can be made even in inaccessible places. However, MLS is

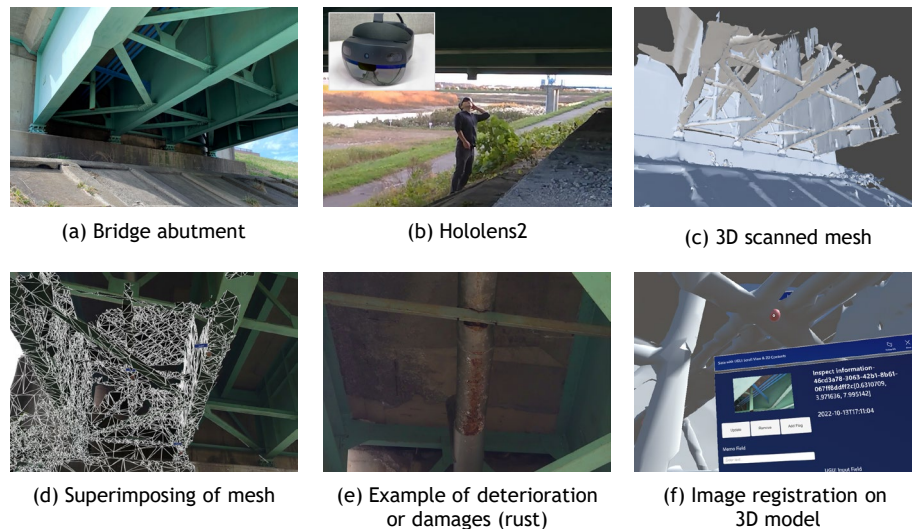


Figure 1. Overview of our developed tool

expensive, and UAV flights require the cooperation of experts, making it difficult to use for maintenance. In addition, in both TLS and MLS, the user must interactively indicate the corresponding location on the 3D model for each inspection image of deterioration and damages to register them on it. This is inefficient and there are cases where the correct position cannot be accurately specified. In contrast, 3D reconstruction processing from images using structure from motion (SfM) can obtain high-quality point clouds or 3D mesh with rich texture with only a relatively inexpensive camera and commercial software (Mohammadi et al., 2021; Tang et al., 2024). In addition, it has the advantage that information on the relative position and orientation on the reconstructed 3D model can be obtained for each captured image, making it easy to register the inspection results using images on the 3D model. However, the disadvantages are that it takes time to capture images and the reconstruction processing.

In particular, in recent years, in addition to mobile devices such as iPads and iPhones, head-mounted 3D scanning devices have also appeared, and 3D scanning is becoming increasingly popular. With such devices, even general users without specialized knowledge can easily perform 3D scanning. Among them, HoloLens 2 is equipped with a wide range of functions, including not only 3D scanning and camera photography, but also augmented reality (AR), internet communication, and gesture recognition (Ungureanu et al., 2018). Therefore, it is expected that a new inspection tool can be developed by making full use of these functions.

1.2 Research Purpose and Overview of Proposed Tool

In this paper, we propose a simple inspection tool using the head-mounted device HoloLens 2. In this report, we implemented core functions essential for maintenance management, such as 1) constructing a 3D model using a depth sensor, 2) photographing deterioration and damage using a camera, and 3) registering the images to a 3D model using hand gestures. We also evaluated its effectiveness through experiments simulating inspections of actual bridges. Furthermore, to evaluate the accuracy of 3D measurement using HoloLens 2, we performed plane fitting and distance evaluation to point cloud for local accuracy evaluation, and point-to-point distance evaluation based on global ICP alignment with the TLS point cloud. Due to the characteristics of the device, this tool is not suitable for large-scale structures, but it can be used to perform simple inspections efficiently and at low cost for relatively small structures such as ditch bridges. And this can be also used for narrow areas where TLS scanning and UAV flight are difficult. The developed tool is a combined implementation of existing 3D scanning and hand-gesture functions, but we believe it demonstrates its potential as a future human-machine collaborative maintenance which contribute to efficient and low-cost inspection. The features and advantages of this tool are summarized as follows.

1) Creating 3D models and registering inspection information at the inspection site

Conventionally, the creation of 3D models and the registration of inspection images to the 3D models were carried out separately in the office after the on-site inspection. In contrast, with this tool, the entire process of creating 3D models using SLAM, camera imaging of deteriorated or damaged areas and registering the images to the 3D model can be carried out at the inspection site. This reduces the amount of work required in the office and makes inspections more efficient.

2) Obtaining location information using hand gestures

Even in places where GPS does not reach or where it is difficult to install QR codes or RFID, the corresponding location information on the 3D model can be easily obtained by simply indicating the relevant location on the actual structure using hand gestures.

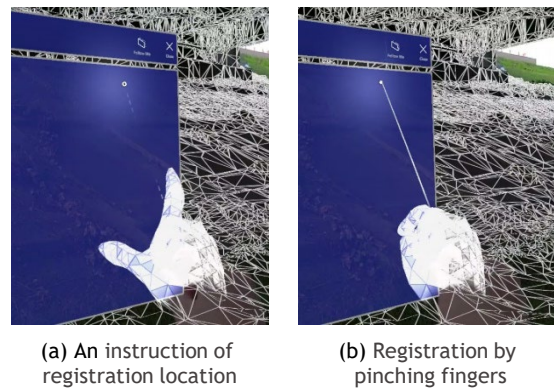


Figure 2. Image registration by hand gesture



Figure 3. Experimental locations for image capture and registration

3) Hands-free inspection

By using a head-mounted device, inspections can be carried out in a hands-free state. And this makes it easier to access areas with poor footing or narrow areas and to inspect areas that were difficult to do using conventional methods.

2. OVERVIEW OF OUR PROPOSED TOOL

An overview of the developed tool is shown in Figure 1. With this tool, for example, the narrow section of a bridge (Figure 1(a)) is targeted. When an inspector moves around the area while wearing the device (Figure 1(b)), 3D scanning of the surrounding environment, including the target object, are taken at any time as the inspector moves (Figure 1(c)). The acquired 3D model and the real environment are superimposed so that their positions match and are displayed on a see-through display (Figure 1(d)). If the inspector finds deterioration or damage in the structure while moving, a photo is taken with the camera mounted on the device (Figure 1(e)). Furthermore, immediately after taking the photo, the inspector can register it at the corresponding position on the 3D model by indicating the damaged area in the real environment with a hand gesture (Figure 1(f)). In the following, we will provide an overview of the SLAM-based 3D scanning and hand gesture-based photo position acquisition functions that are mainly used in this tool.

2.1 3D scanning by SLAM

HoloLens 2 is equipped with a variety of sensors. In particular, the Time-of-Flight (ToF) depth camera can measure depth in the form of distance images. Four grayscale cameras also provide robust self-localization (Weinmann, 2020). During inspections, users can simply wear HoloLens 2 on their head and move around the surroundings to perform 3D scanning of the area within the camera's field of view and obtain a 3D mesh of the target in real time. Compared to conventional TLS and MLS, the scanning distance is a maximum of 3.5 m, so the target is limited to small structures. However, the operation is simple, so it can be used by general users as well.

2.2 Specification of image registration position using hand gestures

When an inspector finds deterioration or damage, a picture is taken with the camera mounted on the HoloLens. Since the image is taken of a local area, it does not have global position information, and the location of the image cannot be identified later. Therefore, immediately after taking the image, the image is registered at the corresponding location on the 3D mesh being constructed. For this registration process, when the location of

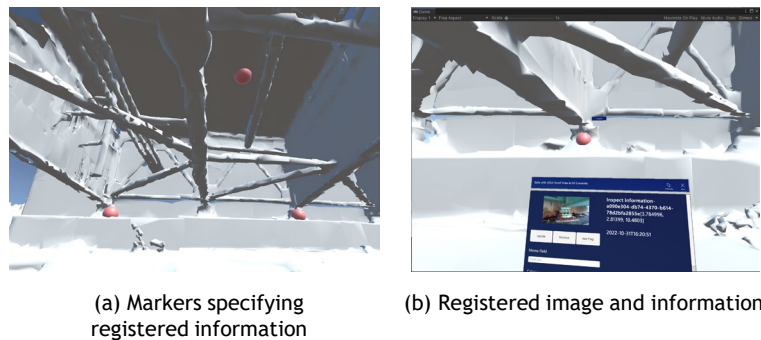


Figure 4. Display of inspection information on PC



Figure 5. Display of 3D model and inspection information on HoloLens 2

the structure to be photographed is specified with a finger as shown in Figure 2(a), the hand tracking function draws the point and the curve leading to it, called a hand ray, on the transparent display. The user can specify the registration location of the image while checking them. Once the position is determined, registration is completed by pinching the tips of the thumb and index finger together as shown in Figure 2(b). In the development tool, a marker is displayed at the registration location as shown in Figure 4(a). In addition, as shown in Figure 4(b), when the marker is clicked after registration, a form is opened on which the registered image can be viewed. It is also implemented so that a simple memo can be written in the form later. Thanks to the graphical guide, even general users can specify the position intuitively and easily.

3. RESULTS

3.1 Test site and overview of experiment

The experiment was conducted at the abutment of a steel girder bridge shown in Figure 1(a). The area around this abutment is sloped, making it difficult to move around with your hands occupied. And GPS is also difficult to reach, making it suitable for performance verification of the developed tool. In the experiment, 3D scanning were taken while moving around the area, and multiple images were taken of areas showing deterioration or damage, and location information for each was given using hand gestures, which were then registered on the measured 3D model.

3.2 Performance evaluation of our developed tool

While scanning the target site, photos of the six locations were taken as shown in Figure 3 and they were registered on the 3D model. The total time required was about 10 minutes, and the obtained 3D mesh was 14.1 MB in size with 84,078 polygons. We confirmed that the operations of scanning, photographing, and position specification were all very simple and could be accurately registered at the location intended by the user on the 3D model. The resolution of the captured images was 3,904 x 2,196 pixels. Figure 5 shows the results of overlaying the entire scanning mesh and structures, as well as the registration results, displayed on HoloLens 2.

3.3 Accuracy evaluation of 3D scanning by HoloLens2

We evaluated the accuracy of 3D scanning using HoloLens2. For the evaluation, we obtained a dense point cloud shown in Figure 6(a) according to the procedure described in (Ungureanu et al., 2018). The number of points was 50,646,562. In the evaluation, we performed two evaluations: local accuracy evaluation based on plane fitting and global evaluation by comparing with TLS point cloud.

First, to evaluate the local accuracy, we cut out a partial point cloud in the planar area of the point cloud shown in red in Figure 6(a), fitted a plane using the least squares method, and calculated the distance to each point

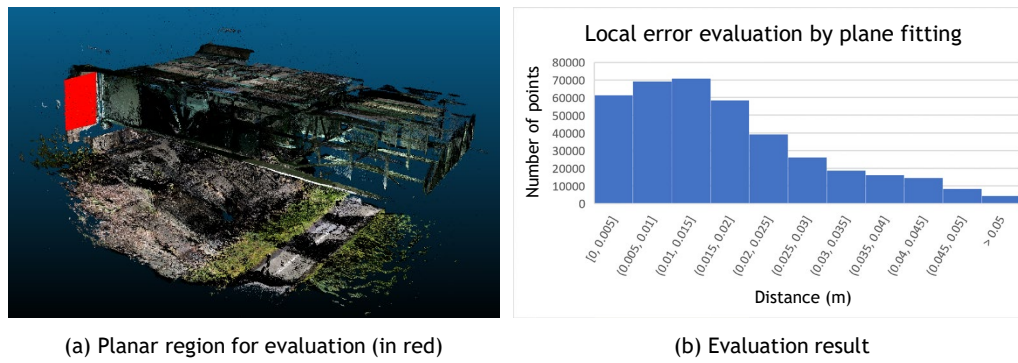


Figure 6. Local error evaluation by plane fitting

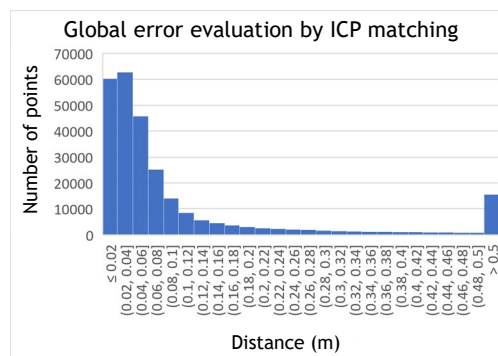
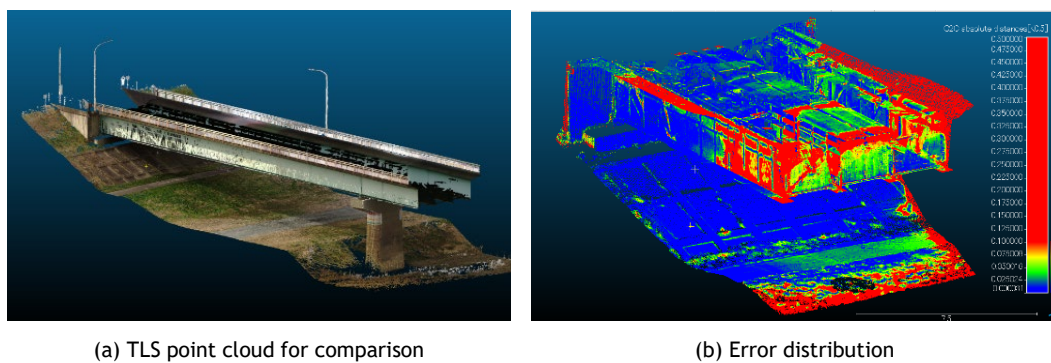


Figure 7. Global error evaluation by ICP matching with TLS point cloud

as the scanning error. Figure 6(b) shows the error distribution of the point cloud. The mean absolute error was 1.71 cm, and about 90% of the points were within an error of 3 cm, so sufficient accuracy was obtained for simple scanning.

Next, to evaluate the global accuracy, we scanned the same area of the bridge using a terrestrial laser scanner. Figure 7(a) shows the point cloud data. The laser scanner used was a FARO Focus 3D Scanner. Its maximum range is 150m and scanning accuracy is $\pm 1\text{mm}$ when scanning is between 10m and 25m. The overlapping parts are cut out from this point cloud, and the closest point distance between the point clouds is calculated after alignment using ICP (Iterative Closest Point) method. The number of points is 16,923,998. For global accuracy evaluation, the point clouds were sampled at 1cm intervals, and the closest point distance was calculated. The color-coded results are shown in Figure 7(b), and the histogram is shown in Figure 7(c). In Figure 7(b), points over 5cm are shown in red. Here, the average point distance is 4.69cm, and approximately 90% of the points are within 10cm, which is sufficient accuracy. The graph also shows points over 50cm, but this is because the scanning distance of HoloLens 2 was insufficient and corresponding points between the data could not be found.

4. CONCLUSIONS AND FUTURE WORKS

In this study, we developed a simple inspection tool using the head-mounted device HoloLens 2 and verified its performance. Through various experiments, we confirmed that the developed tool can perform a series of inspection processes, such as 3D scanning, photographing deterioration and damage, and registering the

photographs on a 3D model with simple hand gestures. In addition, an evaluation of the accuracy of 3D scanning confirmed that scanning with sufficient accuracy is possible for simple inspection.

Future works include a quantitative comparison of the accuracy of position specification by hand gestures and linking with remote locations using communication functions. We also consider using it to supplement areas where 3D measurement by UAV is difficult. In addition, the developed tool can be used widely, not only for infrastructure inspection, but also for maintenance and management of buildings and park trees, and on-site investigations immediately after the disasters. It is also thought that multiple workers can use the tool in parallel and share the obtained information such as 3D models and images on a server in real time via high-speed communication. It is also possible for an expert to analyze this aggregated information and give appropriate instructions to each worker on site based on the analysis results. Such two-way communication between the office and the field via a 3D model is expected to improve the sophistication of inspections and investigations.

ACKNOWLEDGMENTS

We thank Kumonos Corporation for providing the terrestrial laser scanning point cloud used in this study.

REFERENCES

- Blaskow, R. and Maas, H. G. (2024). Structural health monitoring of bridge with personal laser scanning: segment-based analysis of systematic point cloud deformations, *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, X-2-2024, 9-16.
- Khoshelham, K., Tran, H. and Acharya, D. (2019). Indoor Mapping Eyewear: Geometric Evaluation of Spatial Mapping Capability of Hololens, *ISPRS Archives of the Photogrammetry, Remote Sensing and Spatial Information Science*, Volume- XLII-2/W13, 805-810.
- Mizoguchi, T., Mizuno, K. and Taniguchi, O. (2023). Automatic Creation of 3D Textured Simplified Model for Supporting Piled Pier Maintenance, *Proceedings of The Fifth International Conference on Civil and Building Engineering Informatics*, 289-295.
- Mohammadi, M., Rashidi, M., Mousavi, V., Karami, A., Yu, Y. and Samali, B. (2021). Quality evaluation of digital twins generated based on UAV photogrammetry and TLS: bridge case study, *Remote Sensing*, 13, 3499.
- Rashidi, M., Mohammadi, M., Sadeghlou Kivi, S., Abdolvand, M. M., Truong, H. L., and Samali, B. (2020). A Decade of Modern Bridge Monitoring Using Terrestrial Laser Scanning: Review and Future Directions, *Remote Sensing*, 12(22), 3796.
- Sacks, R., Kedar, A., Borrmann, A., Ma, L., Brilakis, I., Hüthwohl, P., Daum, S., Kattel, U., Yosef, R., Liebich, T., Barutcu, B. E. and Muhic, S. (2018). SeeBridges as next generation bridge inspection: overview, information delivery manual and model view definition, *Automation in Construction*, 90, 134-145.
- Shim, C. S., Dang, N. S., Lon, S., and Jeon, C. H. (2019). Development of a bridge maintenance system for prestressed concrete bridges using 3D digital twin model, *Structure and Infrastructure Engineering*, 15(10), 1319–1332.
- Tang, Z., Peng, Y., Li, J. and Li, Z. (2024). UAV 3D modeling and application based on railroad bridge inspection, *Buildings*, 14, 26.
- Ungureanu, D., Bogo, F., Galliani, S., Sama, P., Duan, X., Meekhof, C., Stühmer, J., Cashman, T. J., Tekin, B., Schönberger, J. L., Olszta, P. and Pollefeys, M. (2020). HoloLens 2 Research Mode as a Tool for Computer Vision Research, *arXiv:2008.11239*.
- Weinmann, M., Jäger, M. A., Wursthorn, S., Jutzi, B., Weinmann, M. and Hübner, P. (2020). 3D Indoor Mapping with the Microsoft Hololens: Qualitative and Quantitative Evaluation by Means of Geometric Features, *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Science*, Volume-1-2020, 165–172.
- Yan, Y. and Hajar, J. F. (2021). Automated extraction of structural elements in steel girder bridges from laser point clouds, *Automation in Construction*, 125, 103582.

A PRACTICE IN CHINA FOR COMPUTER AIDED DESIGN AND CONSTRUCTION OF PREFABRICATED ENCLOSURE WALLS

Shen Wei¹, Liu Tao², Wang Guangjian³, Yu Hu⁴, Gao Hang⁵

- 1) Structural Design Software Engineer, Beijing Yjk Building Software Co., Ltd. Beijing. Email: shenwei@yjk.cn
- 2) Structural Design Software Engineer, Beijing Yjk Building Software Co., Ltd. Beijing. Email: liutao@yjk.cn
- 3) Structural Design Software Engineer, Beijing Yjk Building Software Co., Ltd. Beijing. Email: Wangguangjian@yjk.cn
- 4) Structural Design Software Engineer, Beijing Yjk Building Software Co., Ltd. Beijing. Email: Yuhu@yjk.cn
- 5) Director of Structural Engineering, Beijing Yjk Building Software Co., Ltd. Beijing. Email: Gaohang@yjk.cn

Abstract: This article combines the author's past software development and other work practices related to computer-aided design and construction of building enclosure walls, discussing how computer technology can assist in various methods and directions during the design and construction of building enclosure walls. It mainly involves structural design of load-bearing enclosure walls, structural design of non-load-bearing enclosure walls, thermal insulation design of building enclosure walls, waterproof design of building enclosure walls, planning for enclosure wall hoisting, calculation verification for enclosure wall hoisting, and construction of enclosure walls. Whether in the design of the main structure of the building or in the planning and decoration of the building space, designers need to put in a lot of effort for enclosure walls. It is worth studying how to fully utilize the capabilities of computer technology to free up more time for designers. This article has certain reference value for research and application in related fields.

Keywords: Enclosure Walls; CAD; CAE; BIM; Architectural design; Structural engineering; Interior design;

1. INTRODUCTION

The building enclosure walls is an important component of the building system (not necessarily part of the main structural system), playing a pivotal role in building structure and architectural space design. Generally speaking, building enclosure walls can be constructed using brick walls, concrete walls, stone walls, timber walls, glass curtain walls, metal walls, etc. as well as inorganic materials (such as glass fiber-reinforced inorganic materials). Based on their load-bearing characteristics, the enclosure walls of a building can be categorized into load-bearing walls and non-load-bearing walls.

However, there are pain points and problems to be solved for traditional enclosure wall design and construction process. The poor fire resistance, insufficient waterproof and anti-seepage, moisture-proof performance, lousy insulation effect, and low strength are flaws. And the construction process is complex and time-consuming, which increases construction costs and time. The development of prefabricated buildings and building industrialization can solve many problems. Meanwhile prefabricated engineering design and construction process bring new issues that need attention.

For example, prefabricated components require strict dimensional planning, and prefabricated components with incorrect dimensions will become waste; The demand for multi-disciplinary collaborative design is more evident, and precise hole reservation is required; There are added requirements for short-term working condition verification of individual components, such as lifting, delamination, etc.; The number of drawings has greatly increased, and non-structural professional information is also required to be supplemented in the drawings; The problem of data silos in design is becoming increasingly prominent, and the demand for integrated design, production, and construction data is becoming stronger. So new engineering design and construction process requires new computer-aided technology. There is no much systematic and targeted research to this series of multi-disciplinary issues targeted towards prefabricated walls.

The article discusses the trying of one computer software scheme that integrates prefabricated component design and traditional structural engineering design into one big process, called "YJK's built-in prefabricated building design software", which may adapt to the need to support the prefabricated walls design and construction processes of selection, planning, and installation. Involving several targeted aspects, such as structural design, insulation design, waterproof design, and installation engineering of prefabricated enclosure walls.

2. CAD FOR PREFABRICATED LOAD-BEARING ENCLOSURE WALLS

2.1 Technical scheme

Believe in the law that if the solution process is clear, rule-based expert systems or similar systems can be used, else if the solution process is unclear, machine learning is needed to obtain rules.

Because the design, calculation and detailing process can be planned as deterministic programs and user interactive programs in computers, so the final scheme was as two aspects, one as a rule-based systems, the other as a interaction system. All data was embedded within the structural analysis software.

In short, it was based on structural design models, rule-based, and integrated into the main process of structural analysis.

2.2 Wall type

Currently, in the practice of prefabricated building structures in China, the common-used prefabricated walls are based on the "Prefabricated Concrete Shear Wall Exterior Panels" (15G365-1) and "Prefabricated Concrete Shear Wall Interior Panels" (15G365-2).

In addition, there are numerous prefabricated wall systems in China, each with its own advantages. In the YJK's built-in prefabricated building design software, besides supporting the above two reference manuals, it also supports various forms such as single-row sleeve prefabricated walls, double-row full-sleeve prefabricated walls, double-sided composite shear walls, and vertically unconnected prefabricated walls, etc.

2.3 Design Process

In the YJK's built-in prefabricated building design software, the general design process of prefabricated walls includes: establishing structural models, designing prefabricated wall subdivision schemes, performing overall structural calculations, editing planar reinforcement, carrying out detailed design of prefabricated walls, making detailed drawings of prefabricated walls, and exporting the inventory list of prefabricated walls.

According to the current design logic of YJK's built-in prefabricated building design software, the prefabricated wall subdivision scheme is a specialized task based on the existing building structure model. Of course, with the rapid development and maturity of the national building industrialization level, the software may gradually evolve into another systematic workflow following the progress of the industry. Figure 1 shows the subdivision scheme for prefabricated wall in the software (filling parts are prefabricated walls).

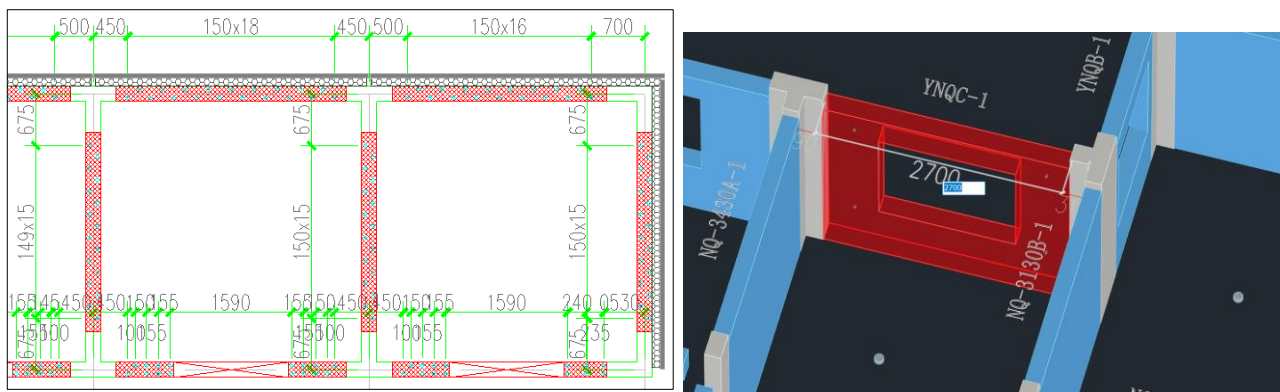


Figure 1: Subdivision scheme for prefabricated walls of 15G365-1 and 15G365-2 in YJK's built-in prefabricated building design software

Based on current practice and understanding, in the proposed amendments to the "Technical Standard for Sawtooth-Shaped Prefabricated Concrete Shear Wall Structures" (T/HNKCSJ011-2024), it is suggested that the original provision 3.0.1, which reads "Design of prefabricated subdivision scheme, including structural selection, structural layout, force transfer path, and production and assembly of prefabricated units of structural components," be revised to read: "Overall scheme design, including structural selection, structural layout, design of prefabricated subdivision scheme, force transfer path, and production and assembly of prefabricated units of structural components." This is because, based on the reading and understanding of this standard, provision 3.0.1 mentions a total of five structural design elements of the building's structural system. It is believed that the first item does not solely encompass the "design of prefabricated subdivision scheme".

2.4 Calculation of horizontal joints

In the integrated prefabricated shear wall structure system, node connections performance is particularly important. The Technical Specification for Prefabricated Shear Wall Structures stipulates that the shear bearing capacity of the horizontal joints at the bottom of prefabricated shear walls should be checked.

In the YJK's built-in prefabricated building design software, there are corresponding functions (Figure 2).

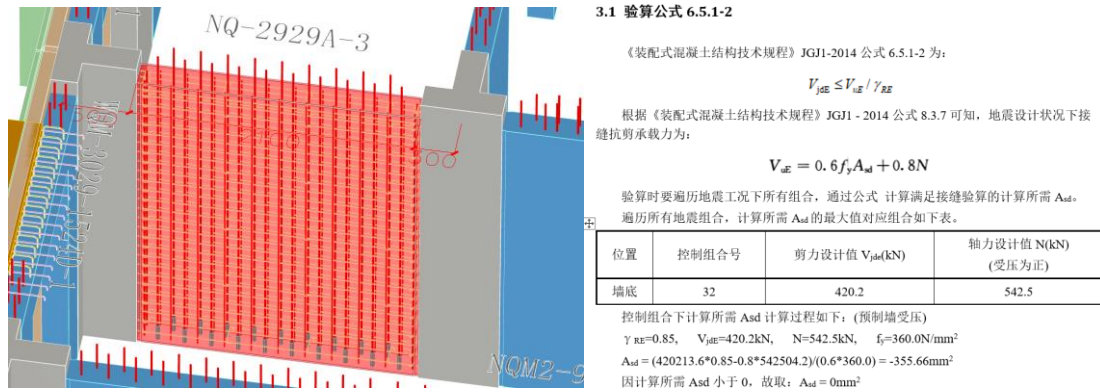


Figure 2: Some result content of the prefabricated wall horizontal joints calculation in YJK's built-in prefabricated building design software

2.5 Component reinforcement and detailed drawings

In the YJK's built-in prefabricated building design software, the principle of "normal structure drawing first" is followed. The design process involves the construction drawing by plane integral representation method and the detailed reinforcement drawing for prefabricated walls. The process of obtaining reinforcement for prefabricated walls is shown in the figure.

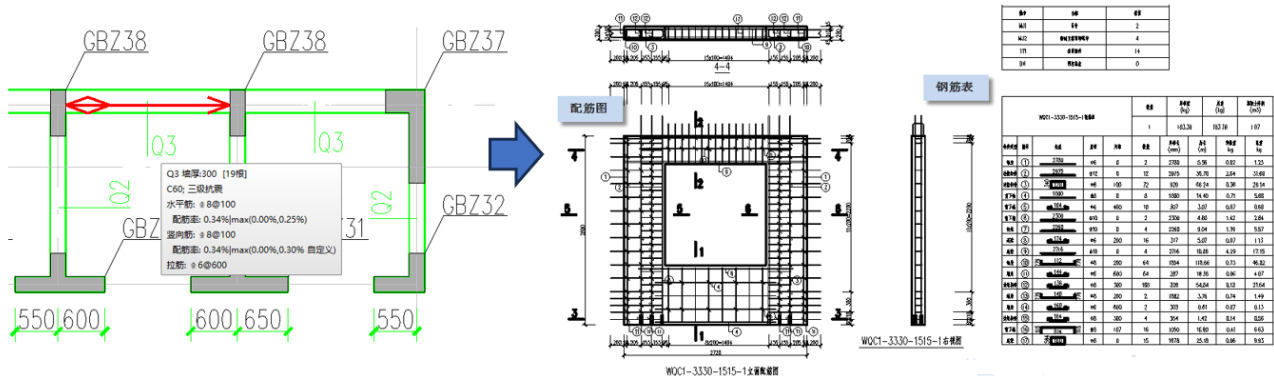


Figure 3: Prefabricated wall reinforcement obtaining process in YJK's built-in prefabricated building design software

After obtaining detailed reinforcement information, you can proceed to review the reinforcement diagram for prefabricated walls.

3. CAD FOR PREFABRICATED NON-LOAD-BEARING ENCLOSURE WALLS

3.1 Technical scheme

Generally speaking, non-load-bearing walls are not parts of structural analysis model. Because current structural analysis only focuses on the load and rough stiffness contributions it causes, without the strong need for precise non-load-bearing wall model data. But some experts do suggest accurately stiffness considering.

Actually in YJK structural software, non-load-bearing walls can be modeled. Simultaneously, prefabricated non-load-bearing walls can be modeled based on the ordinary non-load-bearing wall data.

In short, prefabricated non-load-bearing walls was based on ordinary non-load-bearing wall data in the structural design software, and is rule-based, and integrated into the main process of structural analysis.

3.2 Wall type of ordinary non-load-bearing wall

Non-load-bearing enclosure walls can be categorized based on their stress characteristics as self-supporting walls and partition walls.

And partition walls are classified according to their construction methods as follows:

(1) Block partition wall:

Block partition walls are constructed by block units made of some kind of material such as ordinary bricks, hollow bricks, aerated concrete, etc. The degree of prefabrication is relatively low, and nowadays they are less

commonly used in prefabricated buildings.

(2) Lightweight frame partition wall

The lightweight frame partition wall consists of two parts: the frame and the surface panel layer. The frame can be made of wood or metal, and the surface panel layer can be made of wood, plywood, fiberboard, gypsum board, etc.

(3) Plate partition wall

Plate partition wall refers to the direct installation of lightweight sheets with large dimensions in length and width. The lightweight sheets can be made of steam-added lightweight concrete slab partition walls (autoclaved lightweight concrete strip panels, ALC), composite slab walls, etc. It is easy to assemble and is widely used in prefabricated buildings nowadays.

In special, for frame structure buildings, non-load-bearing enclosure walls can be categorized into infilled walls and curtain walls. And the infilled walls (including internal partition walls and external retaining walls) can be prefabricated.

3.3 Prefabricated Concrete non-load-bearing enclosure wall

The prefabricated concrete non-load-bearing enclosure wall can be either a single piece of wall or an assembled wall made of slats.

For the kind of single piece, the YJK's built-in prefabricated building design software can support various types of prefabricated concrete partition walls such as prefabricated infilled walls, prefabricated beams with partition walls, and prefabricated external wall panels.

For the kind of assembled with slats, software can support prefabricated partition walls in the form of strip panels, primarily made of ALC (autoclaved lightweight concrete strip panels).

3.4 Design process

The design process of prefabricated non-load-bearing enclosure walls in YJK's built-in prefabricated building design software are: establishing a structural model with partition walls, specifying prefabricated partition walls, checking and calculating prefabricated partition walls, editing the detail of prefabricated partition walls, making detailed drawings of prefabricated partition walls, and exporting the inventory list of prefabricated partition walls.

The screenshot shown in the figure depicts a portion of the designated and editable features for prefabricated partition walls in YJK's built-in prefabricated building design software.

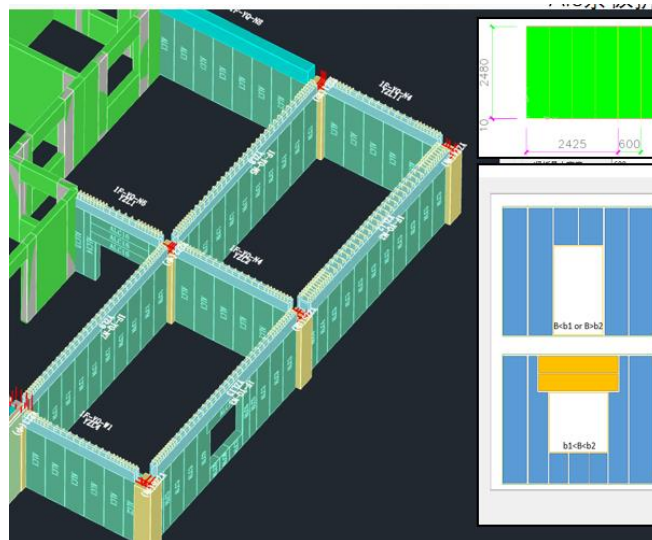


Figure 4: The designation and subdivision functions for prefabricated partition walls (ALC, assembled with slats) in YJK's built-in prefabricated building design software

3.5 Computer-aided calculation related to structural safety

The bearing capacity of the external enclosure wall should be calculated under the effects of wind load and seismic load.

In the YJK's built-in prefabricated building design software, the action effect value of the prefabricated external enclosure wall refers to 10.2 of the Technical Specification for Prefabricated Concrete Structures (JGJ1-2014). The calculation is carried out separately for the two cases of “upper end fixed and lower end hinged”, as well as “upper and lower end hinged”, and the envelope is taken. The bending bearing capacity of the wall body is checked with reference to 6.2.10-1 of the Code for Design of Concrete Structures (GB 50010-2010) and the

Technical Specification for Prefabricated Concrete Structures (JGJ1-2014). The shear bearing capacity of the wall body is checked with reference to 6.3.1-1, 6.3.3-1, 6.3.3-2 of the Code for Design of Concrete Structures (GB 50010-2010) and the Technical Specification for Prefabricated Concrete Structures (JGJ1-2014). The node bearing capacity is checked separately for in-plane shear and out-of-plane tensile resistance. The figure shows part of the results.

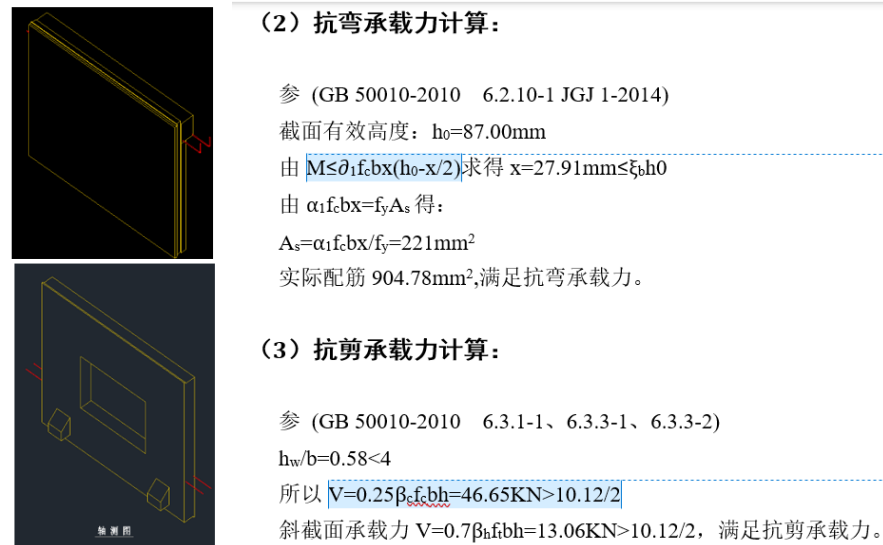


Figure 5: Some part of calculation for prefabricated external non-load-bearing enclosure wall in YJK's built-in prefabricated building design software

4. THE ENCLOSURE WALL IN THE AUTOMATIC ASSEMBLY RATE CALCULATION

4.1 Technical scheme

Assembly rate is currently a reference indicator in China for evaluating the industrialization level of prefabricated buildings. This indicator roughly represents the proportional between prefabricated and non-prefabricated parts. In fact, the calculation methods vary greatly, and there are different calculation regulations in different regions of China.

The calculation process is not completely clear even according to specific regulation. But it is not very suitable for using machine learning methods because the answer to question is essentially a contractual relationship and the rules are still in development and discussion. So the software system was planned as rule-based expert system, permitting different regions switching to different rules.

4.2 Software implementation

The screenshot shown in the figure depicts the UI of load-bearing and non-load-bearing enclosure walls when automatically calculating the assembly rate in YJK's built-in prefabricated building design software, based on the national standard "Evaluation Standard for Prefabricated Buildings" (BG/T51129-2017). Additionally, it supports over twenty other local evaluation standards.

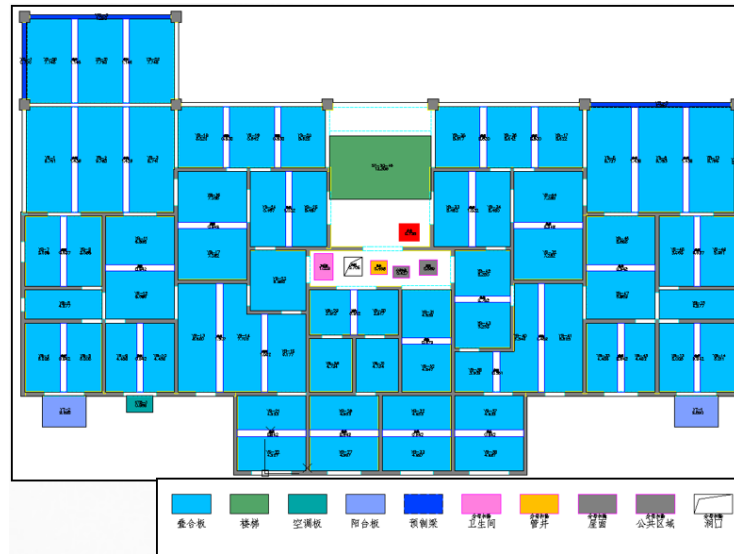


Figure 6: The assembly rate calculation view in YJK's built-in prefabricated building design software

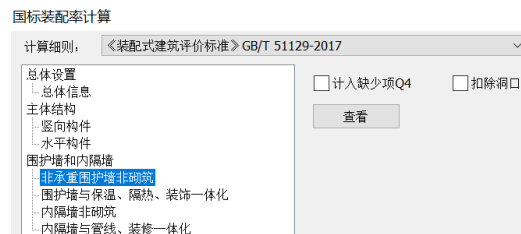


Figure 7: Part of the UI related to non-load-bearing enclosure walls in the calculation of assembly rate using YJK's built-in prefabricated building design software (GB part)

5. THERMAL INSULATION DESIGN AND CAD TECHNOLOGY FOR PREFABRICATED EXTERIOR ENCLOSURE WALLS

5.1 Technical scheme

Heat preservation in winter and thermal insulation in summer are crucial considerations that cannot be overlooked in exterior enclosure wall architectural design.

This issue is solely about the prefabricated enclosure wall itself, and have nothing to do with the underlying structural design or structural models.

Current technical solution categorizes this issue as a deepening design of components. For example, the placement of insulation layers for walls of different types and shapes, the definition method of local insulation components, etc.

5.2 Heat preservation and thermal insulation design for different kind of wall

In YJK's built-in prefabricated building design software, two kinds of insulation layer practices are supported: sandwich prefabricated exterior walls with insulation layer (GB standard load-bearing prefabricated walls, Figure 8) and double-sided laminated walls (including load-bearing walls and non-load-bearing walls) with internal insulation layers.

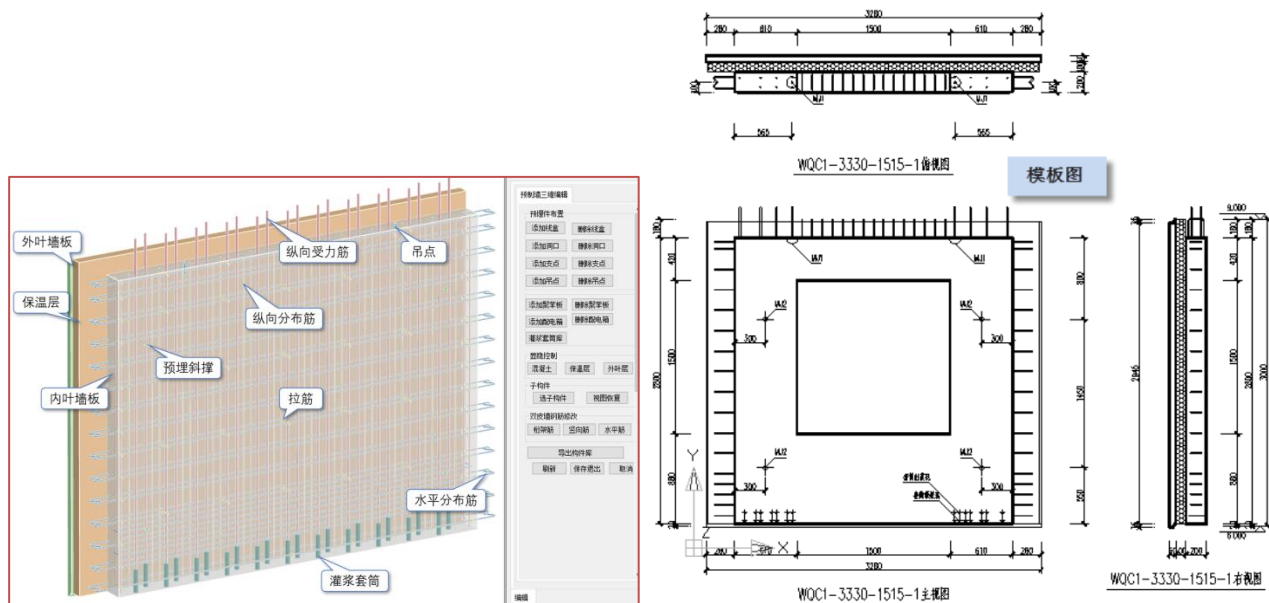


Figure 8: a sandwich prefabricated exterior wall(with a insulation layer which is larger than wall itself)
In YJK's built-in prefabricated building design software (GB part)

5.3 Insulation material

The insulation layer material for the exterior enclosure wall can be spray-on insulation materials such as polyurethane insulation materials; insulation boards made from inorganic fiber materials, such as glass wool and rock wool; and insulation boards made from polystyrene resin, including expanded polystyrene (EPS) boards and extruded polystyrene (XPS) boards.

YJK's built-in prefabricated building design software supports embedding polystyrene boards for prefabricated walls to reduce self-weight (weight-reducing boards), while also providing insulation functionality.

5.4 Fire protection for insulation layer

The "Technical Specification for Prefabricated Concrete Structures" (JGJ1-2014)4.3.2 stipulates that the combustion performance of insulation materials in sandwich exterior wall panels should meet the requirements of Class B₂ in the "Classification for Burning Behavior of Building Materials and Products" (GB8624-2012). According to Chapter 4 of the "Classification for Burning Behavior of Building Materials and Products" (GB8624-2012), Class B₃ refers to flammable materials.

To comply with the Technical Specification for Prefabricated Concrete Structures, according to the provisions of JGJ1-2014, the amendment suggestion for the "Technical Standard for Sawtooth-Shaped Prefabricated Concrete Shear Wall Structures" (T/HNKCSJ011-2024) is to adjust the original provision "The insulation layer should be made of materials with a combustion performance not lower than Class B. The thickness should be determined by calculation" to "The insulation layer should be made of materials with a combustion performance not lower than Class B₂. The thickness should be determined by calculation".

The fire protection standards for the insulation layer of the external enclosure wall should be improved in accordance with regulations, in order to prevent the spread of fire from one part of the building to another along the external wall as much as possible.

Although YJK has the function of fire resistance calculation for structural components, the fire resistance for insulation layer is suggested to be confirmed through the definition method in the previous two sections.

6. WATERPROOF DESIGN AND CAD TECHNOLOGY FOR PREFABRICATED EXTERIOR ENCLOSURE WALLS

6.1 Technical scheme

Waterproof performance is another important design considerations for the exterior enclosure walls.

This issue is solely about the prefabricated enclosure wall itself, and have nothing to do with the underlying structural design or structural models.

Current technical solution also categorizes this issue as a deepening design of components. For example, waterproof details of vertical connection nodes for prefabricated wall outer leaf panels, and treatment for horizontal connection nodes between prefabricated walls.

6.2 Waterproof material

In article 4.3.1-2 of the Technical Specification for Prefabricated Concrete Structures, it is stipulated that

silicone, polyurethane, and polysulfide building sealants should comply with the provisions of Silicone Building Sealants (GB/T14683), Polyurethane Building Sealants (JC/T482), and Polysulfide Building Sealants (JC/T483), respectively.

The Technical Specification for Prefabricated Concrete Structures (JGJ1-2014) does not specify any restrictions on the use of sealants, so proposing amendments to "Technical Standard for Sawtooth-Shaped Prefabricated Concrete Shear Wall Structures" (T/HNKCSJ011-2024) that it is recommended to lift the restrictions on the use of sealants. Specifically, the original provision "The joints of the external protective layer should be sealed with structural sealants. The durability and performance indicators of the structural sealants should meet the relevant provisions of the current national standard 'Silicone Structural Sealants for Building' GB16776" is suggested to be adjusted to "The joints of the external protective layer should be sealed with structural sealants. The durability and performance indicators of the structural sealants should comply with the relevant provisions of the current national standard."

If there are no other considerations, designers should be able to independently weigh and select the appropriate sealant based on project needs.

6.3 Waterproof method

The YJK's built-in prefabricated building design software has been enhanced with the ability to automatically set tongue-and-groove joints, following the criterion method for the national standard of prefabricated shear walls. As shown in the figure.

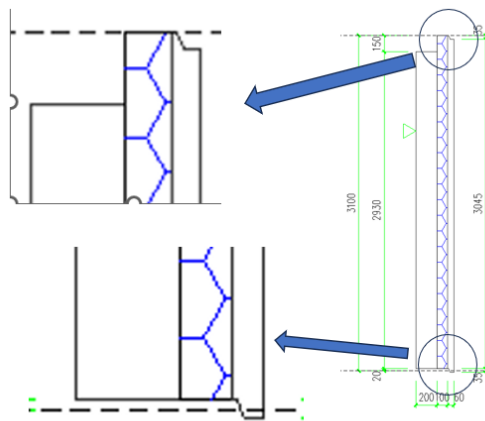


Figure 9: YJK's built-in prefabricated building design software automatically sets up tongue-and-groove joints for prefabricated exterior walls

People also can use YJK's built-in prefabricated building design software to design horizontal nodes of the external enclosure walls mutually, as shown in the figure.

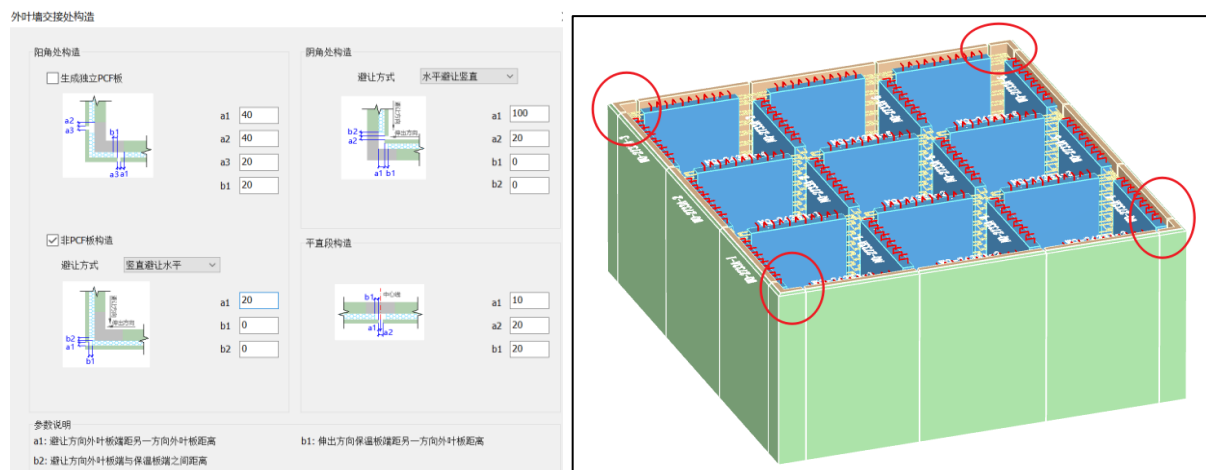


Figure 10: Wall horizontal connection node design in YJK's built-in prefabricated building design software

7. COMPUTER TECHNOLOGY FOR THE CONSTRUCTION OF PREFABRICATED ENCLOSURE WALLS

7.1 Technical scheme

Regarding construction, if we focus on the issue of prefabricated components, it can be assumed that this issue is solely about prefabricated enclosure wall itself, and have nothing to do with the underlying structural design or structural models.

However, there are various issues with different aspects to be solved. Firstly, 3D Model based visualization is required for installation planning function, and the core technology is about Computer Graphics. Secondly, independent lifting safety calculation is needed in the lifting validation, and Material Mechanic is the core related knowledge. Thirdly, to achieve automatic installation, it is required that the design data can match the on-site environment, so the core is about data transmission and sharing, coordinate alignment between computer data and the real world.

7.2 Installation planning

To achieve the goal of facilitating users in planning and simulating the construction installing of wall panels, efforts in two aspects were made.

The hoisting scheme function in YJK's built-in prefabricated building design software is implemented by the way of dynamic display of the model.

On the other hand, reinforcement collision is a troublesome thing. Steel collision detection program provides prompts. In some cases, also providing a certain degree of automatic avoidance solutions (Figure 11).

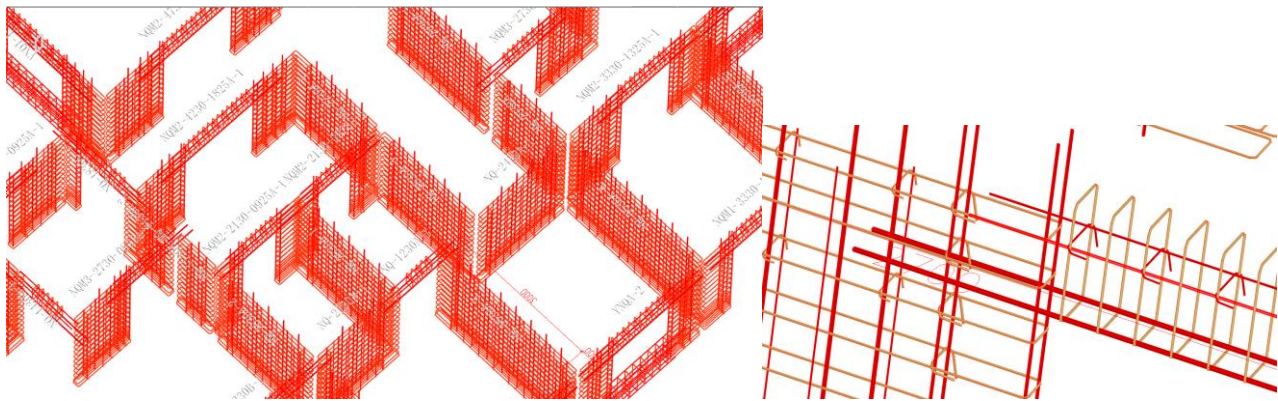
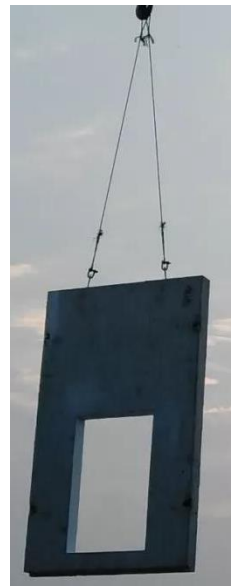


Figure 11: Reinforcement collision detect program in YJK's built-in prefabricated building design software

7.3 Lifting verification

Some part of the calculation content for the hoisting calculation function in the YJK's built-in prefabricated building design software is shown in the figure below. It can automatically perform hoisting calculation verification for wall panels and output a complete hoisting calculation sheet.



3.1 混凝土开裂验算

根据《混凝土结构工程施工规范》GB50666-2011 第 9.2.3 条：

吊点	跨长 L1(mm)	跨长 L2(mm)	跨长 L3(mm)
吊点排布	550	1650	640

取 1 米板带作为计算单元 $L_0=1\text{m}$

计算线荷载 $q=Q_k \cdot L_0/S=57.51 \times 1/7.7=7.50\text{kN/m}$

根据吊点位置，最大跨中弯矩值 $M_{\text{MAX} \pi}=1.22\text{kN} \cdot \text{m}$

最大支座弯矩值 $M_{\text{MAX} \pi}=-1.54\text{kN} \cdot \text{m}$

控制弯矩值 $M_{\text{MAX}}=\max(|M_{\text{MAX} \pi}|, |M_{\text{MAX} \pi}|)=1.54\text{kN} \cdot \text{m}$

截面抵抗矩 $W=B \cdot H^2/6=1 \times 0.2^2/6=0.0067\text{m}^3$

正截面边缘混凝土法向拉应力计算： $\sigma_{\text{sk}}=M_{\text{MAX}}/W=1.54/(0.0067 \times 1000)=0.23\text{N/mm}^2$

脱模时按构件混凝土强度达到抗拉强度标准值的 75.0%，经验算

$\sigma_{\text{sk}} < 75.0\% \cdot f_{\text{tk}}=1.50\text{N/mm}^2$ ，满足要求！

3.2 脱模件承载力验算：

单个吊件脱模吊装荷载计算：

$z=1/\cos \beta=1.00$

$F_D=G \cdot z/n=38.34 \times 1.00/4=9.58\text{kN}$

$N_{\text{ax}}/\gamma_1=36.00/4.00=9.00\text{kN} < 9.58\text{kN}$ 不满足要求！

3.3 脱模混凝土锥体破坏强度验算：

单根锚栓受拉，理想混凝土锥体破坏时的受拉承载力标准值：

Figure 12: Some part of the calculation content for wall hoisting in YJK's built-in prefabricated building design software

7.4 For Wall Installing Robot

The key technologies for wall installing robots are perception (sensors) coordinate positioning, navigation, decision-making, etc.

The integration of wall installation robots with BIM enclosure wall data can be achieved by the alignment of virtual and real spatial coordinates. This enables accurately placing the enclosure wall to the correct position and angle at the construction site.

The robot should do specific tasks such as carrying, setting out, etc. The ALC wall data of YJK's built-in prefabricated building design software can be imported into REVIT for subsequent robotic operations (Figure 13), while other wall data can be imported into AUTOCAD for subsequent robotic operations.

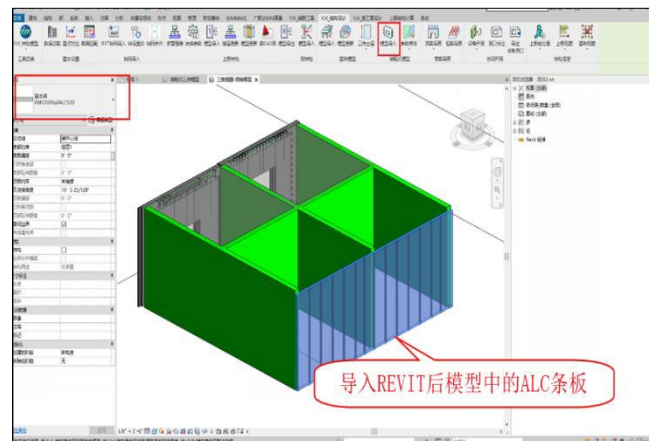


Figure 13: Output ALC data into revit

7.5 Roughly estimating carbon emissions

The entire process of carbon emission calculation can be divided into three stages: the production stage of prefabricated components, the transportation stage of prefabricated components, and the construction stage.

Firstly, the transportation phase mainly considers transportation distance, transportation machinery, and fuel or electricity consumption of transportation machinery. However, prefabricated structures need to consider the issue of empty load during transportation.

Secondly, the emission reduction effect is more significant during the construction stage of prefabricated components compared to the construction stage of cast-in-place structures. Mainly considering quota standards, consumption of labor, water, electricity, machinery, etc.

Thirdly, speaking of carbon emission calculation software during the production phase, the accuracy depends on the precision of the steel reinforcement, and further improvement is needed.

8 SUMMARY

During the entire process of building enclosure walls, from design to installation and construction, computer-aided technology can be leveraged to enhance productivity in many areas.

This article discusses several computer technology solutions. It primarily covers the structural design of load-bearing and non-load-bearing enclosure walls, insulation design for exterior enclosure walls, waterproofing design for exterior enclosure walls, as well as hoisting and construction. It should be noted that not all implemented content has been listed here.

There are both successful experiences and lessons that need to be learned. One of the successful experiences is the benefits brought by close integration. Lessons learned include that detailed design should be based on a more powerful 3d platform, which is currently on the way to implementation. In addition, the refinement of functions is also underway, for example, targeted functional improvements and new multi-disciplinary collaboration solutions.

The next generation of implementation plans is also under consideration, one of which is more scalability and definability of rules, and the other is a feasible solution to reduce user interaction and interference.

REFERENCES

- buildingSMART. (2024). *IFC4.3.2.0*. Retrieved from buildingSMART website: https://standards.buildingsmart.org/IFC/RELEASE/IFC4_3/
- Fujian Provincial Department of Housing and Urban-Rural Development.(2021). *Technical Standard for Prefabricated Building Envelope Wall Structure*: DBJ/T13-343-2020. [S].
- Guoshun Green Building Technology Co., Ltd. (2023). *Design method and system for exterior maintenance wall panels based on BIM* : CN202311117170.3. [P].
- Jang, S. and Lee, G. (2024). BIM Library Transplant: Bridging Human Expertise and Artificial Intelligence for Customized Design Detailing, *Journal of Computing in Civil Engineering*, 38(2), 04024004.
- Wang, Z., Sacks, R., Ouyang, B., Ying, H., and Borrmann, A. (2023). A Framework for Generic Semantic Enrichment of BIM Models, *Journal of Computing in Civil Engineering*, 38(1), 04023038.
- Mengiste, E., Mannem, K.R., Prieto, S.A., and Garcia de Soto, B. (2024). Transfer-Learning and Texture Features for Recognition of the Conditions of Construction Materials with Small Data Sets, *Journal of Computing in Civil Engineering*, 38(1), 04023036.
- Gan, V.J.L. (2022). BIM-Based Building Geometric Modeling and Automatic Generative Design for Sustainable Offsite Construction, *Journal of Construction Engineering and Management*, 148(10), 04022111.
- Gilani, G., Pons, O., and de la Fuente, A. (2021). Sustainability-Oriented Approach to Assist Decision Makers in Building Facade Management, *Journal of Construction Engineering and Management*, 148(1), 4021182.
- Costa, S., Carvalho, M.S., Pimentel, C., and Duarte, C. (2023). A systematic literature review and conceptual framework of construction industrialization, *Journal of Construction Engineering and Management*, 149(2), 03122013.
- Jang, J.Y., Kim, J.I., Koo, C., and Kim, T.W. (2022). Automated components—Vehicle allocation planning for precast concrete projects, *Journal of Management in Engineering*, 38(6), 04022059.

A BIM Component-centred Bridge Digital Twin for Smart and Practical Bridge Maintenance

Guanyu Xiong¹, Yan Gao², Chengzhang Chai³, Kehong Chen⁴ and Haijiang Li^{5*}

1) Ph.D. Candidate, School of Engineering, Cardiff University, United Kingdom. Email: xiongg@cardiff.ac.uk

2) Ph.D., School of Engineering, Cardiff University, United Kingdom. Email: GaoY74@cardiff.ac.uk

3) Ph.D. Candidate, School of Engineering, Cardiff University, United Kingdom. Email: ChaiC1@cardiff.ac.uk

4) Ph.D. Candidate, School of Engineering, Cardiff University, United Kingdom. Email: ChenK43@cardiff.ac.uk

5) Prof., School of Engineering, Cardiff University, United Kingdom. Email: LiH@cardiff.ac.uk

* Corresponding author

Abstract: For bridge visual inspection and maintenance, a fundamental task is to determine the condition of a bridge component from its appearance as captured by images. Then the information of the identified defects is recorded in documents and assessed collectively by bridge practitioners to determine what maintenance activities are required for the component. This engineering practice naturally fits for the idea of IFC components in BIM workflow. So, to automate this labour-intensive and time-consuming process, a smart and practical framework is proposed utilising BIM component-centred bridge digital twin system. In this system, sensors-equipped robotics is integrated with the bridge digital twin to identify any defect. Component-wise defect data is linked to the BIM model for detailed and holistic assessment, ensuring that maintenance decisions are fully informed. This framework is validated by demonstration of key functions including RTK-enabled drone for automatic defect localisation, defect quantification by computer vision, defect data storage in SQL and visualisation of enriched BIM model in interactive web-based platform for maintenance decision-making. It is demonstrated that the framework can streamline defect data transfer from on-site inspection to an online bridge digital twin, supporting decision-making processes by referencing relevant industrial standards.

Keywords: Bridge Maintenance, Digital Twin, BIM, IFC

1. INTRODUCTION

In recent years, the construction industry is experiencing digital transformation and Building Information Modelling (BIM) is a key tool to facilitate the management of full life cycle of built environments, especially the maintenance of assets. The idea of BIM is consistent and continuous use of digital data and exchange digital data between different systems and platforms (Borrmann *et al.*, 2018). As the major file format of BIM, Industry Foundation Classes (IFC) contain information about an asset and every component of the asset is represented as an object with unique identifiers and attributes, which allows for precise identification, referencing, and retrieval.

According to the UK Network Rail standard NR/L3/CIV/006 (Network Rail, 2022a), the defect properties that the bridge practitioners concern include 1) defect location - which bridge component that the defect is developed upon, and 2) defect description in terms of dimensions, directions, profiles, etc. The characteristics of defect are generally documented as schedules, with each entry corresponding to a specific defect instance. However, adding hundreds of defects observed during an inspection to an IFC model of a bridge could be challenging. As IFC is capable of representing 3D bridge models with highly accurate geometry but it is considered insufficient for updating model content due to its static nature (Isailović *et al.*, 2020). Therefore, the proposed solution to this challenge is to create a data schema that can link bridge components on an IFC model to the defects schedule created from bridge inspection.

Digital twinning is an emerging technology for intelligent asset management that provides an up-to-date representation of an actual physical asset in operation. A digital twin for bridge maintenance is not only about creating a digital representation of the physical bridge but can also update as new data is collected, moreover, it can perform analysis for asset risk assessment. There are studies for different aspects of bridge digital twin, such as drone-enabled bridge inspection (Gao *et al.*, 2023; Yoon *et al.*, 2022), bridge information modelling (Adibfar and Costin, 2022; Mohammadi *et al.*, 2023) knowledge-based and optimisation-oriented bridge management systems (Yang *et al.*, 2022; Allah Bukhsh *et al.*, 2020), etc. Nevertheless, there is limited discussion on the full maintenance cycle of bridge digital twin spanning from data collection to maintenance decision-making. So, this study investigates a framework that can streamline on-site inspection, digital twin data model and synchronisation, and Common Data Environment (CDE) that facilitate structural assessment and decision-making on maintenance activities.

2. METHOD

Bridge maintenance involve several steps, from conducting inspection, identification of defects, recording the defects in terms of locations/type/area, assessing the extent and severity of the defects on the bridge structure, to determining the repair activities need to be carried out. There is information flow in different systems and documents, starting from the physical bridge, reality capture devices (i.e. camera), on-site inspection notes, summarised inspection reports, structural assessment and repair proposals. To streamline the data in this process, by utilising the idea of BIM, a framework shown in the Figure 1 is proposed. The centre of the framework is the BIM Component which comprises of IFC object as well as linked defect data from database. The BIM Component is referenced with ComponentID (i.e. GlobalId in IFC terminology) so that it can be traced across different systems and platforms.

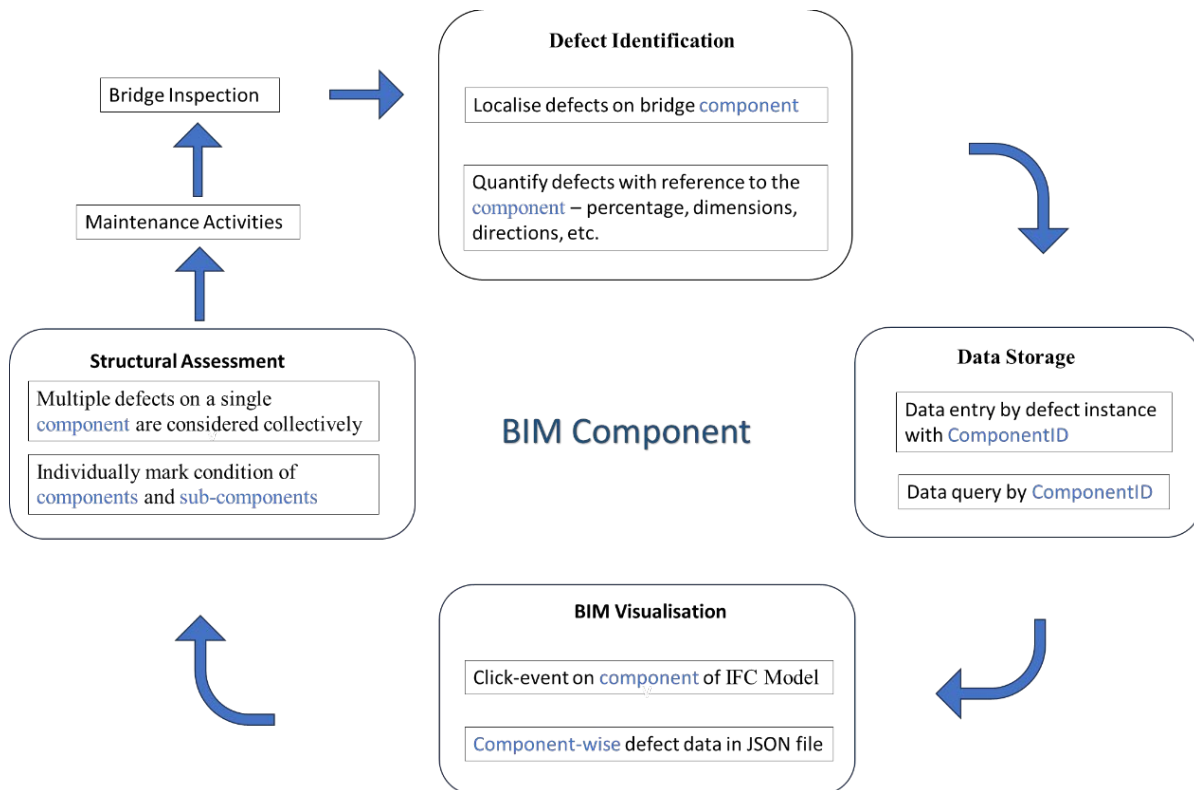


Figure 1: Proposed framework

2.1 Defect Identification on BIM Component

Automated bridge inspection system utilising versatile robotic systems such as drones (Barrile *et al.*, 2019) and climbing robots (Nguyen and La, 2020; Sutter *et al.*, 2018) can not only carry inspection payloads (e.g. camera, LiDAR, etc.) but also provide accurate positioning assisted with control system. For example, mobile robots and drones can locate themselves based on Global Navigation Satellite System (GNSS), obstacle avoidance system (i.e. camera, ultrasonic distance ranger, LiDAR, etc.), and inertial measurement units (IMUs) to help with defect localisation. On the other hand, the detection of defects can be carried out by object detection algorithms such as YOLO, and defect quantification can be realised with semantic segmentation by image processing or deep learning (e.g., DeepLab).

An approach based on Real-Time Kinematic (RTK) is developed to provide defect localisation in the bridge coordinate system, as indicated in Figure 2.

Considering the Earth's surface is nearly planar over short distances of a few kilometers, the distance between the drone and the defect, D_{drone} is calculated as Eq. 1. The coordinates of the defect X_{crack} , Y_{crack} and Z_{crack} can be calculated as Eq.2 - 4.

$$D_{drone} = R\sqrt{(\phi_2 - \phi_1)^2 + (\lambda_2 - \lambda_1)^2} \quad (1)$$

Where L – horizontal distance; R – earth's radius (the parameter that needs to be calibrated); ϕ_1, ϕ_2 – base and drone latitude; λ_1, λ_2 – base and drone longitude.

$$X_{crack} = D_{drone} \sin \theta - D_{crack} \cos \beta \cos \alpha \quad (2)$$

$$Y_{crack} = D_{drone} \cos \theta - D_{crack} \cos \beta \sin \alpha \quad (3)$$

$$Z_{crack} = H_{drone} - H_{base} + h_{equip} - D_{crack} \sin \beta \quad (4)$$

The coordinates can be further linked to bridge components based on geometric information available from IFC model. The crack direction (such as latitudinal or longitudinal) can be determined by drone position and camera angle. Given situations without stable GNSS signals (e.g., underneath a bridge), simultaneous localization and mapping (SLAM) based on IMUs and cameras can be utilised to improve accuracy.

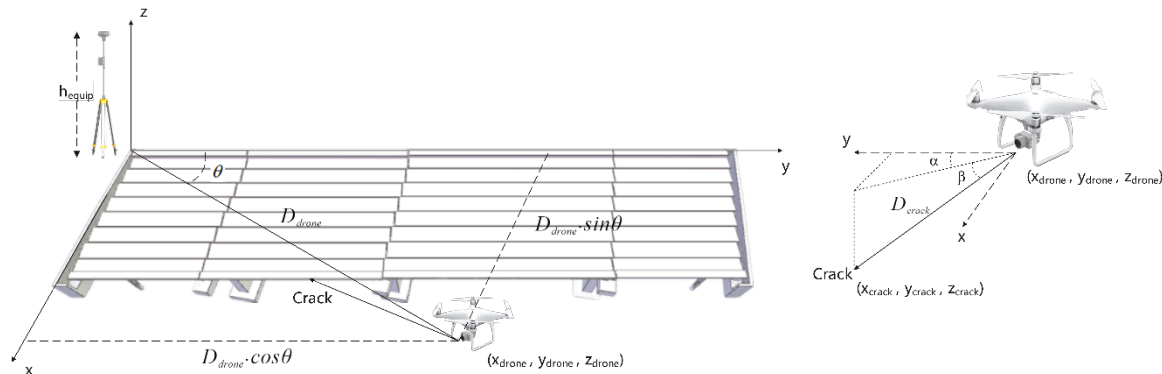


Figure 2: Image capture and defect localisation with RTK-enabled drone (Gao Y, Li H and Xiong G, 2023)

2.2 Defect Data Storage and Visualisation

To enable interoperation between defect data captured from the drone and geometrical data extracted from the IFC model, and to visualize both data streams at the CDE, the workflow shown in Figure 3 has been designed.

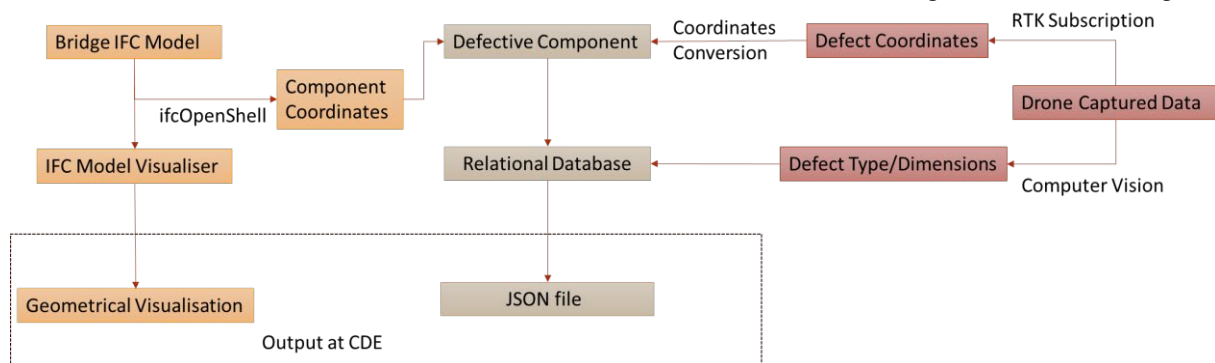


Figure 3: Workflow for defect data storage and visualisation

(1) Data Storage

To achieve interoperability, some functions in the workflow are coded, such as ‘Coordinates Conversion’ module and ‘RTK Subscription’, and some practical tools are utilised – ‘ifcOpenShell’ is required to manipulate geometrical data from the IFC model. Once a defect is captured by RTK-enabled devices in an image, the latitude/longitude/altitude can be extracted from the metadata of the image, then converted to coordinates on 3D model to match the specific IFC component. Finally, the description of each defect instance (e.g. type, dimensions, etc) can be stored in a relational database, ready for retrieving to a JSON file.

(2) Visualisation

The visualisation module shall be able to display not only IFC geometrical model but also defect schedule stored in database. IFC visualiser is employed in the platform for visualisation of 3D model, and also enables users to manipulate each component of the BIM model based on the IFC file. Interactive mechanism shall be allowed at CDE to trigger data query of the defects related to the specific ComponentID, to provide an overview of the

condition of the component, for the ease of review by bridge practitioners.

2.3 Standards-Based Structural Assessment of BIM Component

Multiple defects on a single component shall be considered collectively following NR/L3/CIV/006 (Network Rail, 2022a), where it states that the practitioner shall decide condition rating for each component or sub-component. The component is marked high if the defects identified on a component are localised due to local circumstances while the component is marked low when defects on it are widely spread, as the structure of the component is likely to be severely affected.

Whereas for the whole bridge, the industrial practice of structural assessment in NR/L2/CIV/035 (Network Rail, 2022b) is described as - each bridge is composed of hierarchical levels of components (e.g. components, sub-components, etc.) which are marked individually, then the condition marking of the component is determined against their lowest marked sub-component. Maintenance decisions (i.e. which component shall be repaired) and repair methods are then determined collectively in predefined components group. The decision-making process described above use the idea of decision tree, and it can be implemented as a hierarchical graph structure where some nodes represent a bridge component with its attribute as condition marking, while other nodes represent decision points. The maintenance decision can be determined by traversing the graph structure by following predefined relationships.

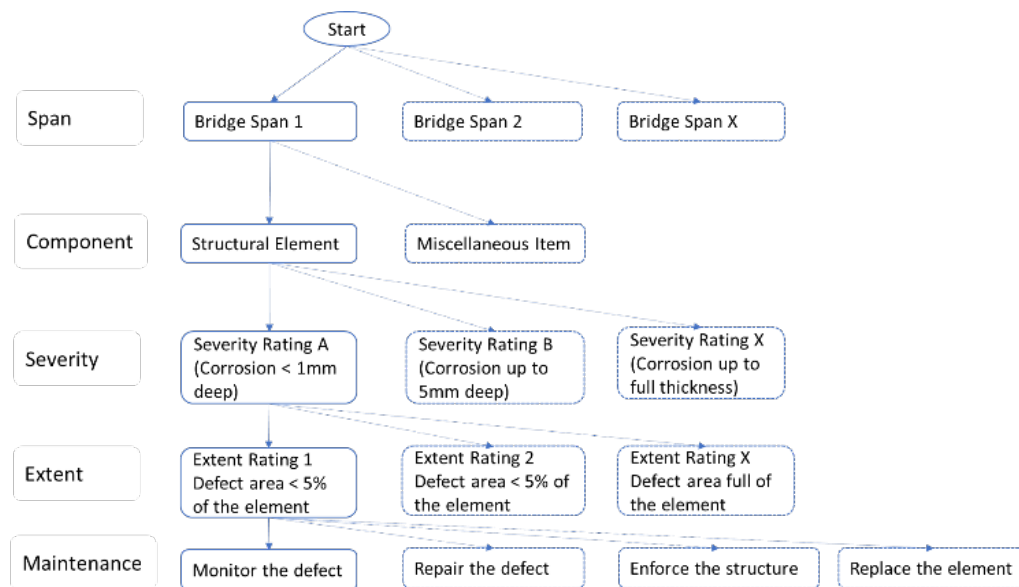


Figure 4: Graph structure for maintenance decision-making

3. RESULTS

The proposed framework was implemented with various tools and software to demonstrate its functionality, i.e., 1) drone-enabled defect localisation and quantification, 2) defect data storage, retrieval and visualisation, and 3) standards-based repair method recommendation and decision-making.

As shown in Figure 5 (a) (b) (c), drone height (i.e. H_{drone} in Equation 1) can be streamed from RTK subscription from OSDK or MSDK of DJI drone, latitude and longitude of drone and base can be obtained from controller and also captured in the metadata of the photo taken. To map drone coordinates from Earth's GPS system to the 3D space of bridge model, the relative position is calculated by determining the difference between the GPS coordinates of the photo and the GPS coordinates of the 3D model's coordinate origin. This establishes the photo's position within the 3D model. The function is implemented as Python programme shown in Figure 5 (d).

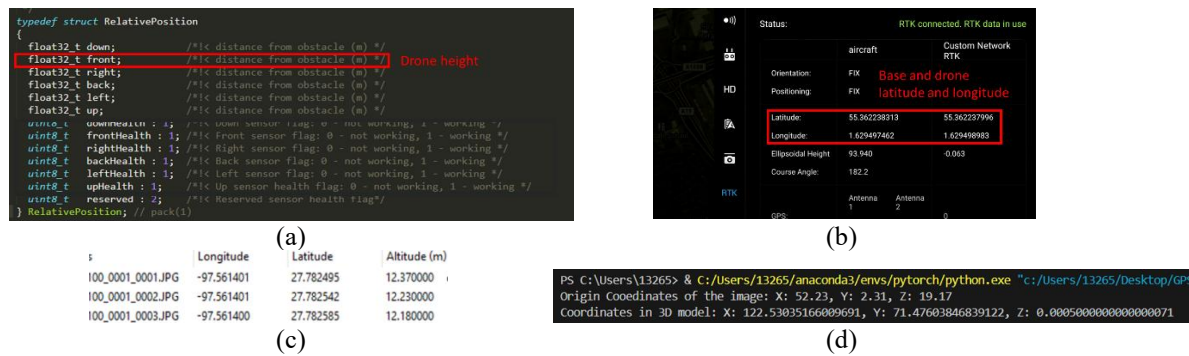


Figure 5: Localisation of defect via RTK subscription(a)(b), metadata of image(c) and coordinates conversion(d)

Defect quantification module follows the principles defined in NR/L2/CIV/035, where corrosion is assessed by percentage of corrosion area, while crack is evaluated by its length, width and direction. The module is deployed in web-services, as illustrated in Figure 6, providing key information after inserting the defect image.

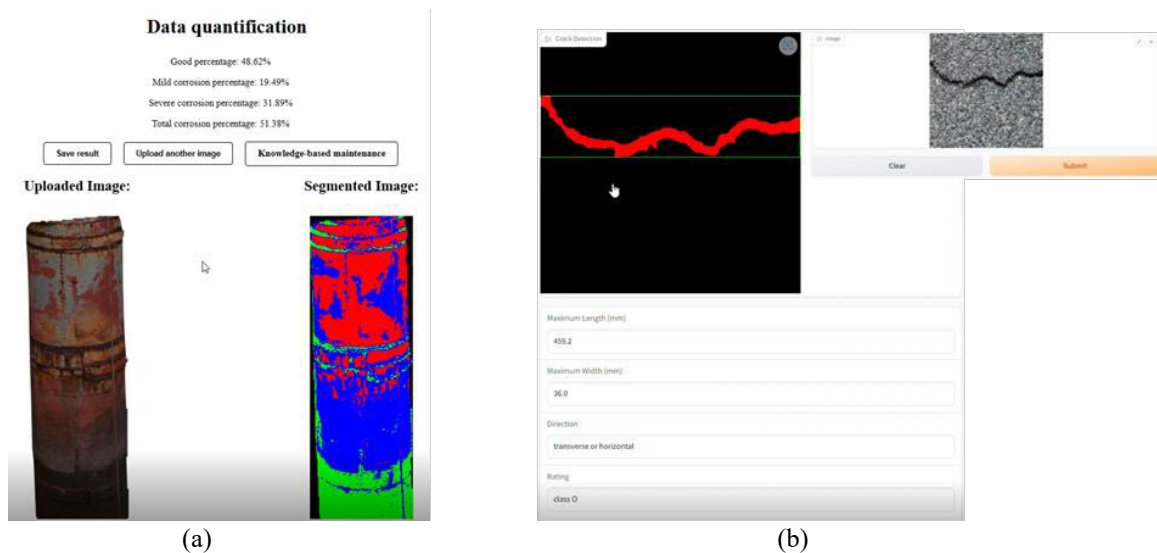


Figure 6: Quantification of corrosion(a) and crack(b) on bridge using computer vision

The bridge digital twin CDE is designed to perform through functions based on the RESTful framework where Node.js and NPM Anywhere server (i.e. a static file server) support the web-based interface. The user-friendly interface can enable retrieval and display of defect information on a specific component by web-based click event, as shown in Figure 7.

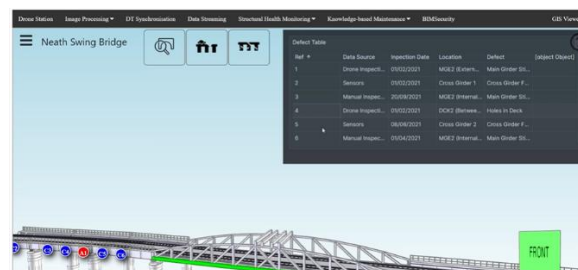


Figure 7: Visualisation of BIM model and defect data

On the other hand, the conceptual design of structural assessment from Figure 4 is implemented in the graph database software Neo4j. As shown in Figure 8, the graph is partitioned into three layers: 1) bridge structure layer which represents graph-based bridge model; 2) defect information layer describing the extent and severity of the defect and 3) maintenance layer which reflects repair proposals. The decision-tree traversal is performed to find the path starting from the bridge structure layer and leading to the node of maintenance layer, with the traversal

being filtered based on the properties of each node along the path.

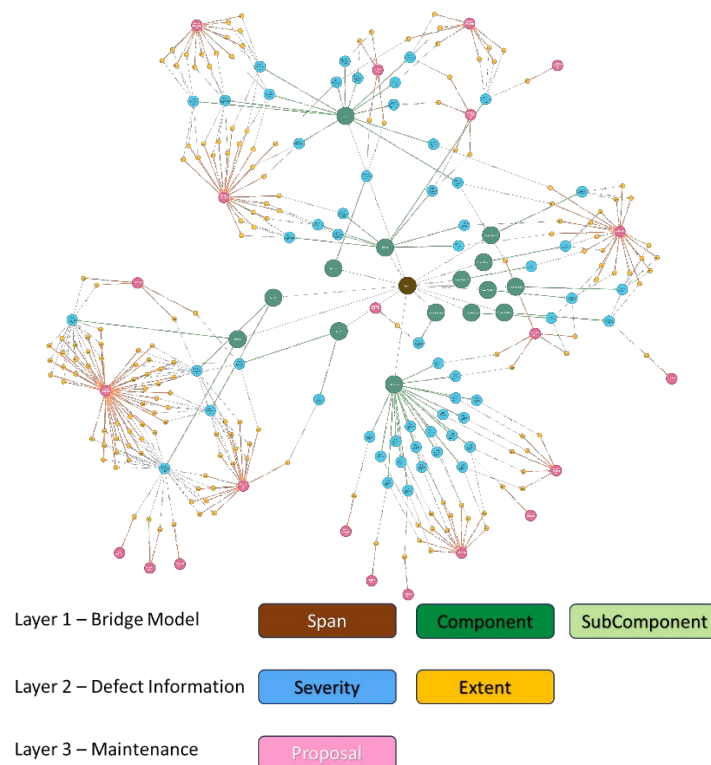


Figure 8: Implementation of graph structure for maintenance decision-making in Neo4j

4. DISCUSSION AND CONCLUSIONS

In this paper, the design and implementation of a BIM component-centred bridge digital twin system purposed for bridge maintenance was discussed and presented. This pipeline for bridge maintenance decision-making is considered smart and practical as it automates and streamlines several labour-intensive and time-consuming process with open-source software toolkits and the structural assessment is based on industrial practices. Specifically, the defects are identified, quantified and localised on the BIM component, the information of the defect is then stored in relational database with the IFC model's ComponentID as its unique identifier. By clicking the visualised IFC geometrical model at web-based CDE, defects schedule for the specific component can be retrieved and displayed at the front-end. After that, bridge practitioners can assess the condition of the component by considering all the defects collectively and provide maintenance proposals based on the condition markings of the whole bridge.

For the future work, multi-source data, information and knowledge (e.g., life-cycle information, dynamic responses captured by sensors, traffic, weather) can be fused to provide holistic feedback (e.g. early warnings, inspection advice, optimised maintenance planning, etc.) to the physical bridge with more sophisticated assessment and analysis, thus improve the resilience of the bridge network.

ACKNOWLEDGMENTS

This work was supported by the project - DigiBridge KTP (InnovateUK Grant Reference Number 10003208) and industrial partner of the project, Centregreat Rail Ltd.

REFERENCES

- Adibfar, A. and Costin, A. M. (2022) 'Creation of a Mock-up Bridge Digital Twin by Fusing Intelligent Transportation Systems (ITS) Data into Bridge Information Model (BrIM)', *Journal of Construction Engineering and Management*, 148(9), p. 04022094. doi: 10.1061/(ASCE)CO.1943-7862.0002332/ASSET/FC314C15-6F20-44EC-9DB3-8026716583E9/ASSETS/IMAGES/LARGE/FIGURE9.JPG.
- Allah Bukhsh, Z. et al. (2020) 'Maintenance intervention predictions using entity-embedding neural networks', *Automation in Construction*, 116, p. 103202. doi: 10.1016/J.AUTCON.2020.103202.

- Barrile, V. *et al.* (2019) 'UAV Survey of Bridges and Viaduct: Workflow and Application', *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11622 LNCS, pp. 269–284. doi: 10.1007/978-3-030-24305-0_21.
- Borrmann, A. *et al.* (2018) 'Building information modeling: Technology foundations and industry practice', *Building Information Modeling: Technology Foundations and Industry Practice*, pp. 1–584. doi: 10.1007/978-3-319-92862-3/COVER.
- Gao, Y. *et al.* (2023) 'AIoT-informed digital twin communication for bridge maintenance', *Automation in Construction*, 150, p. 104835. doi: 10.1016/J.AUTCON.2023.104835.
- Gao Y, Li H and Xiong G (2023) 'An Efficient and Resilient Digital-twin Communication Framework for Smart Bridge Structural Survey and Maintenance', *29 th International Workshop on Intelligent Computing in Engineering*. doi: 10.7146/aul.455.c207.
- Isailović, D. *et al.* (2020) 'Bridge damage: Detection, IFC-based semantic enrichment and visualization', *Automation in Construction*, 112, p. 103088. doi: 10.1016/J.AUTCON.2020.103088.
- Mohammadi, M. *et al.* (2023) 'Integration of TLS-derived Bridge Information Modeling (BrIM) with a Decision Support System (DSS) for digital twinning and asset management of bridge infrastructures', *Computers in Industry*, 147, p. 103881. doi: 10.1016/J.COMPIND.2023.103881.
- Network Rail (2022a) *Level2 Business Process Management of Structures*.
- Network Rail (2022b) *Level3 Manual Structures , Tunnels and Operational Property Examinations Approvals*.
- Nguyen, S. T. and La, H. M. (2020) 'A Climbing Robot for Steel Bridge Inspection'. doi: 10.1007/s10846-020-01266-1.
- Sutter, B. *et al.* (2018) 'A semi-autonomous mobile robot for bridge inspection', *Automation in Construction*, 91, pp. 111–119. doi: 10.1016/J.AUTCON.2018.02.013.
- Yang, J. *et al.* (2022) 'Intelligent bridge management via big data knowledge engineering', *Automation in Construction*, 135(January), p. 104118. doi: 10.1016/j.autcon.2021.104118.
- Yoon, S. *et al.* (2022) 'Seismic fragility analysis of deteriorated bridge structures employing a UAV inspection-based updated digital twin', *Structural and Multidisciplinary Optimization*, 65(12), p. 346. doi: 10.1007/S00158-022-03445-0.

A Distributed Twin System for Managing Asset-process Interactions in Highway Infrastructure Systems

Mengtian Yin¹, Tom Kelly¹, Pieter Pauwels², Liming Xu¹, Brian Sheil¹, Junxiang Zhu¹, Yue Xie¹, Ioannis Brilakis¹

1) Department of Engineering, University of Cambridge, United Kingdom. Email: my424@cam.ac.uk, twk22@cam.ac.uk, lx249@cam.ac.uk, bbs24@cam.ac.uk, jz652@cam.ac.uk, yx388@cam.ac.uk, ib340@cam.ac.uk

2) Department of the Built Environment, Eindhoven University of Technology, Netherlands. Email: p.pauwels@tue.nl

Abstract:

Highway agencies face challenges managing scattered asset data across maintenance processes and information systems, obstructing efficient retrieval of dynamic cross-system road information for timely interventions. This paper presents a Digital Twin (DT)-based data federation framework to effectively manage fragmented systems and dispersed data for highway infrastructure operation and maintenance. The framework middleware can decompose users' queries and requests from different subsystems based on a metadata database and a distributed system architecture. The connected data ecosystem enables dynamic communication between different asset systems, ensuring efficient maintenance planning and process coordination between different users and teams. The presented framework is demonstrated based on datasets and synthesised systems conforming to asset management practices adopted by United Kingdom (UK) National Highways.

Keywords: Highway infrastructure maintenance, Digital Twins, information management, distributed systems, digital transformation, process optimisation

1. INTRODUCTION

Highway infrastructure plays a vital role in a nation's transportation system. However, road networks are deteriorating rapidly with increasing traffic flows (Department for Transport, 2022). Degraded highway assets, such as pavement potholes and drainage blockages, pose significant safety issues for road users. As a result, managing effective maintenance of ageing road infrastructure is becoming increasingly vital yet challenging, necessitating crucial decision-making and timely maintenance interventions to prevent asset failures and traffic incidents.

High-quality data and functioning information systems are critical determinants for maintenance planning and execution. Unfortunately, poor information management (IM) has emerged as a barrier for transportation authorities in effectively executing maintenance operations. In the context of highway maintenance, IM refers to a systematic process of data definition, collection, storage, sharing, and usage of data related to the maintenance and operation of highways (Highways England, 2016). In the current practice, transportation agencies collect highway asset data for various business processes and store the data in different information systems. These information systems were developed by different software vendors and procured in different historical periods, so they often use incompatible data schemas and system architectures (National Academies of Sciences and Medicine, 2021). Due to the lack of interoperability and communication between these systems (see Figure 1(a)), asset data is inaccurate, incomplete, conflicting, and less reliable for use in analytics. As introduced in our previous survey (Yin, Reja, *et al.*, 2024), National Highways (NH) manages most motorways and A roads in the United Kingdom (UK); it currently employs over 50 information systems and data repositories for various asset systems and processes. For example, NH uses the CONFIRM system for routine maintenance management, the Pavement Asset Management System (P-AMS) for pavement systems, and Highways Agency Drainage Data Management System for managing drainage system. These systems manage the data of the same asset in different business logic; each uses different software data models that operate independently from each other. For example, a change of asset status in one system either cannot be synchronized to other systems or entails substantial manual efforts to update other database systems.

Poor quality of asset data also hinders effective maintenance activities. As illustrated in Figure 1(c), if a change of asset information from a business process cannot be synchronised from one system to other systems, other processes scheduled for the same asset in other systems may be interrupted. For example, a traffic sign is replaced in an emergent repair task (Process A) after a traffic accident, but this information is not sent to another system for routine maintenance management (System B). This causes other engineering teams to receive instructions to inspect or clean the traffic sign that has already been removed (data redundancy effect). The uncoordinated asset data management and maintenance operations lead to economic loss and safety risks to both road users and site workers. It would be beneficial to (a) have a service for end users to retrieve information of assets and processes produced in other systems for planning, design, and coordination; and (c) establish linkages between systems for information sharing.

To address the problems of system fragmentation and data redundancy, this paper presents a novel Digital

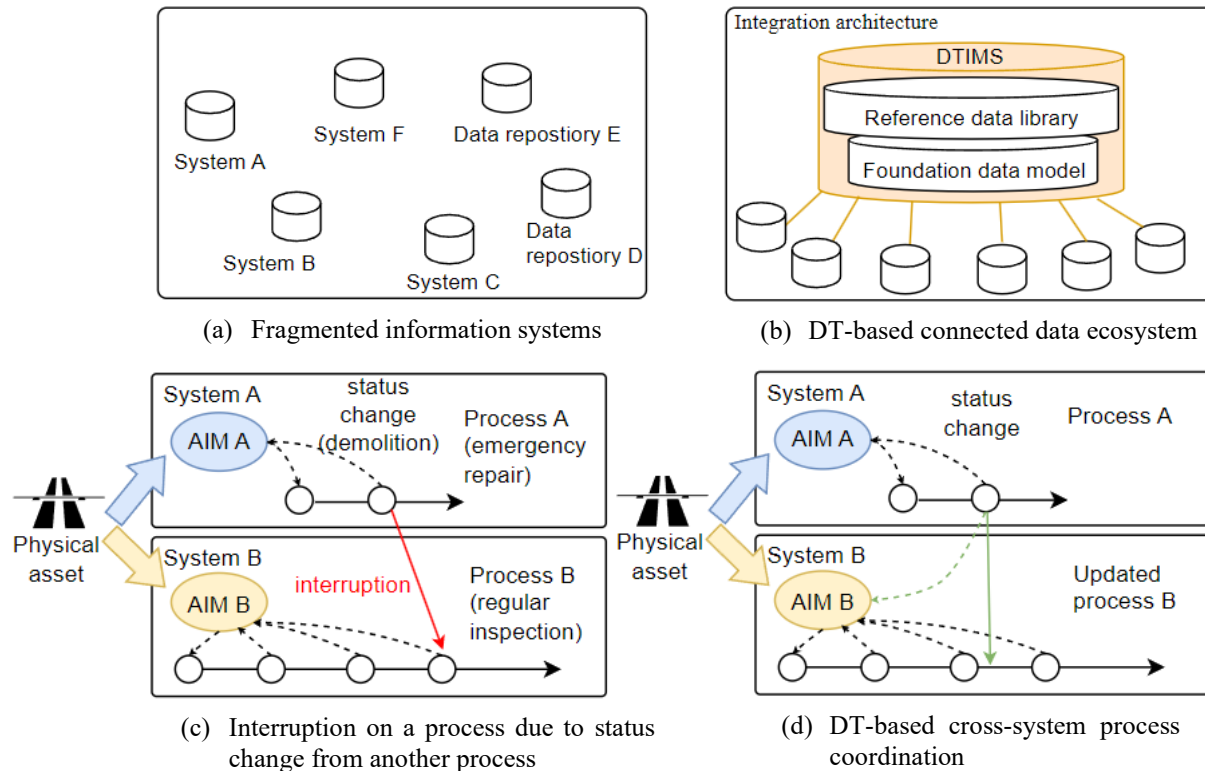


Figure 1. Problem statement (left) and the proposed DT-based solution (right).

Twin-based Data Federation (DTDF) framework for managing asset-process interactions across heterogeneous information systems in highway infrastructure maintenance. Grieves first describes the concept of Digital Twin (DT) in manufacturing domain: “a set of virtual constructs that fully describes a potential or actual physical product from the micro atomic level to the macro geometric level”. A twin system consists of a physical entity, a DT, and there are bi-directional data flow between physical and digital spaces. According to the UK National Digital Twin Programme (2020) of the UK, a DT in the built environment field refers to a digital representation of assets, processes, or systems. Our proposed DTDF framework offers a method to connect the fragmented information systems within an integration architecture, as presented in Figure 1(b).

The DT-based connected data ecosystem facilitates seamless data integration and sharing in federated Asset Information Models (AIMs), therefore resolving challenges in cross-system process coordination, as shown in Figure 1(d). This study is among one of the first studies to tackle the system fragmentation problem in highway asset management domain. The contributions of this research can be outlined as follows:

- 1) This study presents a new data federation method for highway IM based on ontology-based metadata modelling. The DTDF framework generalises the main layers and components of the distributed system architecture to handle data queries towards heterogeneous highway asset systems.
- 2) This study demonstrates how the proposed system integration approach can effectively check the conflicts between asset interventions at different levels of maintenance processes.

The rest of this paper is structured as follows. Section 2 reviews the relevant studies about the existing IM practices in highway maintenance domain and the DT concept. Section 3 presents the proposed DTDF framework. Section 4 presents a case study in a real-world highway road. Section 5 discusses and concludes this research.

2. INFORMATION MANAGEMENT IN HIGHWAY INFRASTRUCTURE O&M

Highway agencies started to implement asset management principles (AMPs) in the 1990s (Haas and Hudson, 2015), which represented a systematic coordination of investment, planning, design, construction, operation, maintenance, and demolition of infrastructure assets to maximise their lifecycle values (Uddin, Hudson and Haas, 2013). Asset management principles emphasise asset lifecycle data integration based on Asset Management Systems (AMSs), which refers to a set of interrelated and interacting elements of an organization and the processes needed to achieve Asset management objectives (Hastings and Hastings, 2021). IM takes place within the AMS, or a project management framework (Davidson *et al.*, 2022), where Project Information Models (PIMs) and AIMs are produced and exchanged during the project delivery phase and operational phase, as specified in ISO-19650 (Davidson *et al.*, 2022).

Modern transportation AMSs are becoming increasingly complex, involving many different subsystems (Asghari and Hsu, 2021). In general, there are two kinds of information systems that manage AIMs in the highway domain. The first kind of system looks after a specific asset class and supports professional activities regarding an individual asset system, such as condition investigation and rehabilitation. Typical asset systems include pavement management systems (PMS), bridge management systems (BMS), drainage management systems (DMS), geotechnics management systems, and electricity monitoring systems (Haas and Hudson, 2015). Each asset system integrates relevant contextual information to support its unique workflows. For example, DMSs often incorporate regional flooding data to assist in the decision-making of drainage system maintenance (National Highways, 2023). The second kind of systems concentrate on specific maintenance processes. There are different levels of maintenance processes managed by various agencies, departments, and contractors, such as project-level maintenance, routine maintenance, emergent maintenance, and natural hazard management. Developing a software system that handles all relevant data and operations is common practice for process management in highway maintenance.

Although AMS frameworks aim for seamless interconnection between all subsystems, the integration of existing asset systems and maintenance systems is inadequate and challenging. This causes asset data located in different systems to be inconsistent. According to a report by the Danish Road Directorate (Ebbesen, 2021), the existing AMSs are out-of-date and fragmented, obstructing a systematic coordination across different assets. Data warehousing and data distribution are two primary means of system integration (Li, 2018). Data warehousing is the process of gathering all information from different sources into a centralised data repository (Oti and Gharaibeh, 2020). A centralised database simplifies data management and data retrieval, but it necessitates significant development costs and induces problems with system complexity and data storage. Therefore, real-world AMSs only reach a certain level of data warehousing for specific asset types or processes. On the other hand, data distribution refers to an architectural approach that enables the access and retrieval of data distributed in multiple, disparate sources (Haas, Lin and Roth, 2002). Even though distributed systems allow flexible data sharing, synchronisation, and cross-database queries to support various asset management use cases, it is difficult to handle disparate data formats, referencing systems, and the heterogeneity of various software components.

Digital Twins are emerging as promising solutions for tackling such IM problems. The DT-driven Information Management Framework (IMF) proposed by the UK National DT programme (Hetherington and West, 2020) consists of three essential components: Foundation Data Model, Reference Data Library, and Integration Architecture. Foundation Data Model is an upper-level ontology that addresses problems like space-time, actuality, and granularity. So far, National DT programme has already published a 4D ontology to represent the temporal properties of data (NDT, 2023). Reference Data Library offers controlled vocabularies with a set of classes and attributes to describe domain-specific knowledge. Integration Architecture stands for a digital system that manages multiple asset information systems and offers protocols for data authorisation, sharing, discovery and retrieval across distributed users and providers.

3. PROPOSED DATA FEDERATION SOLUTION

3.1 Overview of the proposed DTDF framework

The proposed DTDF framework provides a systematic approach to integrating dispersed asset data and the associated processes, which includes two parts: (a) a distributed system architecture to integrate fragmented asset systems; and (b) workflows for highway organisations and practitioners to operate twin systems.

The main purpose of the design and development of the distributed twin system is to integrate fragmented AIMs and associated processes stored in different information systems. Specifically, integration refers to the ability to (a) discover and query data stored in different database and file systems and (b) synchronise the related data elements across different subsystems. The DTDF framework focuses on two important functional features of data and system integration for managing asset-process interactions:

- 1) Retrieval of information of assets and processes for users to efficiently coordinate, planning, and design maintenance activities. For example, maintenance planners may need to know the scheduled activities of nearby assets, so they can identify conflicts or coordinate with other teams to share working zones.
- 2) Change propagation management to ensure consistency across systems. In other words, modifying asset data during project or operational phase in one system can influence AIMs and processes managed by other systems. Therefore, it is necessary to compute the impacts of data modification and notify the influenced systems.

These two functional features are closely related to the problems of asset-process interactions discussed in Section 1. Therefore, we devised the system architecture and workflows based on the requirements of these two features, and we abstract the components of the system architecture to ensure the scalability for different application scenarios and use cases.

3.2 System architecture design

The overview of the DTDF system architecture is shown in Figure 2. The distributed twin system contains three interconnected layers: user access layer, federation middleware layer, and data source layer. Additionally, there is a component for security and access control that interacts with all three layers.

User access layer interacts with end users (e.g., maintenance planners) and applications that require retrieval of federated data. End users operate a graphical user interface (GUI) to discover published DT models (i.e., asset systems), query the required data (e.g., nearby assets and recent events). The incoming requests are routed to the federation middleware layer for processing.

Federation middleware layer manages the data flow between the user's request and the various data sources (subsystems). The federation middleware breaks down the queries and interacts with different systems using APIs (Application Programming Interfaces). The ontology and data standardisation layer provides a common ontology to clearly define the domain knowledge (e.g., classification of assets), a metadata schema, and a metadata database. Metadata refers to data that describes data of assets and process stored in subsystems. Metadata includes key description information, such as asset ID, process ID, and system ID of the associated information system. A metadata schema is a structural representation that describes the data schema of metadata. The data aggregation and transformation engine perform ETL (extraction, transform, load) to convert the raw data objects into standardised formats (e.g., Resource Description Framework (RDF)) and send them back to user access layer. Raw data objects may be validated using predefined constraint rules.

The data source layer contains all underlying information systems and data repositories that need to be integrated within the connected data ecosystem. The connected subsystems include different asset systems, such as PMS, BMS, and DMS. Due to the heterogeneous software architecture of different subsystems, a connector is needed for each subsystem to connect with the federation middleware layer.

Figure 3 shows the structure of a metadata schema (ontology) and data examples in the metadata database in form of RDF instance graphs. The minimal required metadata includes (a) traffic elements (e.g., road number, road section); (b) road network assets; (c) processes associated with different assets; and (d) system. The metadata database stores linkage data between AIMs, processes, and systems to allow federation middleware to communicate with correct subsystems for a given data query. For example, the routine maintenance information of pavement assets can be found in a system called Confirm.

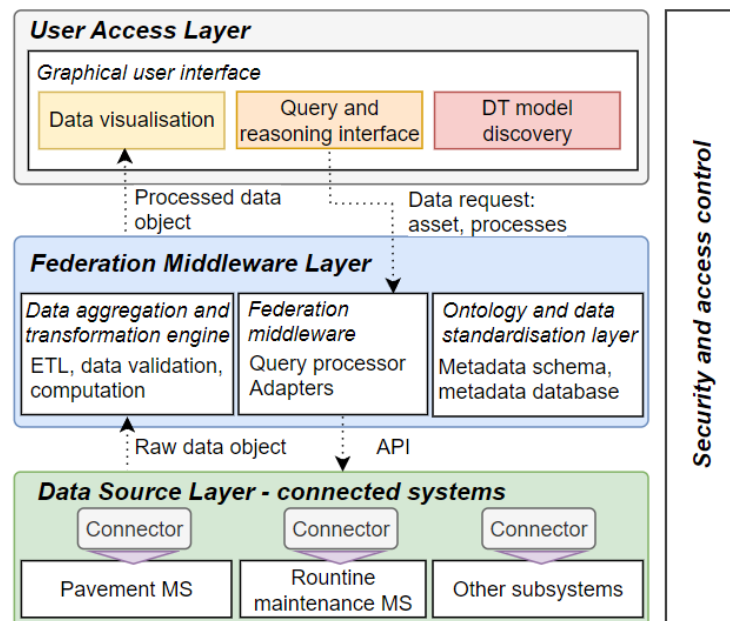


Figure 2. System architecture of the distributed twin systems for integrated highway asset management. Dotted line denotes information flow.

3.3 Workflows

The relevant workflows within the DTDF framework are shown in Figure 4. We consider two kinds of workflows: (a) the workflow to facilitate the data federation; and (b) the workflow for end users to access cross-system data and perform their processional maintenance activities.

(1) Data federation workflow

Apart from necessary software development according to the system architecture described in Figure 2,

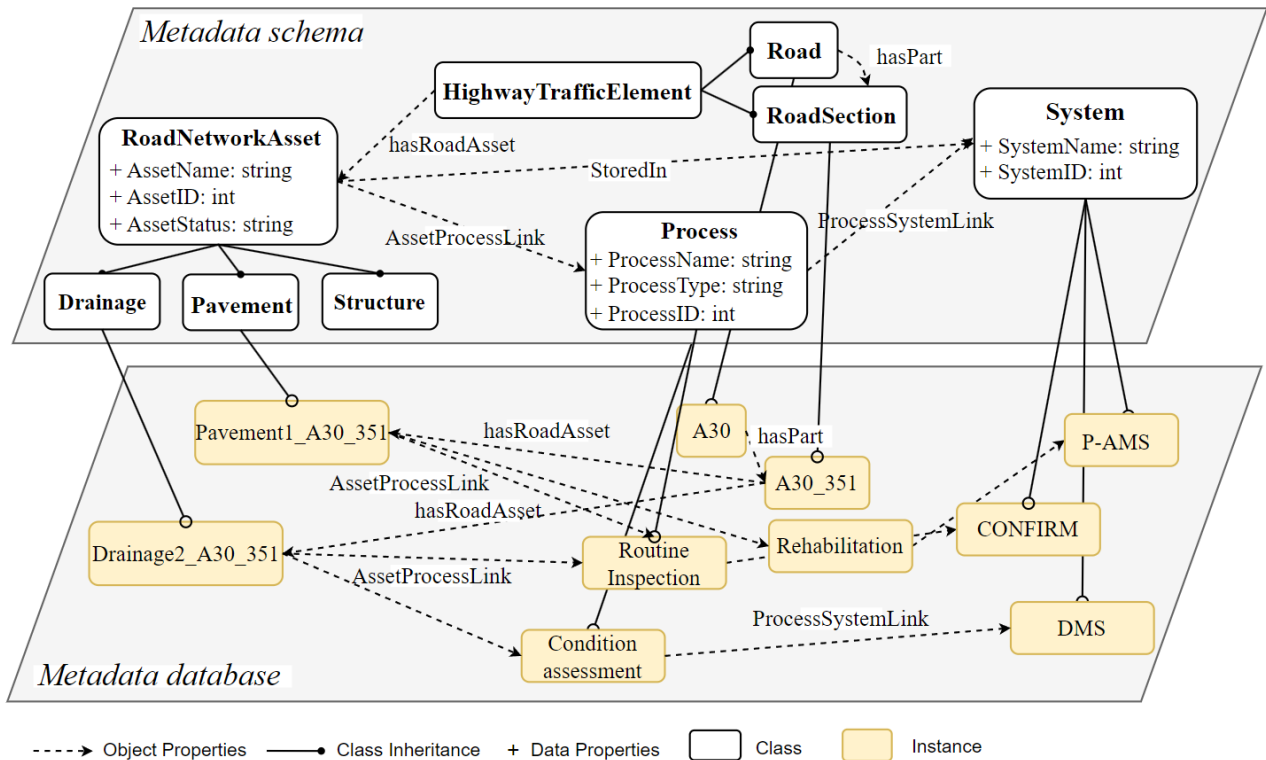


Figure 3. Ontology-based metadata schema and metadata database in the federation middleware of the highway twin system.

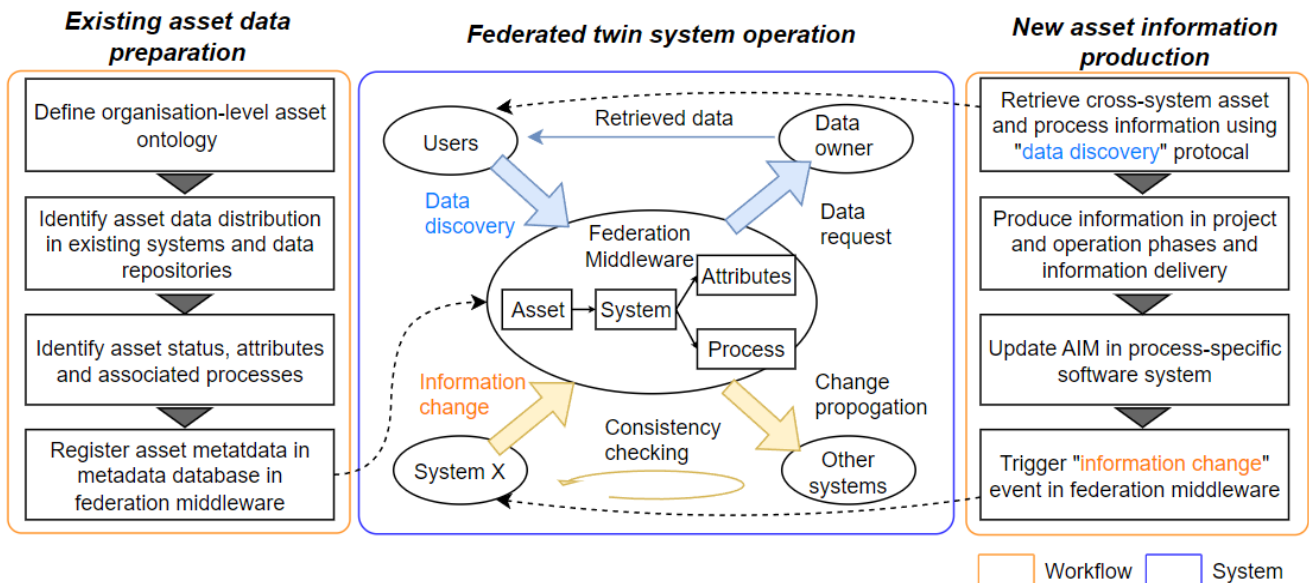


Figure 4. Workflows for facilitating DTDF framework (left) and for end users to access cross-system data (right).

Highway organisations should appoint a party (e.g., data management department) to organise the existing datasets and systems to achieve data federation and integration. This IM process is accompanied by the system development process. The workflow for data preparation is as follows.

- Develop a common ontology to clearly define the classification and hierarchy of assets, processes, and systems. The structure of the ontology is highly dependent on the existing asset management and IM practices in organisations.
- Identify the existence of all road assets and the distribution of asset data in different software systems, file systems, and data repositories. For example, pavement asset data may be managed by PMS, Maintenance management systems (MMS), and traffic management systems simultaneously.

- c) Identify asset attributes, such as construction and condition data, as well as the associated maintenance processes in different systems. For instance, a routine MMS may schedule regular visual condition surveys for gullies, and a DMS may arrange professional inspection activities for the same drainage system.
- d) Collect different subsystems' data models of the identified assets and relevant processes. Software vendors collaborates with highway organisations to develop more specific metadata schema based on the common ontology developed in stage (a), information requirements of use cases, and the existing data models, following the IM framework of ISO 19650 (Davidson *et al.*, 2022). Eventually, metadata is generated and registered in the metadata database of federation middleware. The metadata stores key information to allow federation middleware to be aware of which subsystem the data is stored for processing a query. For example, when a user submits a query: *"retrieve all assets in A30 and the schedules of their maintenance activities in November 2024"*, the metadata database would offer information of identifiers of assets, processes and systems, then API requests were sent to respective systems for querying detailed schedules.

(2) Workflows for end users

As shown in Figure 2, the DTDF provides a user access layer with GUIs for end users to query cross-system data. The design of GUIs and query interfaces depends on the service provided to end users. The generic workflow can be described as follows:

- a) Users who log in to the federated twin system operate the GUI to discover published DT models (e.g., pavement network models from pavement asset systems) in 2D/3D geospatial scenes and request the required data for their professional tasks. The design of the data query protocol is based on metadata schema, showing all available data elements shared by different subsystems. The data request is partitioned into sub-queries in the federation middleware and sent to subsystems. The returned raw data objects are processed into standardised formats in a data aggregation and transformation engine before sending back to end users. Apart from data retrieval, inference and computation are also be operated to check the clash between maintenance activities.
- b) Users optimise their work plans after acquiring the asset and process data retrieved from different systems. For example, a maintenance engineer may adjust the lane closure dates to work with other teams concurrently after obtaining the information of recent activities planned by other departments. Afterwards, users perform their professional tasks and produce new information in project and operational phases.
- c) Asset data managers handle the new information delivery and update AIMs in the corresponding information system. For example, the information of a repair task in a routine maintenance is updated in a routine MMS.
- d) Subsystems periodically communicate with federation middleware to update the change of information. To achieve so, an event system is developed in the federation middleware to control propagation of information change in one system to other systems. The event system first register duplicated attributes between different asset data models and dependencies between AIMs and processes in different systems. For example, the status information of assets can influence many cyclic maintenance activities. When an information change event occurs, the event system identifies the related asset attributes and processes using metadata, and notifies other influenced subsystems.

The above workflows allow highway participants to effectively retrieve information that is generated in other business processes. This information helps them coordinate with other departments, agencies, and contractors for maintenance planning and execution.

4. CASE DEMONSTRATION

We conducted a case study on a real-world highway to demonstrate the proposed DTDF framework. Specifically, we implemented a prototype called a Highway twin system (HTS) that contains a user access layer, a federation middleware layer, and a data source layer. We developed several synthesised asset information management systems and MMSs in the data source layer, all working independently. We then collected asset data of a long road segment in the UK - Chiverton to Carland Cross, of the A30. The selected road segment includes several roundabouts, overpasses, underpasses, and intersections with other roads.

In this case study, we consider a lighting facility outage event. In this situation, maintenance managers should arrange an engineering team to inspect the associated underground utilities and repair problematic asset components. This activity involves lane closures and the excavation of pavements. In this scenario, the HTS is used to retrieve assets in the same road section and the associated maintenance activities scheduled in different asset systems. This helps different teams coordinate intervention activities considering the dependencies of different assets.

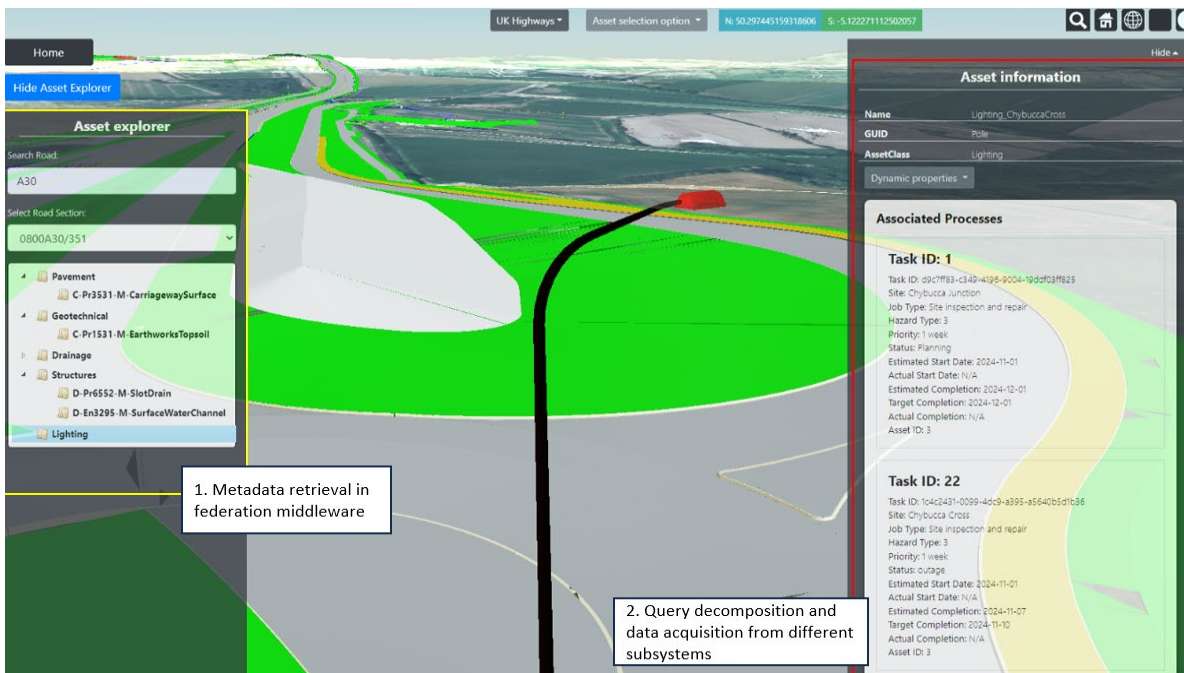
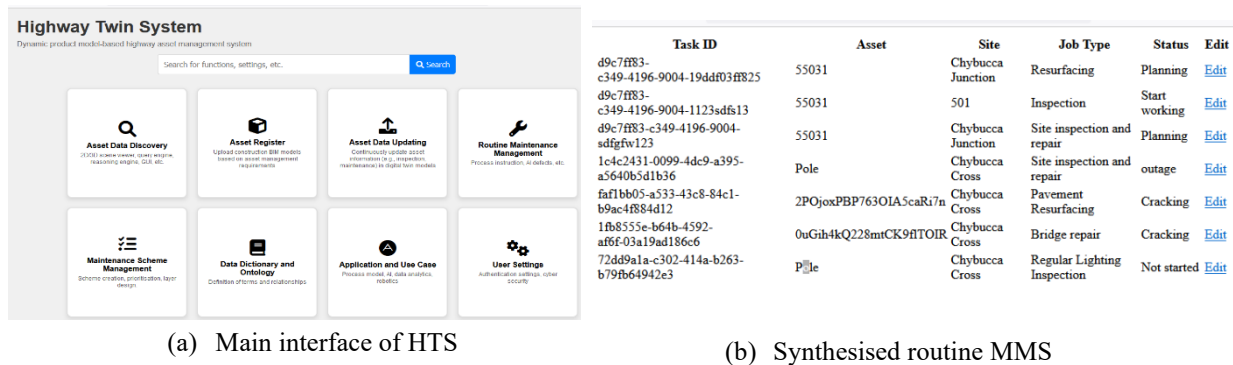
4.1 Implementation details

We implemented the HTS using Django MVC (Model-View-Controller) framework. We created a Django project to manage all components in the user access layer and the federation middleware layer, and several different Django projects that simulate the operation of subsystems. Every project uses a unique port, and they communicate with each other using HTTP protocols. Within the federation middleware, we use the official National Highways ontology (Department for Transport, 2020) as a common organisation-level ontology, and we deployed an ontology-based metadata schema by extending the National Highways ontology with constructs of the existing data models used by different subsystems, including P-AMS (TRL Software, 2022) and CONFIRM (Highway Agency, 2009).

In the user access layer, we use Cesium (2018), a web-based 3D geospatial development framework, to display data and allow end users to interact with the GUI. The implemented prototype is presented in Figure 5. The HTS offers a main interface (see Figure 5 (a)) for asset data managers to manage the connection with subsystems, metadata registration, and data specifications within the federation middleware. End users can view and retrieve distributed asset data in a GUI in the user access layer (see Figure 5 (c)). Figure 5 (b) shows an interface of a synthesised routine MMS in the data source layer, listing work orders for different assets.

4.2 Data collection

We collected asset data of the selected road segment in the forms of Industry Foundation Classes (IFC) specifications (buildingSMART International Ltd., 2019) from industry partners. The collected data involves a diverse range of asset classes, including pavement, earthwork, drainage, utilities, line markings, lightings, traffic signs, and structures. Due to the data confidentiality issues, we generated synthesised maintenance work orders



(c) Data query and visualisation interface in HTS.

Figure 5. Prototype of the federated twin system for highway maintenance management.

that conform to the data models of the MMSs used by the NH.

4.3 Results

The resulting RDF instance graph (metadata database) contains 1498 triples, 17 class counts, and 228 instances. We operate the HTS prototype to query asset and process data distributed in different subsystems, as illustrated in Figure 5. Given an example of a lighting outage event, we first retrieve the road number and road section where the lighting facility is placed. As shown in the left-hand side of the interface, different classes of assets located at the selected road section can be returned from the metadata database. Afterward, we can select a specific asset instance or asset group and request its associated process information stored in distributed subsystems, as shown in the right-hand side of the interface in Figure 5. This involves a multi-level data retrieval process as follows:

- An ontology-based metadata retrieval is first operated in the federation middleware to retrieve any processes linked to the selected asset and systems linked to the processes, as shown in Listing 1. The query processor extracts key information, such as ProcessID and SystemID, from SPARQL queries to perform next-level data query in subsystems.
- Having obtained a list of AssetID, ProcessID, and SystemID, the query processor decomposes the query into sub-queries for different subsystems before an adapter sends these sub-queries to subsystems using HTTP GET protocol.
- When a subsystem, such as a routine MMS, receives the data request, a connector transforms the data request into queries that comply with the local database settings and schema. For instance, a query for the synthesised MMS based on the Django MVC framework is based on Python and SQL commands, as presented in Listing 2. The retrieved results, such as scheduled tasks, are sent back to the federation middleware.
- The data aggregation engine gathers the requested JSON objects from different subsystems, populates RDF instance graphs, and conduct data inference before sending the processed data to the front end.

```
SELECT ?process ?processid ?processname ?system ?systemid ?systemname
WHERE {{
    ?asset rdl:GUID "{guid}"^^xsd:string .
    ?asset rdl:AssetProcessLink ?process .
    ?process rdl:ProcessID ?processid .
    ?process rdl:ProcessName ?processname .
    ?process rdl:ProcessSystemLink ?system .
    ?system rdl:SystemID ?systemid .
    ?system rdl:SystemName ?systemname .
}}
```

Listing 1. Metadata retrieval in the federation middleware to acquire the associated processes and systems of an asset.

```
work_orders =
WorkOrder.objects.filter(asset__asset_id=asset_id, tasktypeid = process_id)
```

Listing 2. Data retrieval in the distributed subsystems using localised query mechanism.

```
SELECT ?task1 ?task1type ?task2 ?task2type
WHERE {{
    ?task1 rdf:type rdl:Task.
    ?task2 rdf:type rdl:Task.
    ?task1 rdl:ScheduleStart ?start1 .
    ?task1 rdl:ScheduleEnd ?end1 .
    ?task2 rdl:ScheduleStart ?start2 .
    ?task2 rdl:ScheduleEnd ?end2 .
    FILTER (?start1 <= ?end2 && ?start2 <= ?end1) .
    FILTER (str(?task1) < str(?task2)) .
    ?task1 rdl:RelatedSite ?site.
    ?task2 rdl:RelatedSite ?site.
    ?task1 rdl:TaskType ?task1type .
    ?task2 rdl:TaskType ?task2type.
}}
```

Listing 3. Clash detection of tasks retrieved from different asset systems.

We searched the assets and processes that are spatially close to the broken lighting and perform clash

detection, as shown in Listing 3. Table 1 demonstrates an example of searching for assets and processes from other systems that occur at the same time as the lighting utility inspection and repair task. Different engineering teams can coordinate their work based on the interdependencies of assets and processes. For example, a utility repair activity should precede a pavement resurfacing activity because the former may involve excavation of pavements.

Table 1. Conflicted processes of nearby assets in A30 across different systems.

Asset class	Associated process	Schedule start	Schedule end	System
Lighting	Utility inspection and repair	2024-11-01	2024-11-07	Routine MMS
Pavement	Pavement resurfacing	2024-10-30	2024-11-03	Pavement MS
Structure	Bridge deck repair	2024-11-01	2024-11-10	Structure MS

Finally, the lighting asset information is updated in the routine MMS after maintenance. This triggers an “information change” event in the federation middleware layer based on HTTP POST protocols. The change propagation can be inferred by predefining the dependency relationships between asset attributes and processes across all systems. In this case, the change of an attribute’s “maintenance date” in the MMS would influence the regular structure integrity inspection activities scheduled in an electricity management system (EMS). Therefore, HTS would be able to send messages to that EMS to notify the change of asset status.

Table 2 shows a qualitative comparison between the HTS and the existing AMSs and data lake method adopted by the UK National Highways. The HTS allows cross-system retrieval of process information, but the existing AMSs fail to provide functions to obtain task data from different asset systems. Based on this, the HTS also offers semantic rules for conflict checking across various maintenance activities, providing opportunities for improved coordination. At last, the existing AMSs do not update or periodically update the latest asset information across information systems. In contrast, our event-driven design of the information updating mechanism allows more timely data synchronization.

Table 2. Qualitative comparison between HTS and existing AMSs.

Category	Cross-system process retrieval	Automatic asset-process constraint checking	Asset information update
HTS	Yes	Ontology-based conflict checking	Event-driven dynamic update
Existing AMSs	No	No programs for multi-system consistency checking	No update or periodically update

5. CONCLUSION

Data and system integration is a longstanding challenge for highway agencies to efficiently plan and perform maintenance activities. The difficulty originates from heterogeneous structures of different software components, incompatible data models, and poor communication protocols. As a result, the road asset data is often inaccurate and inconsistent across different systems and the associated processes are often uncoordinated. There is a lack of clear information management processes and enabling systems to effectively integrate dispersed datasets and fragmented systems in an economic way.

This paper presents a novel Digital twin-based data federation (DTDF) framework based on the UK NDT Information management framework. The DTDF framework includes a distributed system architecture and two workflows. The distributed system architecture offers a federation middleware to decompose queries from user access layer into sub-queries and communicate with subsystems. A distributed data query is enabled by (a) metadata schema and metadata database to retrieve the linkage between assets, processes, and systems; and (b) communication protocols and connectors to activate data retrieval and processing in subsystems. On the other hand, the DTDF framework also illustrates IM workflows for (a) preparing and registering the existing asset data in the federation middleware; and (b) end users to discover published DT models, retrieve cross-system data, and produce new information. Compared with the existing studies, such as (Le and Jeong, 2016), that purely focuses on data modelling for decision-making, this study provides a holistic solution to integrate fragmented asset systems, including system architecture, metadata modelling, and communication protocols.

We demonstrate the DTDF framework by implementing a system prototype called Highway Twin System (HTS), which consists of (a) a web-based software system that supports interactions with end users and processing federated queries; (b) several independent software systems that simulate the operation of fragmented asset information systems and maintenance management systems; and (c) communication protocols to connect different systems. We tested the HTS based on a case study on real-world highway road in United Kingdom. Different kinds of road assets, including pavements, drainage, utilities, and structures, are registered in the HTS. We then operate the HTS to handle a simulated lighting outage event. The results show that HTS can allow end users to effectively

search nearby assets that are located at the same road section and retrieve their attributes, as well as the associated processes distributed in different subsystems. These results can help different departments, engineering teams, and agencies in highway organisations coordinate their activities, avoiding clashes and improving productivity by scheduling co-intervention procedures.

This research benefits the highway industry by offering a practical and cost-effective data federation solution to manage fragmented systems and dispersed datasets without the need for re-development, restructuring, and migration of existing asset database systems. The connected data ecosystem can help different practitioners within HOs efficiently retrieve asset data and perform data analytics for different data-driven asset management processes. It may also be used by users from external parties to plan, design, construct, operate, and maintain other infrastructure assets, considering the interdependencies between infrastructure systems. Furthermore, a federated twin system makes it possible to cross-validate asset data and manage the influence of information change in one system on other systems, improving the data quality, data accuracy, and data completeness.

There are several limitations of this research. First, the scope of the distributed data query is small, and it only includes basic asset data attributes and process information, such as schedule start date and schedule end date. Second, it is inevitable to make efforts to develop connectors for different subsystems. It may be laborious to investigate the existing data schemas and construct appropriate data queries within subsystems. Furthermore, mapping data schemas of different subsystems to ontology-based metadata schema may be difficult when handling larger road networks or more complex asset systems. Third, the proposed framework does not provide a complete cybersecurity mechanism to restrict the user access and data usage rights. In the future work, the DTDF framework will be further extended and refined with more real-world use cases implemented and evaluated. Large language model-based ontology learning models (Yin, Tang, *et al.*, 2024) and autonomous agents may be explored to improve the automation level of the data model mapping and database retrieval processes.

ACKNOWLEDGMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 101034337. We wish to thank Dr. George Economides, Dr. Stefano Cavazzi, Dr. Matt Peck, Yogesh Patel, Kieran Mulvey, Simon Hayton, James Hewson, Sameer Kesava, and Alex Gillies for their contributions to this paper.

REFERENCES

- Asghari, V. and Hsu, S.-C. (2021) 'An open-source and extensible platform for general infrastructure asset management system', *Automation in Construction*, 127, p. 103692.
- buildingSMART International Ltd. (2019) 'Industry Foundation Classes (IFC)'. Available at: <https://technical.buildingsmart.org/standards/ifc>.
- CDBB (2020) 'The approach to delivering a National Digital Twin for the United Kingdom', p. 20. Available at: https://www.cdbs.cam.ac.uk/files/approach_summaryreport_final.pdf.
- Cesium (2018) 'Cesium - an open source Javascript library for world-class 3D globes and maps'. Available at: <https://cesiumjs.org/>.
- Davidson, S. *et al.* (2022) 'ISO 19650 Guidance 1: Concepts, Edition 3'.
- Department for Transport (2020) *Highways England and Local Authority Data Discovery*. Available at: <https://assets.publishing.service.gov.uk/media/5fb52dc4d3bf7f63d7075a44/DfT-and-Highways-England-data-discovery-report-accessible.pdf>.
- Department for Transport (2022) *Road traffic statistics*. Available at: <https://roadtraffic.dft.gov.uk/summary#:~:text=Road traffic by road type&text=motorway traffic increased by 13.3,billion vehicle miles of traffic>.
- Ebbesen, M. (2021) 'Data-driven asset management on the state roads of Denmark', *Bridge Maintenance, Safety, Management, Life-Cycle Sustainability and Innovations - Proceedings of the 10th International Conference on Bridge Maintenance, Safety and Management, IABMAS 2020*, pp. 3215–3219. Available at: <https://doi.org/10.1201/9780429279119-436>.
- Haas, L.M., Lin, E.T. and Roth, M.A. (2002) 'Data integration through database federation', *IBM Systems Journal*, 41(4), pp. 578–596. Available at: <https://doi.org/10.1147/sj.414.0578>.
- Haas, R. and Hudson, W.R. (2015) *Pavement asset management*. John Wiley & Sons.
- Hastings, N.A.J. and Hastings, N.A.J. (2021) 'ISO 55000 series standards', *Physical Asset Management: With an Introduction to the ISO 55000 Series of Standards*, pp. 595–621.
- Hetherington, J. and West, M. (2020) 'The pathway towards an information management framework-A "commons" for digital built Britain'.
- Highway Agency (2009) *Network Management Manual*. Available at: https://assets.publishing.service.gov.uk/media/5a8000cf40f0b62302690edd/nmm_part_Composit_0_9.pdf.

- Highways England (2016) *Information Management and Technology Strategy*. Available at: https://assets.publishing.service.gov.uk/media/5a7fffb840f0b62302690e77/Information_Management_and_Technology_Strategy.pdf.
- Le, T. and Jeong, H.D. (2016) 'Interlinking life-cycle data spaces to support decision making in highway asset management', *Automation in construction*, 64, pp. 54–64.
- Li, Z. (2018) *Transportation asset management: Methodology and applications*. CRC Press.
- National Academies of Sciences and Medicine, E. (2021) 'Guidebook for Data and Information Systems for Transportation Asset Management'.
- National Highways (2023) *GDMS - Geotechnical & Drainage Management Service*. Available at: <https://gdms.assetia.cloud/home>.
- NDT (2023) *Introduction to 4D Ontology*. Available at: <https://github.com/dstl/IES4/blob/master/4dOntologyIntro.md>.
- Oti, I.C. and Gharaibeh, N.G. (2020) 'A Note on Assessment of Transportation Infrastructure Data Management Practices', *Public Works Management & Policy*, 25(2), pp. 132–150.
- TRL Software (2022) *iROADS: infrastructure asset management system*. Available at: <https://trlsoftware.com/products/digital-asset-management/iroads/>.
- Uddin, W., Hudson, W.R. and Haas, R.C.G. (2013) 'Public infrastructure asset management', *(No Title)* [Preprint].
- Yin, M., Tang, L., *et al.* (2024) 'A deep natural language processing-based method for ontology learning of project-specific properties from building information models', *Computer-Aided Civil and Infrastructure Engineering*, 39(1), pp. 20–45.
- Yin, M., Reja, V.K., *et al.* (2024) 'How Can Digital Twins Be Used in Highway Maintenance? A Questionnaire Survey for Industry Practitioners'.

FINITE ELEMENT MODEL UPDATING ENHANCED BY DEEP LEARNING FOR REINFORCED CONCRETE CONTAINMENT VESSEL DIGITAL TWIN

Jiaji Wang¹, Si Fu², and Chawit Kaewnuratchadasorn³

1) Ph.D., Asst. Prof., Department of Civil Engineering, the University of Hong Kong, Hong Kong. Email: cewang@hku.hk

2) Ph.D. Candidate, Department of Civil Engineering, the University of Hong Kong, Hong Kong. Email: civilessfu@connect.hku.hk

3) Ph.D. Student, Department of Civil Engineering, the University of Hong Kong, Hong Kong. Email: chawit.ka@connect.hku.hk

Abstract: Model updating serves as the foundation for constructing a digital twin. In this study, a series of unidirectional seismic excitation tests were carried out on a 1/20-scaled reinforced concrete containment vessel (RCCV) specimen using a 6-DOF shake table. Based on the experimental setup, a finite element (FE) model of the RCCV was established and analyzed in Abaqus. Leveraging the experimental measurement data, a parallel FE-based nonlinear model updating framework was developed, and the model updating process was conducted to capture the complex behavior and failure modes of reinforced concrete shear walls under seismic loading with high fidelity. The model updating resulted in a significant improvement in the accuracy of the simulation results for the structural acceleration and displacement responses. During this process, a finite element simulation database consisting of 3400 pieces of data was generated. However, FE-based model updating is constrained by its high computational cost, long processing time, and challenges with convergence. To address this, we developed a Fourier Neural Operator (FNO) model to learn from the FE simulation database and predict the FE simulation results efficiently. It is observed that FNO exhibits excellent performance in predicting structural responses with various material and damping parameters. Compared to a single FE simulation, FNO achieved a speedup of over 5,000 times while maintaining acceptable accuracy. By employing the FNO model, the model updating process was significantly enhanced. The time required for FE model updating enhanced by FNO can be reduced from three days to less than 30 minutes. The findings highlight that the strategy of enhancing FE model updating with deep learning holds great potential in the field of constructing digital twins for large and complex structures.

Keywords: Shake table tests, Finite element (FE) simulation, Model updating, Fourier Neural Operator (FNO), Digital twin

1. INTRODUCTION

In the event of a nuclear reactor accident resulting in the release of radioactive material into public space, containment structures would serve as the last barrier (Bachev, 2014). Therefore, reinforced concrete containment vessels (RCCVs) are crucial for the safety of Nuclear Power Plants (NPPs). RCCVs are designed to withstand extreme seismic events. To this end, researchers conducted seismic excitation tests on a 1/20-scaled RCCV specimen using a 6-DOF shake table in Taiwan, enriching the RCCV experimental dataset. In addition to experimental studies, numerical simulation is also a main way of investigating the structural seismic performance of RCCV. The hysteresis performance of reinforced concrete tends to exceed the elastic range in practice. Despite substantial advances in current numerical analysis methods for predicting and simulating structural mechanical behavior, directly and accurately capturing nonlinear responses remains a challenge. Researchers often use the FE method to calculate and simulate structural responses of RCCV scientifically. However, uncertainties and errors in FE modeling, including (1) uncertainties due to the modeling strategy, material properties, and boundary conditions; (2) grid sensitivity; and (3) errors caused by the computational methods, etc. (Marwala, 2010; Lin, 2020), make it demanding to build high-fidelity FE models and obtain satisfactory structural results.

In this research, considering the geometrical properties of the RCCV specimen where the thickness is much smaller than the circumference and the height, it can be regarded as a shell structure consisting of planar stress (membrane) elements. Studies have shown that employing 2D shell elements with biaxial material constitutive models can significantly improve the computational efficiency of FE analysis while maintaining accuracy compared to 3D elements with elaborate models (Zhong, 2005; Wang et al., 2021). Over the last four decades, various constitutive models based on different theories and assumptions have been developed for simulating RC shell structures, including the smeared crack model (Mosler, J. & Meschke, 2004), the nonlinear elastic constitutive models (Polak & Vecchio, 1993; Li & Ansari, 1999), the elastoplastic constitutive models (Yu, et al., 2010), the damage plasticity models (Syed & Gupta, 2015; Lee & Fenves, 2001), etc. Wang et al. (2021) have proposed a biaxial reinforced concrete constitutive model which has the potential to predict the overall mechanical behavior of RC shell structures with high fidelity, especially for shear failure.

To match nonlinear performance and further improve simulation accuracy, in practical applications, the constitutive model needs to be calibrated and validated. FE model updating methods have been developed for this purpose. FE model updating refers to transforming the task of adjusting and selecting the physical parameters of

a structure into a multi-objective optimization problem. This is achieved by identifying key physical parameters as optimization variables and formulating a suitable objective function tailored to specific requirements. Researchers have investigated the use of FE model updating for a wide range of complex structures. For instance, Lin et al. (2021) conducted a nonlinear FE model updating for a long-span cable-stayed bridge model, and the simulation results aligned well with the experimental measurements. Kurent et al. (2023) performed FE model updating on a 5-story hybrid timber-concrete building by using modal properties, effectively reducing the discrepancy between the measured and predicted natural frequencies. Girardi et al. (2020) introduced FE model updating in the modal analysis of ancient masonry buildings and integrated the analysis and updating process in the NOSA-ITACA code. Previous research has primarily concentrated on using approximate models instead of refined models, with a predominant focus on frequency responses, modal parameters, or responses under static conditions. This highlights the necessity of using high-fidelity models to study complex nonlinear dynamic responses in the time domain, which is crucial for a more comprehensive understanding of structural performance.

In addition to computational accuracy, FE model updating also faces challenges such as computational complexity, time cost, and convergence. Recent advancements in Graphical Processing Units (GPUs) and Artificial Intelligence (AI) have opened up promising avenues for combining FE model updating with AI techniques. Specifically, Neural Operators, as an innovative deep learning architecture, are emerging as powerful tools to reduce the computational time and avoid convergence issues of the FE analysis and model updating. Neural operators can directly learn a mapping from an arbitrary combination of the input field (including material and damping parameters and ground motion history) to the solution for an entire family of Partial Differential Equations (PDEs) in infinite-dimensional space (Li et al., 2020). Li et al (2020) introduced a novel Fourier Neural Operator (FNO) via direct parametrization of the integral kernel in Fourier space. FNO has an efficient architecture and can serve as a data-driven digital twin of engineering structures. With sufficient data, FNO can approximate the target solution field with arbitrary accuracy at any spatial and temporal discretization and show a huge speed advantage over traditional numerical PDE solvers. By leveraging the FE simulation data, enhancing model updating with data-driven FNO models has the possibility to speed up the task considerably. This research makes efforts to the following key points:

- (1) Developing a FE model updating framework for the RCCV digital twin to exactly reflect the actual structural behavior.
- (2) Validating the accuracy of the selected constitutive model by comparing the finite element simulation results with the experimental data from the RCCV shaking table tests. Furthermore, validating the rationality of the selected updating factors, objective function, and optimization algorithm.
- (3) Proposing an FNO model to predict the selected update parameter combinations and the RCCV structural responses. Using FE simulation data and experimental data to train this deep learning model.
- (4) Enhancing FE model updating through deep learning to establish digital twins. Revealing the enormous potential of integrating model updating with AI in structural engineering.

2. EXPERIMENTAL STUDY AND FE SIMULATION OF RCCV

2.1 Test Program Overview

A 1/20 scaled RCCV specimen was designed for conducting tests under both gravity and seismic effects using a 6-DOF shake table. This RCCV specimen was designed based on a prototype of an Advanced Boiling Water Reactor (ABWR) nuclear containment structure, and the 1/20 scale was chosen to ensure that the seismic inputs would cause sufficient damage to the specimen within the actual capabilities of the shake table. This specimen can be divided into four parts from top to bottom: the counterweights, the top slab, the main containment vessel, and the bottom slab. The seismic excitations input to the shake table is scaled from the 1999 Chi-Chi earthquake record and is loaded only in a horizontal N-S direction. The first step of the loading protocol procedure is to warm up by applying a white noise with a peak ground acceleration (PGA) of 0.02g for identification of the predominant period and damping ratio of the specimen. Then, a test sequence consisting of four seismic excitations, each lasting 20 seconds, is input to the shake table. The PGAs of these four seismic excitations are 0.28g, 0.54g, 0.81g, and 1.07 g, respectively.

During the experimental process, the seismic excitations input to the shake table, the output accelerations measured on the shake table surface, the output accelerations on the top of the counterweights, and the displacements on the top of the main containment vessel were measured and recorded. After the tests were completed, the condition of the RCCV specimen was checked and evaluated. It was observed that the vertical reinforcement yielded, and shear and flexural cracks appeared, but the concrete did not crush in the main containment vessel. Fu et al., (2025) provided a clear description of the setup, procedure, and results of these RCCV shake table tests, accompanied by figures.

2.2 FE Modeling and Simulation

In this research, we utilize the commercial FE software Abaqus to perform numerical simulations. We modeled the counterweights, the top slab, and the main containment vessel. For the parts of the counterweights

and the top slab, a linear solid element (C3D8) with a mesh size of 320 mm was applied. It is ensured that these parts are always kept in a linear state under dynamic excitations, in line with the actual scenario. For the main containment vessel part, a multi-layer element (S4R) with a mesh size of 144 mm was applied. A tie constraint was adopted to bind different parts of the model, and a fixed boundary condition was used to simulate the fixed connection between the RCCV specimen and the shake table. In this research, the in-plane horizontal and vertical accelerations measured on the shake table surface were set as the FE simulation inputs. Keeping the FE model geometry, boundary conditions, interaction constraints, and loading settings consistent and reasonable with the tests helps to carry out high-fidelity FE simulations and model updating. As for the choice of the constitutive model, a biaxial reinforced concrete constitutive model developed in the constitutive model toolbox THUC2 was adopted to precisely predict structural behavior through the User Material (UMAT) subroutine for ABAQUS Implicit solver (Wang et al., 2021). Wang et al. (2021) have an explicit explanation of the set of reinforcement and concrete parameters being defined.

After modeling, FE simulations were successfully carried out. Using Intel Core i9-13900KF, a single FE simulation on a single core took 2100 seconds.

3. FE MODEL UPDATING FOR RCCV DIGITAL TWIN

3.1 FE Model Updating Framework

For complex and delicate structures, the FE model updating includes three important concepts: the determination of the selected update parameters, the setting of the objective function, and the selection of the optimization algorithm. Applicable to the constitutive model used in this research, there are seven selected parameters used for model updating, which are three concrete material parameters (α_c , η_c , γ_u), two steel material parameters (α_s , α_u), and two Rayleigh damping proportionality factors (α_{dM} , β_{dK}). The combined relative error for accelerations and displacements based on the widely used root mean square error (RMSE) form is taken as the objective function to fully consider the differences between FE simulation and experimental data. For the optimization algorithm, both Bayesian optimization and heuristic optimization can be applied to solve complex optimization problems. Heuristic global optimization algorithms are able to give optimal solutions to cases within an acceptable time range, which matches the purpose of this research. Therefore, we adopted the Simulated Annealing (SA) algorithm, a heuristic approach that seeks the global optimum by exploring states with relatively small objective values in the neighborhood. In addition, multi-process parallelism was implemented using Python to improve the speed of model updating work. The FE model updating framework can be referred to in Fig. 9 in the paper by Fu et al. (2025).

3.2 FE Model Updating Process and Results

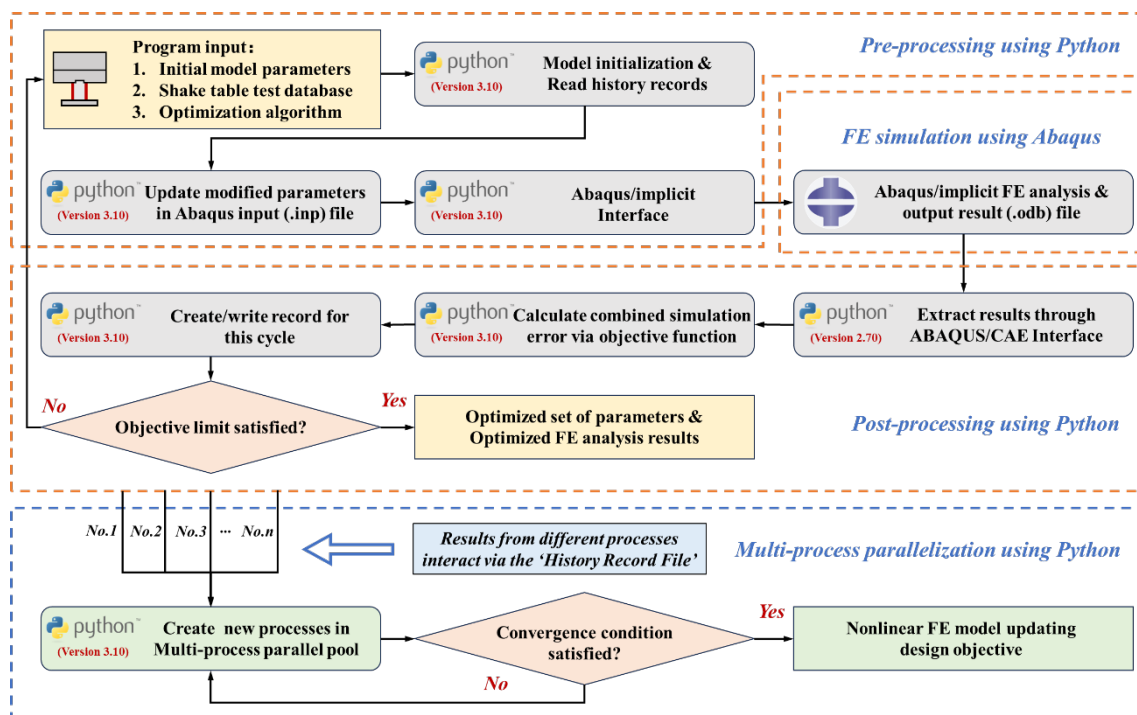


Figure 1. FE model updating process for RCCV digital twin

Figure 1 shows the FE model updating process for RCCV digital twin. Conducting 10-process parallel computing with Intel Core i9-13900KF, the FE model updating process took a total of 210 hours, forming a database of 3400 FE simulations. After the FE model updating work, the RCCV simulated response results had combined relative errors below 15%, and the results were improved by an average of 7.7% over the initial simulated response results. For additional comparison and context, using the equation for calculating the combined relative error, the maximum experimental measurement difference from the different measuring instruments is 3.0%. The RCCV shake table test design results, the test observations, and the updated FE simulation results are validated with each other to prove the effectiveness of the FE model updating.

3. FE MODEL UPDATING ENHANCED BY FNO FOR RCCV DIGITAL TWIN

3.1 FNO Model Development

A data-driven neural operator model containing six Fourier layers is trained using the FE simulation database to predict the seismic response based on material parameters and ground motions. Pre-trained FNO models provide a novel framework for efficient nonlinear model updating and response prediction. In the database, the 3400 data are divided into training set, validation set, and test set in the ratio of 73%: 15%: 12%. The training loss and the validation loss stabilized at 0.038 and 0.039, respectively with 500 epochs of training. Using NVIDIA GeForce RTX 4090, the total training time was 50 minutes and 22 seconds. It is observed that the data-driven FNO model can predict the dynamic structural responses with combined relative errors below 3% compared to the best FE simulation results, while the FNO inference time (0.4 seconds) achieves 5,250 times speedup compared to a single FE simulation time (2100 seconds). The results show that FNO has excellent performance in predicting the structural responses with different sets of update parameters.

3.2 FE Model Updating Enhanced By FNO Process

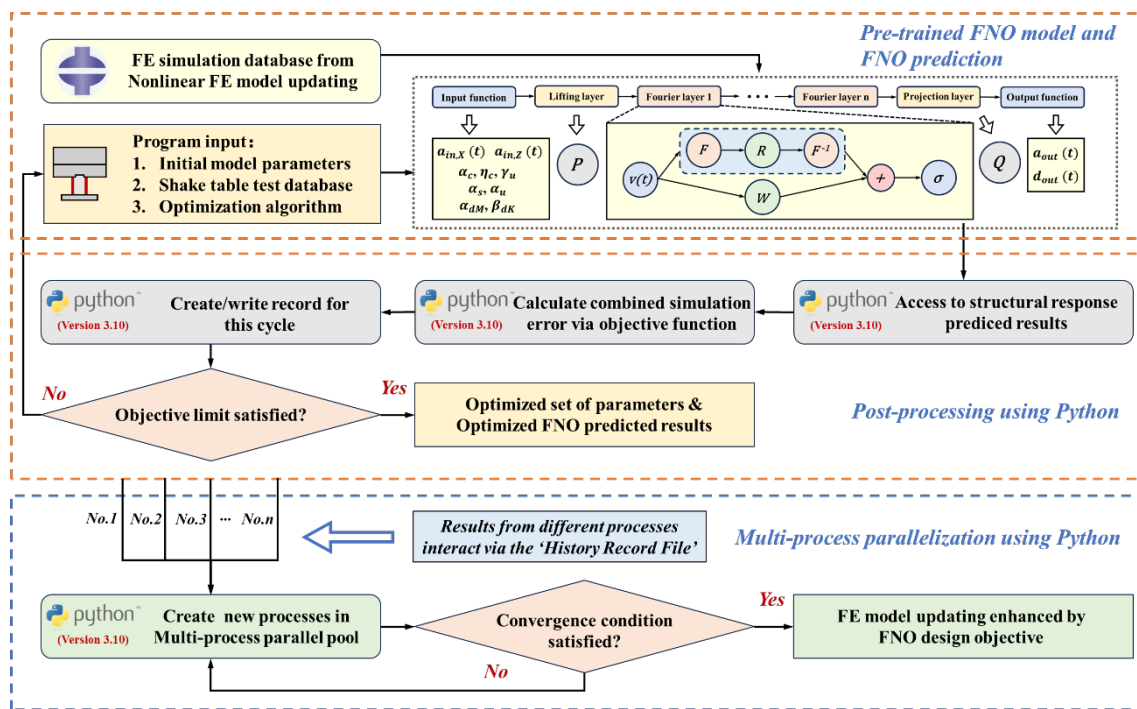


Figure 2. FE model updating enhanced by FNO process for RCCV digital twin

Figure 2 shows the FE model updating enhanced by FNO process for RCCV digital twin. Pre-trained FNO model was applied to replace the FE simulation process during model updating. Using the SA algorithm, the typical time for this process is 10 to 20 minutes, and the number of samples is 160000 to 300000. After obtaining the optimal update parameter combination, FE simulation was performed again using these parameters, proving the effectiveness of FE model updating enhanced by FNO. The results demonstrate that enhancing FE model updating by deep learning can significantly accelerate the model updating process while maintaining an acceptable level of accuracy, thereby enabling the rapid establishment of an RCCV digital twin.

4. CONCLUSIONS

Researchers conducted unidirectional seismic excitation tests on a 1/20 scaled RCCV specimen using a 6-DOF shake table aiming to better understand the actual performance and complex behavior of RCCV structures

under seismic effects. In this research, we combined Abaqus software with Python to establish a nonlinear FE model updating framework for RCCV digital twin, performed the model updating, and obtained the FE simulation database. Subsequently, we proposed an FNO model to predict the selected update parameter combinations and the RCCV structural responses. Then, we used the pre-trained FNO model to enhance the FE model updating process and successfully practiced it. The following conclusions can be drawn:

- (1) The biaxial reinforced concrete constitutive model developed by Wang et al. (2021) can predict and reflect the physical properties of the RCCV and provide a highly accurate simulation of the structural nonlinear behavior.
- (2) The developed FE model updating algorithm is highly effective for simulating large and complex structures and constructing the digital twins, enabling researchers to obtain high-fidelity structural responses with the combined relative error below 15% and the corresponding updated parameter combinations.
- (3) The application of a data-driven FNO model can significantly improve prediction speed by over 5,000 times compared to the FE simulation speed, predict the dynamic structural responses with a combined relative error below 3% compared to the FE simulation, and capture the selected update parameter trends with satisfactory accuracy.
- (4) The strategy of enhancing FE model updating through deep learning to establish digital twins have proven to be effective. Through high-performance GPU-based parallel computing, integrating model updating with AI has significant potential in the field of structural engineering.

5. FUTURE WORK

In the future, we aim to incorporate physical information and integrate the research frontier of time-series models in the AI field to make deep learning models accurate, efficient, concise, and generalizable. We will validate the effectiveness of the model updating strategy on longer time scales and larger scale structures, with the aim of realizing the development of a real-time digital twin model that reflects the structural state. We hope to contribute to driving innovation and digital transformation in the field of structural engineering.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial supports provided by the NCREE, Taiwan, and the University of Hong Kong Start-up Fund for New Staff and the Seed Fund for Basic Research for New Staff.

REFERENCES

- Bachev, H. (2014). Impacts of March 2011 earthquake, tsunami and Fukushima nuclear accident in Japan. *Tsunami Fukushima Nucl. Accid. Jpn*, 1-125
- Marwala, T. (2010). Finite element model updating using computational intelligence techniques: applications to structural dynamics. *Springer Science & Business Media*, 1-44.
- Lin, K., Xu, Y. L., Lu, X., Guan, Z., and Li, J. (2020). Cluster computing-aided model updating for a high-fidelity finite element model of a long-span cable-stayed bridge. *Earthquake Engineering & Structural Dynamics*, 49(9), 904-923.
- Zhong, J. (2005). Model-based simulation of reinforced concrete plane stress structures. *Dissertations & Theses, University of Houston*, 3186318, 1-24.
- Wang, J. J., Liu, C., Nie, X., Fan, J. S., and Zhu, Y. J. (2021). Nonlinear model updating algorithm for biaxial reinforced concrete constitutive models of shear walls. *Journal of Building Engineering*, 44, 103215.
- Mosler, J., and Meschke, G. (2004). Embedded crack vs. smeared crack models: a comparison of elementwise discontinuous crack path approaches with emphasis on mesh bias. *Computer methods in applied mechanics and engineering*, 193(30-32), 3351-3375.
- Polak, M. A. and Vecchio, F. J. (1993). Nonlinear analysis of reinforced-concrete shells. *Journal of Structural Engineering*, 119, 3439-3462.
- Li, Q. and Ansari, F. (1999). Mechanics of damage and constitutive relationships for high-strength concrete in triaxial compression. *Journal of Engineering Mechanics*, 125(1), 1-10.
- Yu, T. T. J. G., Teng, J. G., Wong, Y. L., and Dong, S. L. (2010). Finite element modeling of confined concrete-I: Drucker-Prager type plasticity model. *Engineering structures*, 32(3), 665-679.
- Syed, S. and Gupta, A. (2015). Seismic fragility of RC shear walls in nuclear power plant part 1: characterization of uncertainty in concrete constitutive model. *Nuclear Engineering and Design*, 295, 576-586.
- Lee, J. and Fenves, G. L. (2001). A return-mapping algorithm for plastic-damage models: 3-D and plane stress formulation. *International Journal for Numerical Methods in Engineering*, 50(2), 487-506.
- Lin, K., Xu, Y. L., Lu, X., Guan, Z., and Li, J. (2021). Collapse prognosis of a long-span cable-stayed bridge based on shake table test and nonlinear model updating. *Earthquake Engineering & Structural Dynamics*, 50(2), 455-474.
- Kurent, B., Ao, W. K., Pavic, A., Pérez, F., and Brank, B. (2023). Modal testing and finite element model updating of full-scale hybrid timber-concrete building. *Engineering structures*, 289, 116250.

- Girardi, M., Padovani, C., Pellegrini, D., Porcelli, M., and Robol, L. (2020). Finite element model updating for structural applications. *Journal of Computational and Applied Mathematics*, 370, 112675.
- Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., and Anandkumar, A. (2020). Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv*, 2010.08895.
- Fu, S., Chang, C. C., Kaewnuratchadasorn, C., Wang, J., Wu, C. L., Mo, Y. L., and Hsu, T. T. (2025). Nonlinear finite element model updating and Fourier neural operator for digital twin of reinforced concrete containment vessel under seismic excitations. *Engineering Structures*, 323, 119234.

13



Part 13

Education and Training in Construction

LIMITED DATA-ORIENTED WORKER INTENTION RECOGNITION METHOD IN WORKER-ROBOT COLLABORATION FOR CONSTRUCTION

Zhaoxin Zhang¹, Zaolin Pan², Yantao Yu³, and Liang Liu⁴

1) Zhaoxin Zhang, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: zhangzhaoxin@ust.hk

2) Zaolin Pan, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: zpanaq@connect.ust.hk

3) Yantao Yu, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: ceyantao@ust.hk

4) Liang Liu, Hong Kong INDICS International Information Technology Co., Ltd. Email: liuliangyork@sina.com

Abstract: Timely and accurate identification of workers' intentions in construction scenarios is crucial for seamless worker-robot collaboration. However, limited worker behavior due to varying behavioral styles and difficulties in collecting worker action data limit the practical application of existing methods that rely heavily on extensive worker action data. This paper addresses the dynamic nature of construction environments by proposing a few-shot worker intention recognition method. The proposed approach constructs worker intention query features using randomly sampled frame combinations and then applies metric learning to develop a few-shot worker intention recognition model. To validate the effectiveness of this method, a worker scaffolding installation action video dataset was used for the experiments on worker intent recognition. Given five categories with five worker action samples, the method achieved an accuracy of 71% in recognizing workers' intentions. The results demonstrate that the proposed method can effectively learn and detect novel worker actions with a minimal number of classified action videos, thereby improving model performance while reducing the number of required training videos. This approach not only reduces the labor needed for data labeling but also enhances the practicality of worker-robot collaboration in construction scenarios.

Keywords: Worker-robot collaboration, Intention recognition, Limited data, Meta learning

1. INTRODUCTION

High-risk construction environments and frequent safety incidents pose significant obstacles to the high-quality and sustainable development of the construction industry [1]. These accidents subject workers to severe injuries and life-threatening hazards, while poor working conditions and safety risks exacerbate labor shortages in the sector. Therefore, there is an urgent need to improve safety protocols and address labor shortages.

Automation and robotics have emerged as critical solutions to these industry challenges [2]. However, the construction environment is highly unstructured and dynamic, with frequent changes even during task execution. Single-task robots lack the flexibility required to handle such uncertain workplaces, relying instead on human workers' decision-making and problem-solving abilities. To ensure the effective application of robots in this domain, Worker-Robot Collaboration (WRC) should be established—a novel collaborative framework that leverages the strengths of both workers and robots. WRC technology offers a promising solution to the pressing needs faced by the construction industry [3]. In the realm of construction, such collaboration can liberate construction workers from tasks characterized by repetitiveness and high physical demands, thereby enhancing construction safety and productivity, as well as mitigating labor shortages and the aging workforce crisis [4,5]. To facilitate seamless worker-robot collaboration, robots should possess the capability to intelligently perceive, comprehend, and adapt to the intentions of workers [6], enabling assistance in task execution within dynamic construction environments.

However, most advanced technologies rely on conventional deep learning methods [7], which require large amounts of training data with accurately labeled worker intention information to develop high-performance vision-based algorithms. Consequently, numerous construction scene images must be collected, and worker intention type labels must be annotated for each video [8]. This manual process is time-consuming, costly, and labor-intensive, making the development of comprehensive and high-quality robotic perception databases a significant challenge [9]. The issue is further complicated by the need to recognize different perception intentions that frequently change with construction phases. Each time a new type of worker intention arises, the training database must be updated. Due to these factors, the performance of vision-based monitoring systems often deteriorates in practice.

To overcome the challenges of worker intention recognition with limited data, this study proposes a limited data-oriented worker intention recognition method in worker-robot collaboration for construction, which aims to minimize the required training data and reduce data labeling costs. The proposed method focuses on worker intention feature extraction and a feature learning strategy tailored to limited data scenarios. To validate the approach, experiments were conducted using a dataset of worker scaffold installation actions. Remarkably, even with only five worker action samples, the method achieved 71% accuracy in recognizing worker intentions.

The remainder of this paper is organized as follows: Section 2 provides a literature review of existing studies. Section 3 explains the worker intent recognition method under limited data in detail. Section 4 presents

the recognition performance of the proposed method. Section 5 discusses the results, and Section 6 concludes the study.

2. LITERATURE REVIEW

2.1 Worker Intention Recognition

Human-robot collaboration (HRC) research explores the physical and cognitive interactions between humans and robots to accomplish shared tasks, such as closely collaborating with cooperative robots to safely and effectively complete various joint activities [10]. In HRC, robots need to understand the current work status and worker intention [11] to determine their own tasks and movements. Typical tasks include collaborative assembly [12][13] and object handover [14][15]. For instance, in [12], the robot must comprehend the current steps of the assembly process and select the appropriate components for assembly. In [13], the robot provides assistance based on predictions of human trajectories. For object handover, it is essential for the robot to understand human intention, as smooth collaboration requires the robot to accurately interpret the object desired by the worker [15].

In the construction domain, HRC has gained increasing attention as a promising solution to free workers from repetitive and physically demanding tasks, thereby enhancing construction safety and productivity. It can potentially be applied to various construction activities, such as bricklaying [16], object handover [17], and wooden component assembly [18]. It is well known that the ability to understand ongoing work progress and interpret human intention is crucial for robots to adaptively collaborate and execute required tasks. Existing research has developed methods to guide robots in responding and executing tasks based on various sensors (e.g., tactile sensors [17], cameras [19], and electroencephalography [16]). Specifically, human intention (e.g., target of interest), human posture, or the movement of specific body parts (e.g., hands) has been extensively studied through a range of sensory inputs, including visual [20], accelerometric [21], muscular [22], and neural activities [23].

2.2 Sample Learning Methods under Limited Data

Although traditional machine learning algorithms typically require large training datasets, humans demonstrate the ability to learn and recognize new objects from only a few examples. To address this gap, researchers have explored various few-shot learning algorithms to efficiently acquire construction-related knowledge. Few-shot learning has been applied to computer vision tasks, including few-shot image classification, few-shot object detection, and few-shot video action recognition [24]. Few-shot video action recognition aims to achieve video classification capabilities by providing only a limited number of video sequences. Existing studies have effectively explored transfer learning [25–27], meta-learning [28–31], and metric learning [32–35] for few-shot data. Transfer learning algorithms operate on the assumption that knowledge obtained from data in the source domain can be transferred to the target domain through model fine-tuning. Meta-learning algorithms aim to learn generalizable knowledge by training models on several different classification tasks. Metric learning algorithms focus on creating feature embeddings and effective distance measures.

Few-shot action recognition has been validated on public computer vision datasets; however, there is a research gap in conducting few-shot action recognition on challenging construction sites. These sites are characterized by high dynamism and privacy sensitivity, making it difficult to collect large datasets of labeled video sequences. Notably, labeling data related to worker action intentions requires substantial expertise and additional time costs. Due to factors such as complex background variations, similarities between different actions, and occlusion of objects on-site, recognizing worker intent remains challenging [6]. Worker intent may change depending on the construction environment and the assembly components involved. Recent studies have applied few-shot learning techniques to construction site data. For example, Cui et al. [36] proposed using few-shot classification and contrastive learning to identify facade defects. Xu et al. [37] developed a meta-learning-based few-shot classifier for recognizing structural damage from images. Kim et al. [38] applied few-shot object detection to learn and monitor emerging targets, such as excavators on construction sites. Liang et al. [39] proposed a CLIP-based few-shot learning algorithm for temporary object recognition on construction sites. Wang et al. [40] proposes a deep learning-based method to detect fall-related site objects and their associated attributes to better capture site conditions for supporting field compliance checking.

3. METHODOLOGY

The proposed method aims to design a few-shot model for worker intention recognition that can learn and classify with only a limited dataset of worker intention.

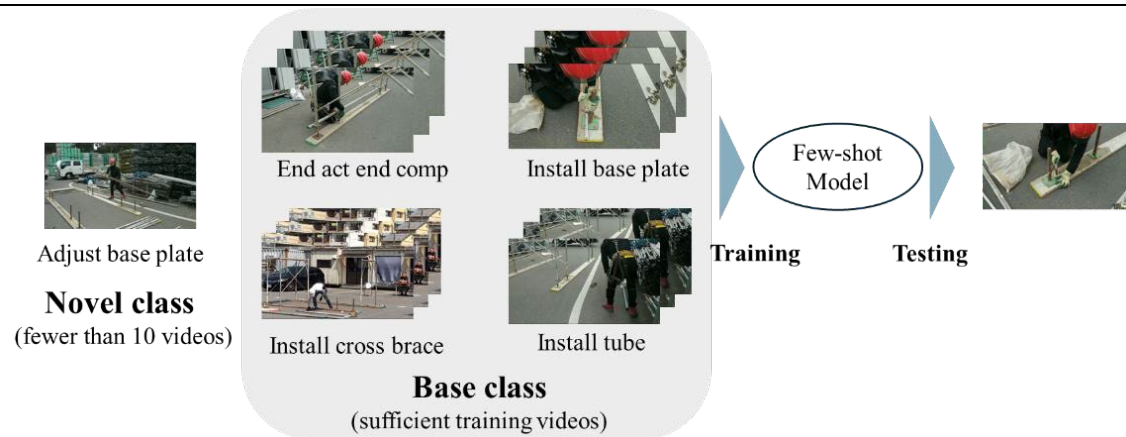


Fig. 1 Concept of few-shot learning

As illustrated in Figure 1, the proposed method aims to generate a model that can learn and classify worker intention based on a limited amount of video data. First, we define the variables used in this study. The objective of this paper is to accurately classify worker intentions that the model has not previously encountered. The small amount of labeled data during the model training phase is referred to as the 'support set.' We categorize each class in the few-shot scenario into k intention categories and n video data, commonly known as a k -way n -shot detection scenario. Let $Q = \{q_1, q_2, \dots, q_F\}$ denote a single query video with F frames. The goal is to classify Q into one of the classes $c \in C$. For class c , its support set S^c contains K videos, with the k^{th} video defined as $S_k^c = \{s_{k1}^c, s_{k2}^c, \dots, s_{kF}^c\}$.

3.1 Model Design

This process designs a few-shot learning model capable of acquiring meta-knowledge from training videos labeled with only base classes (i.e., how to learn worker intention from video data) and leveraging this knowledge to learn new types of worker intention. The proposed method consists of video feature extraction and prediction modules.

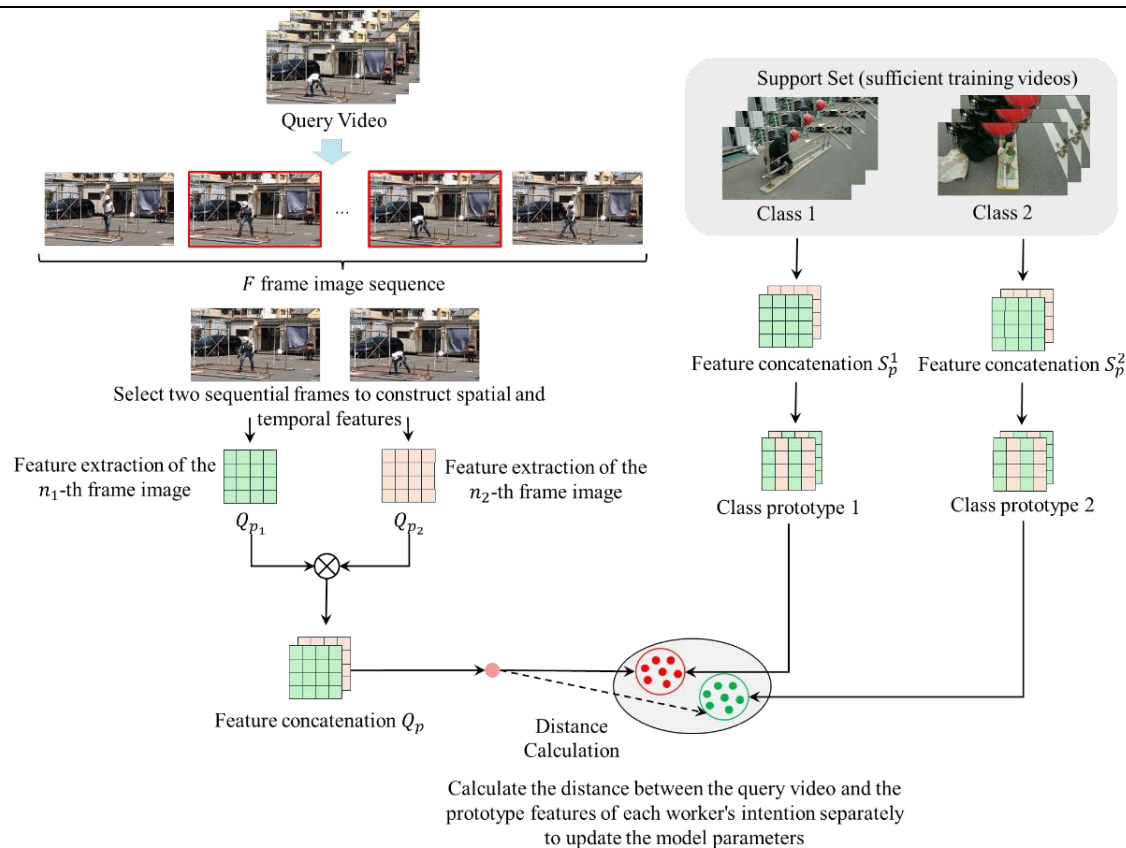


Fig. 2 The architecture of the proposed few-shot worker intention recognition model.

First, the video feature extraction module learns to extract generalized visual features that can represent various worker intentions. These generalized features, referred to as meta-features, delineate and elucidate the

overall characteristics of worker intention types within the dataset, as shown in Figure 2.

In this study, the module consists of two components. In the first step, convolutional neural networks are employed to extract features from individual frames of worker operations. For this single-frame image feature extraction, ResNet-34 is utilized for image feature computation.

In the second step, worker intention may exhibit different feature representations at different moments, making it challenging to capture complex features using only single frames. Furthermore, due to the complexities and dynamic nature of construction scenes, such as worker occlusion, worker intention features may not be fully captured by the robot. This leads to issues of discontinuity in the image sequences of worker intention and variations in action speed from the robot's perspective. To address the inter-frame sequence feature extraction, this study adopts a method of randomly combining two frames from the worker intention video to extract inter-frame sequence features, thereby mitigating the impact of discontinuous video sequences on the model.

Regarding the query video Q , we randomly slice the video frame sequence and set $p = (p_1, p_2)$. The video feature extraction calculation formula is as follows:

$$Q_p = [\Phi(q_{p_1}), \Phi(q_{p_2})] \quad (1)$$

where $\Phi()$ represents the convolutional neural network.

For the support video S_{km}^c , the feature representation is given by:

$$S_{km}^c = [\Phi(s_{km_1}^c), \Phi(s_{km_2}^c)] \quad (2)$$

where $m = (m_1, m_2)$.

3.2 Similarity Calculation

Similarity calculation is a crucial algorithm in few-shot models based on metric learning. This study aims to learn an appropriate similarity measure that enables better differentiation between worker intention types in the original task. Such methods do not rely on extensive parameter updates but instead make predictions through a fixed metric approach. They guide the Few-Shot model in learning how to (1) extract worker intention image features and inter-frame features from input videos using only labeled training data from the support set, and (2) compute their category prototypes when provided with some support images, subsequently determining the current worker intention type by calculating the distance between the query video and the category prototypes.

The metric learning-based few-shot worker intention recognition model plays a significant role in learning how to obtain an appropriate metric space from video frame images. Therefore, to enhance the model's feature generalization performance, during the training phase, the model utilizes sufficiently labeled training data to simulate few-shot detection scenarios. The model is then trained based on the provided few-shot data. In this study, the worker intention category prototype S_c is obtained using the temporal CrossTransformer computational method [35]. Let Q_i denote the query worker intention features. The distance for each query is taken negatively as the loss function for model parameter updates.

We utilize the Euclidean metric to calculate the distance between support features S_c and query features Q_i :

$$d = \|Q_i - S_c\|^2 \quad (3)$$

Based on the distance d , the model's loss function in the C-way K-shot task can be computed as:

$$\hat{d}_i^c = \frac{e^{-d_i^c}}{\sum_{c=1}^C e^{-d_i^c}} \quad (4)$$

During the training phase, we minimize the loss function to update the parameters of the proposed network, repeating this process across all randomly sampled tasks.

$$L = -\frac{1}{C} \sum_{i=1}^C \log(\hat{d}_i^c) \quad (5)$$

During the testing phase, we first need to configure different video categories, and a limited amount of data based on the input, which constitutes the k-way n-shot detection scenario. Using the few-shot model parameters obtained during the training phase, we derive the category prototypes for k worker intentions. The class with the minimum distance is then selected as the identified worker intention type. The worker intention recognition during the testing phase can be expressed as $\arg \min_c \hat{d}_i^c$. Unlike the training phase, new worker intention categories do not appear during training; they represent entirely new worker intention categories. In the testing phase, the worker intention category prototypes for the new classes are calculated solely using the trained model weights.

4. EXPERIMENTAL RESULTS AND ANALYSIS

To validate the proposed method, two distinct experiments were conducted: one involving few-shot learning and the other involving traditional supervised learning. Supervised learning is a conventional method for recognizing worker intent, utilizing numerous support images for model training, while few-shot learning focuses on acquiring general knowledge of worker intent by learning a feature prototype from the provided support video set. This study employed SLOWFAST [41] and X3D [42] architectures as baseline models for traditional

supervised learning. Comparing with the architectures of conventional supervised learning allows for a more comprehensive assessment of the impact of few-shot learning. Consequently, the authors can confirm the applicability and practicality of few-shot learning and objectively evaluate its beneficial effects.

For the experiments and result analysis, the research team utilized a dataset from the literature, which consists of scenes collected from YouTube, involving a worker operating on the same type of scaffolding with varying backgrounds and camera angles. In the authors' experiments, a total of 22 worker operation video data were randomly divided into training and testing datasets. Using this data, the research team trained and tested the model in different k-way n-shot scenarios, where the number of categories k was set to 2, 3, 4, 5, 10, and 20, and the number of support images n (given for each new category) was set to 1, 2, 3, 5, and 10. To minimize the loss function, stochastic gradient descent with a batch size of 8 was employed to train the few-shot worker intent recognition model, with a total of 30,000 iterations and a learning rate of 0.001. The performance of the trained model was evaluated using classification accuracy, which is one of the most widely used metrics for vision-based action classification networks.

4.1 Performance of the Proposed Method

Table 1 presents the experimental results based on different few-shot detection scenarios (k=2,3,4,5,10,20 and n=1,2,3,5,10). The developed model achieved a worker intent recognition accuracy of 71% in the 5-way 5-shot scenario. When 10 labeled videos of new worker intent classes were provided, the accuracy for these new classes reached 80.6%. Furthermore, as the amount of video data for worker intent recognition increased, the accuracy for recognizing new classes progressively improved.

Table 1: Experimental Results of the Proposed Method

	1-shot	2-shot	3-shot	5-shot	10-shot
2-way	71.8	77.9	81.2	87.8	91.3
3-way	58.8	69.2	72.9	80.9	86.6
4-way	53.2	62.1	67.8	75.5	82.5
5-way	47.7	58.1	64	71	80.6
10-way	36.2	45.7	51.8	60.7	71.3
20-way	28.5	37.8	44.8	53.2	64.6

4.2 Performance with and Without Few-Shot Learning

The authors also observed a significant positive impact of the proposed few-shot learning approach from the experiments. As shown in Figure 3, under the 10-way conditions, the recognition performance for new classes using few-shot learning consistently outperformed existing supervised learning methods (i.e., SLOWFAST and X3D). These results align with the established knowledge in computer vision and deep learning, which suggests that complex models cannot be effectively trained when training data is limited. The performance gap between few-shot learning and the best-performing supervised learning model (i.e., X3D) averaged approximately 35.0%, with these differences increasing as the values of k and n rise. This may be attributed to the requirement of supervised learning methods for a substantial amount of labeled training data across all classes, while learning new classes with limited video data poses challenges. In contrast, few-shot learning can leverage a few provided samples to "learn how to recognize new classes." These findings indicate that few-shot learning not only reduces the amount of training data necessary for generating deep learning models but also enables rapid learning for detecting new categories, even with minimal training data. Given these advantages, it may be applicable in real construction environments where numerous novel and foundational objects coexist.

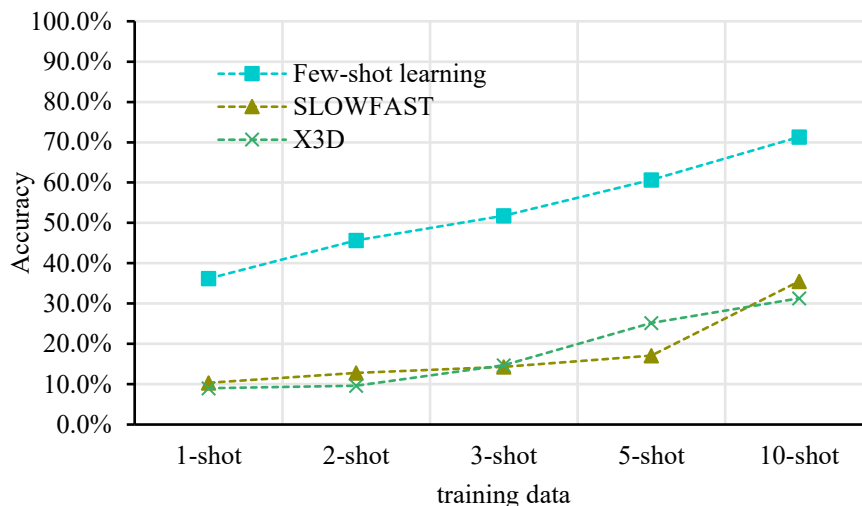


Fig. 3 Comparative analysis for the accuracy of few-shot and supervised learning

5. RESULTS AND DISCUSSION

The research findings indicate that the proposed method can learn meta-knowledge (i.e., how to learn target features) from a limited amount of labeled training data and utilize this knowledge to detect new worker intents when provided with some support videos. The model performed well despite variations in camera positions and viewpoints that presented different worker postures and visual features (e.g., color, shape, size).

Furthermore, this study shows that a promising deep learning model can be constructed using only a limited amount of training data. These findings could positively impact the implementation of vision-based monitoring in real-world construction sites. Specifically, the proposed few-shot learning method can quickly adapt to new human-robot collaboration needs arising at different construction stages, enabling the monitoring of new worker intentions without requiring extensive training data. This capability allows administrators to reduce the time and effort needed for additional data collection and annotation, which can be highly labor-intensive.

The technical advantages of this approach will be particularly useful during the phased development and updating of training databases. On-site managers can implement vision-based monitoring more effectively, as the development and updating of datasets represent one of the critical barriers to the use of visual systems in their projects. Additionally, the theoretical insights from this study can provide valuable guidance and future directions for researchers in the field of vision-based human-robot collaboration.

6. CONCLUSIONS

This study proposes a limited data-oriented worker intention recognition method in worker-robot collaboration for construction, aimed at minimizing the amount of training data and the cost of data labeling. The proposed method includes worker intention feature extraction and an intention classification strategy for limited data. As observed in the experiments, the few-shot model can learn meta-knowledge from a limited amount of labeled training data, and meta-learning successfully transfers this knowledge to detect new categories when provided with several video samples. In the 10-way 10-shot scenario, the accuracy for detecting new categories reached 71.3%, whereas the performance of supervised learning was limited to 35.5%. This indicates that the proposed method can train a more robust worker intention recognition model with the same amount of training data compared to existing methods.

The benefits of the proposed method contribute to the following advancements: First, this research introduces a novel technical framework that significantly reduces the number of required training images while maintaining performance in vision-based worker intention recognition. Second, the framework saves time and costs associated with data labeling, enhancing the practical acceptability of vision systems on construction sites. The authors customized few-shot learning for construction environments and provided quantitative results demonstrating its technical feasibility. Finally, the findings of this study lay the groundwork for future research in vision-based construction human-robot collaboration, as well as for other research areas, such as model-based construction automation monitoring.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China [grant number NSFC22EG19]; The Hong Kong University of Science and Technology 30 for 30 Research Fund [grant number 3030007]; the Research Grants Council (Hong Kong) [grant number 26208323]; University Grants Committee of

The Government of the Hong Kong Special Administrative Region Collaborative Research Fund [grant number C6044-23GF].

REFERENCES

- [1] Z. Pan, Y. Yu, Learning multi-granular worker intentions from incomplete visual observations for worker-robot collaboration in construction, *Autom Constr* 158 (2024). <https://doi.org/10.1016/j.autcon.2023.105184>.
- [2] S.K. Baduge, S. Thilakarathna, J.S. Perera, M. Arashpour, P. Sharafi, B. Teodosio, A. Shringi, P. Mendis, Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications, *Autom Constr* 141 (2022). <https://doi.org/10.1016/j.autcon.2022.104440>.
- [3] S. Eaves, D.E. Gyi, A.G.F. Gibb, Building healthy construction workers: Their views on health, wellbeing and better workplace design, *Appl Ergon* 54 (2016). <https://doi.org/10.1016/j.apergo.2015.11.004>.
- [4] S. Park, H. Yu, C.C. Menassa, V.R. Kamat, A Comprehensive Evaluation of Factors Influencing Acceptance of Robotic Assistants in Field Construction Work, *Journal of Management in Engineering* 39 (2023). <https://doi.org/10.1061/jmenea.meeng-5227>.
- [5] J. Cai, A. Du, X. Liang, S. Li, Prediction-Based Path Planning for Safe and Efficient Human–Robot Collaboration in Construction via Deep Reinforcement Learning, *Journal of Computing in Civil Engineering* 37 (2023). [https://doi.org/10.1061/\(asce\)jcp.1943-5487.0001056](https://doi.org/10.1061/(asce)jcp.1943-5487.0001056).
- [6] Y. Zhang, K. Ding, J. Hui, J. Lv, X. Zhou, P. Zheng, Human-object integrated assembly intention recognition for context-aware human-robot collaborative assembly, *Advanced Engineering Informatics* 54 (2022). <https://doi.org/10.1016/j.aei.2022.101792>.
- [7] S. Li, P. Zheng, J. Fan, L. Wang, Toward Proactive Human-Robot Collaborative Assembly: A Multimodal Transfer-Learning-Enabled Action Prediction Approach, *IEEE Transactions on Industrial Electronics* 69 (2022). <https://doi.org/10.1109/TIE.2021.3105977>.
- [8] Y. Tian, J. Chen, J.I. Kim, J. Kim, Lightweight deep learning framework for recognizing construction workers' activities based on simplified node combinations, *Autom Constr* 158 (2024). <https://doi.org/10.1016/j.autcon.2023.105236>.
- [9] J. Teizer, Status quo and open challenges in vision-based sensing and tracking of temporary resources on infrastructure construction sites, *Advanced Engineering Informatics* 29 (2015). <https://doi.org/10.1016/j.aei.2015.03.006>.
- [10] F. Semeraro, A. Griffiths, A. Cangelosi, Human–robot collaboration and machine learning: A systematic review of recent research, *Robot Comput Integr Manuf* 79 (2022). <https://doi.org/10.1016/j.rcim.2022.102432>.
- [11] Y. Pan, C. Chen, Z. Zhao, T. Hu, J. Zhang, Robot teaching system based on hand-robot contact state detection and motion intention recognition, *Robot Comput Integr Manuf* 81 (2023). <https://doi.org/10.1016/j.rcim.2022.102492>.
- [12] A. Cunha, F. Ferreira, E. Sousa, L. Louro, P. Vicente, S. Monteiro, W. Erhlagen, E. Bicho, Towards collaborative robots as intelligent co-workers in human-robot joint tasks: What to do and who does it?, in: 52nd International Symposium on Robotics, ISR 2020, 2020.
- [13] E.C. Grigore, A. Roncone, O. Mangin, B. Scassellati, Preference-Based Assistance Prediction for Human-Robot Collaboration Tasks, in: IEEE International Conference on Intelligent Robots and Systems, 2018. <https://doi.org/10.1109/IROS.2018.8593716>.
- [14] S. Choi, K. Lee, H.A. Park, S. Oh, A Nonparametric Motion Flow Model for Human Robot Cooperation, in: Proc IEEE Int Conf Robot Autom, 2018. <https://doi.org/10.1109/ICRA.2018.8463201>.
- [15] D. Shukla, O. Erkent, J. Piater, Learning semantics of gestural instructions for human-robot collaboration, *Front Neurorobot* 12 (2018). <https://doi.org/10.3389/fnbot.2018.00007>.
- [16] Y. Liu, M. Habibnezhad, H. Jebelli, Brainwave-driven human-robot collaboration in construction, *Autom Constr* 124 (2021). <https://doi.org/10.1016/j.autcon.2021.103556>.
- [17] H. Yu, V.R. Kamat, C.C. Menassa, W. McGee, Y. Guo, H. Lee, Mutual physical state-aware object handover in full-contact collaborative human-robot construction work, *Autom Constr* 150 (2023). <https://doi.org/10.1016/j.autcon.2023.104829>.
- [18] X. Wang, S. Wang, C.C. Menassa, V.R. Kamat, W. McGee, Automatic high-level motion sequencing methods for enabling multi-tasking construction robots, *Autom Constr* 155 (2023). <https://doi.org/10.1016/j.autcon.2023.105071>.

- [19] H. Wu, H. Li, H.L. Chi, Z. Peng, S. Chang, Y. Wu, Thermal image-based hand gesture recognition for worker-robot collaboration in the construction industry: A feasible study, *Advanced Engineering Informatics* 56 (2023). <https://doi.org/10.1016/j.aei.2023.101939>.
- [20] Z. Liu, Q. Liu, W. Xu, Z. Liu, Z. Zhou, J. Chen, Deep learning-based human motion prediction considering context awareness for human-robot collaboration in manufacturing, in: *Procedia CIRP*, 2019. <https://doi.org/10.1016/j.procir.2019.04.080>.
- [21] H. Liu, L. Wang, Human motion prediction for human-robot collaboration, *J Manuf Syst* 44 (2017) 287–294. <https://doi.org/https://doi.org/10.1016/j.jmsy.2017.04.009>.
- [22] W. Wang, R. Li, Y. Chen, Y. Sun, Y. Jia, Predicting Human Intentions in Human-Robot Hand-Over Tasks Through Multimodal Learning, *IEEE Transactions on Automation Science and Engineering* 19 (2022). <https://doi.org/10.1109/TASE.2021.3074873>.
- [23] J. Lyu, A. Maýe, M. Görner, P. Ruppel, A.K. Engel, J. Zhang, Coordinating human-robot collaboration by EEG-based human intention prediction and vigilance control, *Front Neurobot* 16 (2022). <https://doi.org/10.3389/fnbot.2022.1068274>.
- [24] Y. Song, T. Wang, P. Cai, S.K. Mondal, J.P. Sahoo, A Comprehensive Survey of Few-shot Learning: Evolution, Applications, Challenges, and Opportunities, *ACM Comput Surv* 55 (2023). <https://doi.org/10.1145/3582688>.
- [25] S.X. Hu, D. Li, J. Stuhmer, M. Kim, T.M. Hospedales, Pushing the Limits of Simple Pipelines for Few-Shot Learning: External Data and Fine-Tuning Make a Difference, in: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2022. <https://doi.org/10.1109/CVPR52688.2022.00886>.
- [26] H. Wang, H. Zhao, B. Li, Bridging Multi-Task Learning and Meta-Learning: Towards Efficient Training and Effective Adaptation, in: *Proc Mach Learn Res*, 2021.
- [27] L. Khacef, V. Gripon, B. Miramond, GPU-Based Self-Organizing Maps for Post-labeled Few-Shot Unsupervised Learning, in: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2020. https://doi.org/10.1007/978-3-030-63833-7_34.
- [28] X. Zhong, C. Gu, W. Huang, L. Li, S. Chen, C.W. Lin, Complementing representation deficiency in few-shot image classification: A meta-learning approach, in: *Proceedings - International Conference on Pattern Recognition*, 2020. <https://doi.org/10.1109/ICPR48806.2021.9412416>.
- [29] C. Finn, P. Abbeel, S. Levine, Model-agnostic meta-learning for fast adaptation of deep networks, in: *34th International Conference on Machine Learning, ICML 2017*, 2017.
- [30] M. Patacchiola, J. Turner, E.J. Crowley, M. O’Boyle, A. Storkey, Bayesian meta-learning for the few-shot setting via deep kernels, in: *Adv Neural Inf Process Syst*, 2020.
- [31] J. Xu, J.F. Ton, H. Kim, A.R. Kosiorek, Y.W. Teh, MetaFun: Meta-learning with iterative functional updates, in: *37th International Conference on Machine Learning, ICML 2020*, 2020.
- [32] E. van der Spoel, M.P. Rozing, J.J. Houwing-Duistermaat, P. Eline Slagboom, M. Beekman, A.J.M. de Craen, R.G.J. Westendorp, D. van Heemst, Siamese Neural Networks for One-Shot Image Recognition, *ICML - Deep Learning Workshop 7* (2015).
- [33] X. Li, X. Yang, Z. Ma, J.H. Xue, Deep metric learning for few-shot image classification: A Review of recent developments, *Pattern Recognit* 138 (2023). <https://doi.org/10.1016/j.patcog.2023.109381>.
- [34] K. Cao, J. Ji, Z. Cao, C.Y. Chang, J.C. Niebles, Few-shot video classification via temporal alignment, in: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2020. <https://doi.org/10.1109/CVPR42600.2020.01063>.
- [35] T. Perrett, A. Masullo, T. Burghardt, M. Mirmehdi, D. Damen, Temporal-Relational CrossTransformers for Few-Shot Action Recognition, in: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2021. <https://doi.org/10.1109/CVPR46437.2021.00054>.
- [36] Z. Cui, Q. Wang, J. Guo, N. Lu, Few-shot classification of façade defects based on extensible classifier and contrastive learning, *Autom Constr* 141 (2022). <https://doi.org/10.1016/j.autcon.2022.104381>.
- [37] Y. Xu, Y. Bao, Y. Zhang, H. Li, Attribute-based structural damage identification by few-shot meta learning with inter-class knowledge transfer, *Struct Health Monit* 20 (2021). <https://doi.org/10.1177/1475921720921135>.
- [38] J. Kim, S. Chi, A few-shot learning approach for database-free vision-based monitoring on construction sites, *Autom Constr* 124 (2021). <https://doi.org/10.1016/j.autcon.2021.103566>.
- [39] Y. Liang, P. Vadakkepat, D.K.H. Chua, S. Wang, Z. Li, S. Zhang, Recognizing temporary

- construction site objects using CLIP-based few-shot learning and multi-modal prototypes, *Autom Constr* 165 (2024) 105542. <https://doi.org/https://doi.org/10.1016/j.autcon.2024.105542>.
- [40] X. Wang, N. El-Gohary, Few-shot object detection and attribute recognition from construction site images for improved field compliance, *Autom Constr* 167 (2024) 105539. <https://doi.org/https://doi.org/10.1016/j.autcon.2024.105539>.
- [41] C. Feichtenhofer, H. Fan, J. Malik, K. He, Slowfast networks for video recognition, in: *Proceedings of the IEEE International Conference on Computer Vision*, 2019. <https://doi.org/10.1109/ICCV.2019.00630>.
- [42] C. Feichtenhofer, X3D: Expanding Architectures for Efficient Video Recognition, in: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2020. <https://doi.org/10.1109/CVPR42600.2020.00028>.

SIMULATION OF RESPONSE TIME REDUCTION OF OHCA CONSIDERING DRONE BASE SELECTION FOR AED DELIVERY

Yu Wu¹, Kuan-Chen Chin², Jen-Tang Sun³, Ying-Chih Ko⁴,
Wen-Chu Chiang⁵, Matthew Huei-Ming Ma⁶, and Albert Y. Chen⁷

- 1) Graduate Research Assistant, Department of Civil Engineering, National Taiwan University, Taipei City, Taiwan. Email: r12521503@ntu.edu.tw
- 2) MD. MPH., Department of Emergency Medicine, Taipei Hospital, Ministry of Health and Welfare, New Taipei City, Taiwan. Email: chinkuanchen@gmail.com
- 3) MD. MS., Department of Emergency Medicine, Far Eastern Memorial Hospital, New Taipei City, Taiwan. Email: tangtang05231980@hotmail.com
- 4) MD., Section of Emergency Medicine, Department of Medicine, National Taiwan University Cancer Center, Taipei City, Taiwan. Email: b99401027@ntu.edu.tw
- 5) MD. PhD., Department of Emergency Medicine, National Taiwan University Hospital, Yun-Lin Branch, Yunlin County, Taiwan. Email: drchiang.tw@gmail.com
- 6) MD. PhD., Department of Emergency Medicine, National Taiwan University Hospital, Yun-Lin Branch, Yunlin County, Taiwan. Email: mattma.tw@gmail.com
- 7) Prof., Department of Civil Engineering, National Taiwan University, Taipei City, Taiwan. Email: albertchen@ntu.edu.tw

Abstract: Out-of-Hospital Cardiac Arrest (OHCA) is a critical medical emergency that requires immediate intervention. With technological advancements, integrating drones into the Emergency Medical Services (EMS) has emerged as a promising solution. This study examines OHCA cases in Taipei and New Taipei City, Taiwan, from September to December 2019. A two-phase optimization model was employed to identify the effective locations for drone bases. Among the scenarios tested, settings with a maximum flight time of 7 minutes, 90% OHCA case coverage, and 13 drone bases, demonstrated the best overall performance in terms of system-wide and area-specific usage rates, response time reduction, and cost-effectiveness. These findings provide valuable insights into the feasibility of drone deployment and offer practical reference for real-world emergency applications.

Keywords: OHCA, AED, EMS, Drones, Response time

1. INTRODUCTION

Out-of-Hospital Cardiac Arrest (OHCA) is a leading cause of global mortality (Myat et al., 2018), and the most important predictive factor for survival is response time (O’Keeffe et al., 2011). Rapid intervention is essential to the survival of OHCA patients.

As technological advancements progress and labor resources diminish, integrating technology into emergency medical response systems may offer significant advantages. During rush hours, there could be delays for ambulance arrival, while rural areas could be harder to reach. To address this, the use of drones as a supplement to traditional Emergency Medical Services (EMS) may provide a viable solution for improving response times and patient outcomes.

This study focuses on OHCA cases in Taipei and New Taipei City, Taiwan, from September to December 2019. It investigates the integration of a drone system into existing EMS. Spatial analysis will first be conducted to determine efficient drone base locations, followed by simulations of the integrated system. The performance of the combined drone-EMS system will be compared with that of the conventional EMS system without drone support, with service effectiveness analyzed across different area types.

The primary outcome of this study will be the difference in Automated External Defibrillator (AED) response time after integrating the drone system. Additionally, the study will assess the usage rate and response time differences across various area types, including urban, suburban, and rural regions. The findings of this research can be used to evaluate the feasibility of drone base deployment and its practical application in real-world emergency scenarios.

2. LITERATURE REVIEW

In this part, we will first explore the feasibility of drone applications, followed by an analysis of the methodologies for drone base selection. Finally, we will identify the research gaps that this study aims to address.

2.1 Drone Application in AED delivery

Previous work suggests that deploying drones for AED delivery in rural areas during OHCA events is both safe and feasible (Claesson et al., 2016a). The feasibility of AED delivery by drone during nighttime was also evaluated, showing no significant difference in performance between day and night operations (Scholz et al., 2023). Through a geographical model, assessment of the potential usage of drone systems in response time reduction for AEDs has been conducted. Numerical simulations in high-density urban areas indicated that 26% of OHCA cases in

Greater Paris could benefit from drone-delivered AEDs before the arrival of basic life support teams (Derkenne et al., 2021).

2.2 Optimal Drone Base Selection

The selection of drone bases should account for ambulance response time, incidences of OHCA cases, distance, and density of incidents. In their work, different weights were assigned to urban and rural areas, and a Geographic Information System (GIS) combined with the Multi-Criteria Evaluation (MCE) method were utilized to facilitate base selection (Claesson et al., 2016b). In a related approach, a mathematical optimization model was used to minimize the number of drone bases while ensuring that at least $f\%$ of OHCA cases were covered. The model iteratively adjusted the coverage rate (f) and the maximum flight time (t) to achieve targeted reductions in response time (Boutilier et al., 2017). Another two-stage optimization model was employed for base selection: Phase 1 maximized the number of OHCA cases reached within 5 minutes, and Phase 2 minimized the mean response time while maintaining the same coverage level (Starks et al., 2024).

2.3 Research Gap

Current studies primarily focus on single-city implementations, limiting the potential benefits of cross-regional integration. Furthermore, research that combines spatial analysis with simulation is relatively scarce. This study will address these gaps by applying spatial analysis to the Greater Taipei Area, covering both Taipei and New Taipei's OHCA cases, to identify efficient drone base locations using an optimization model. Simulation analysis will then be conducted to evaluate response time reductions. The results will provide insights into how different base configurations affect response times across various geographical settings.

3. METHOD

In the methodology section, we will first describe the role of drones within the EMS system, followed by detailed assumptions related to drone operations. Next, we will outline the data preprocessing steps and the optimization model employed for selecting optimal drone base locations.

3.1 Overview of Approach

After OHCA occurs, the dispatch center will receive the report information and decide whether to dispatch a drone for the OHCA. For OHCA cases, it is essential that there are at least two people present to assist the victim, as one needs to perform CPR while the other prepares and uses the AED. Thus, if there is no bystander performing CPR, the drone will not be dispatched.

To validate the influence of integrating drones into the existing EMS, a simulation model will be used. The historical EMS response time, based on real-world data, will be compared to the drone simulation response time, which represents the time taken for a drone to transport an AED. The shorter of the two times will be selected as the system (EMS + Drone) response time.

3.2 Assumption

The flight distance of the drone will be calculated using the linear distance between the selected drone base and the OHCA location. Considering the practical limitations, it takes time to connect from the EMS center to dispatch the drone, and door-to-door drone delivery is not feasible in urban areas. Therefore, in the simulation, we need to account for the dispatch connection time T_c and AED retrieval time at different accident locations i , denoted as $T_{a,i}$, in addition to the drone flight time. For high-rise buildings, higher floors require more time; thus, for every floor added, the time increases by T_f .

For the performance of the drone itself, we referred to the drone type 'Ceptor' from GlobeUAV, Germany, which has been used in both emergency delivery simulations and real-life situations in the past (Baumgarten et al., 2022; Scholz et al., 2023). Consequently, several time related parameters (in seconds) are set:

Table 1. Parameters

parameters	value
T_c (s)	60
$T_{a,i}$ (s)	30, $i \in \{\text{workplace, street, nursing home, public building, school}\}$ 60, $i \in \{\text{residential area, unclear}\}$
T_f (s)	10
Ascend acceleration (m/s^2)	1.0
Descend acceleration (m/s^2)	1.0
Horizontal acceleration (m/s^2)	19.8
Horizontal speed (km/h)	50
Cruise height (m)	100

3.3 Data Preprocess

To conduct the simulation, it is essential to preprocess the data by discarding entries with suspected misclassified or missing values. Additionally, certain geographical data is indispensable; however, some key information, such as latitude and longitude, may be absent from the raw dataset despite the presence of addresses. In such cases, geocoding will be performed using addresses and an API to obtain the corresponding latitude and longitude coordinates. Furthermore, the classification of areas into urban, suburban, and rural zones will be completed to support the analysis.

3.4 Optimization Model for Drone Base Selection

Given the practical constraints and associated costs, establishing drone stations across the entire city is not cost effective. Therefore, we decided to use a limited number of drones and identify optimal station locations to restrict cost expenditures. It is essential to determine both the number of drones to be deployed and their precise placement. This selection process will be carried out using a two-phase optimization model, using Python with the Gurobi solver employed to ensure efficient problem-solving.

(1) Phase 1: Optimizing the Number of Drone Station

To ensure that the drone service can reach OHCA cases within a specified time limit while maintaining sufficient coverage across the city, the model explores various combinations of maximum flight time t and OHCA case coverage rate f . This approach involves performing a grid search to identify the most feasible and optimal number of drone bases. In each combination of t and f , the following parameters are adjusted and pre-processed, and the model should be solved.

a. Set and Index

- \mathbb{I} : the set of drone base candidates, index by i
- \mathbb{J} : the set of OHCA cases, index by j

b. Parameters

- f : Indicates the percentage of covered OHCA.
- t : Indicates the maximum drone flying time.
- R : Defines the coverage radius
- a_{ij} : Binary parameters indicating whether OHCA case j is covered by drone base candidate i .
- I : The number of fire stations, representing the number of drone base candidates.
- J : The number of OHCA cases in the training set.

c. Decision variables

- z_{ij} : Binary variable indicating whether OHCA case j is covered by drone base i .
- y_i : Binary variable indicating whether a drone base is established at candidate location i .

d. Model

$$\text{Minimize} \quad \sum_{i=1}^I y_i \quad (1)$$

$$\text{Subject to} \quad \sum_{i=1}^I z_{ij} \leq 1 \quad \forall j \in \mathbb{J} \quad (2)$$

$$\sum_{i=1}^I \sum_{j=1}^J z_{ij} \geq \frac{f}{100} \times J \quad (3)$$

$$z_{ij} \leq a_{ij} y_i \quad \forall i \in \mathbb{I}, \forall j \in \mathbb{J} \quad (4)$$

$$y_i, z_{ij} \in \{0,1\} \quad \forall i \in \mathbb{I}, \forall j \in \mathbb{J} \quad (5)$$

(2) Phase 2: Optimizing the Location of Drone Station

Focusing solely on the coverage rate may result in multiple feasible solutions in some cases. Therefore, phase two considers both coverage and distance costs to provide a balanced solution that ensures adequate coverage while minimizing the total travel distance to reach OHCA cases.

a. Set and Index

- \mathbb{I} : the set of drone base candidates, index by i
- \mathbb{J} : the set of OHCA cases, index by j

b. Parameters

- R : The coverage radius of drone station.
- a_{ij} : Binary parameters indicating whether OHCA case j is covered by drone base candidate i .
- d_{ij} : The linear distance between drone base candidate i to OHCA case j .
- n : number of base selected
- α : the weight of coverage. $0 \leq \alpha \leq 1.0$.

c. Decision variables

- y_i : Binary variable indicating whether drone base candidate i is selected/established.
- x_j : Binary variable indicating whether OHCA case j is covered by at least one drone base's service area.
- z_{ij} : Binary variable indicating whether OHCA case j is covered by drone base i .

d. Model

$$\text{Maximize} \quad \alpha \sum_{j \in \mathbb{J}} x_j - (1 - \alpha) \sum_{i \in \mathbb{I}, j \in \mathbb{J}} d_{ij} z_{ij} \quad (6)$$

$$\text{Subject to} \quad a_{ij} y_i \geq z_{ij} \quad \forall i \in \mathbb{I}, j \in \mathbb{J} \quad (7)$$

$$\sum_{i \in \mathbb{I}} z_{ij} \geq x_j \quad \forall j \in \mathbb{J} \quad (8)$$

$$\sum_{i \in \mathbb{I}} y_i \leq n \quad (9)$$

$$x_i, y_j, z_{ij} \in \{0,1\} \quad \forall i \in \mathbb{I}, j \in \mathbb{J} \quad (10)$$

4. RESULTS

The Results section will be divided into three parts: the data preprocessing; the outcomes of drone station selection, including a sensitivity analysis of multiple combinations of f and t as well as the top eight configurations for the number and locations of drone bases; and the performance evaluation of the drone-enhanced system across different area types and the overall system.

4.1 Data

The data used in this study comprise OHCA cases from Taipei, New Taipei City, and Taiwan, collected between September and December 2019. Before conducting the simulations, several data preprocessing steps were completed, as outlined in Figure 1. The area classification method proposed by Claesson et al. (2016) was adopted, where urban areas are defined as having $\geq 6,000$ inhabitants/km², and rural areas as having < 250 inhabitants/km². Population data from the Taiwan's Ministry of the Interior's Socioeconomic Data Platform were combined with these criteria to categorize the areas as urban, suburban, or rural. The final dataset included 2,724 urban cases, 481 suburban cases, and 62 rural cases, totaling 3,267 OHCA incidents. The average AED response time for Taipei City is 357.76 seconds, for New Taipei City 339.17 seconds, and for both cities combined, 345.62 seconds.

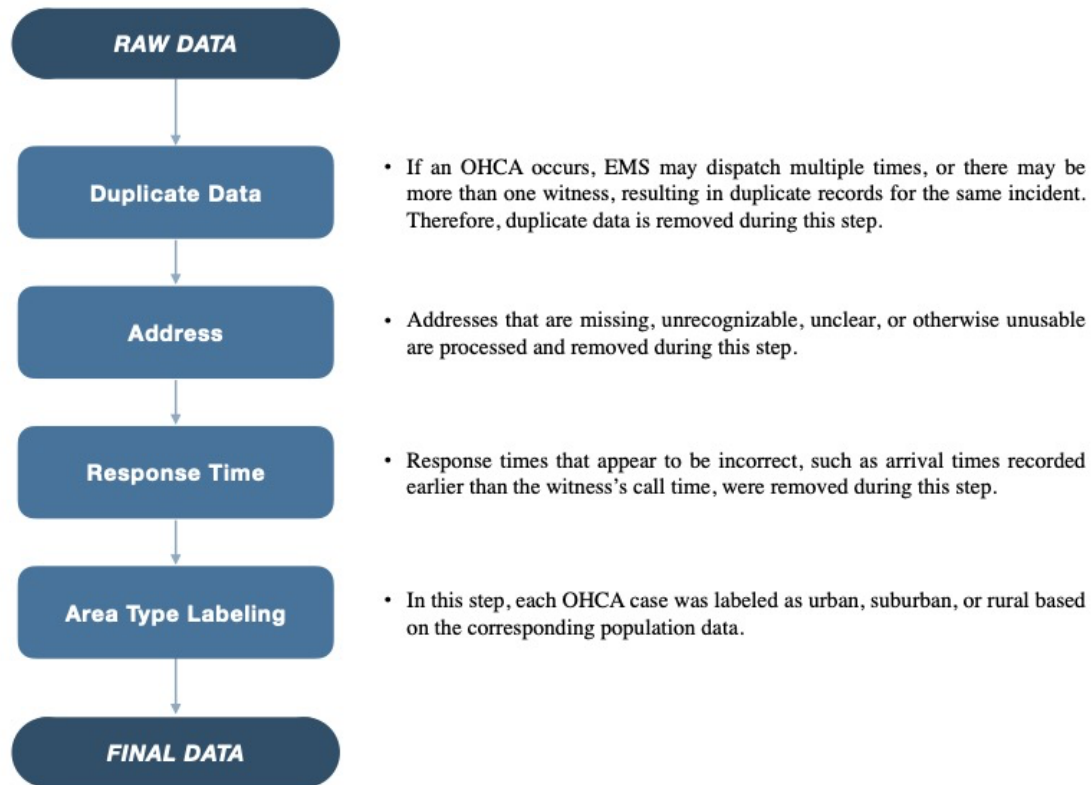


Figure 1. Data Preprocessing

4.2 Drone Bases

To select drone base locations that optimize both rescue efficiency and cost-effectiveness, various combinations of maximum flight time t and coverage rate f were evaluated using the Phase 1 model. The maximum flight time of the Ceptor drone is 60 minutes. Therefore, we set the parameter t to 30 minutes and down to 5 minutes. Since a higher coverage rate is always preferred, we evaluated the coverage rate f from 100% to 50% and did not continue further below 50%. The output N , which represents the minimum number of necessary drones (equivalent to the number of drone bases, as each base supports one drone), is presented in Table 2.

Table 2. Sensitivity Analysis of Different (t, f)

Test	t	f	N	Test	t	f	N	Test	t	f	N	Test	t	f	N	Test	t	f	N
1	30	100%	3	8	15	100%	-	15	10	100%	-	22	7	100%	-	29	5	100%	-
2	30	95%	1	9	15	95%	3	16	10	95%	6	23	7	95%	18	30	5	95%	-
3	30	90%	1	10	15	90%	2	17	10	90%	5	24	7	90%	13	31	5	90%	-
4	30	80%	1	11	15	80%	1	18	10	80%	3	25	7	80%	8	32	5	80%	41
5	30	70%	1	12	15	70%	1	19	10	70%	2	26	7	70%	6	33	5	70%	26
6	30	60%	1	13	15	60%	1	20	10	60%	2	27	7	60%	4	34	5	60%	18
7	30	50%	1	14	15	50%	1	21	10	50%	1	28	7	50%	3	35	5	50%	13

The options that provide sufficient coverage with a limited number of drone bases are highlighted in blue. Eight combinations of t and f were used as parameters for the Phase 2 optimization. The corresponding geographical plot, which illustrates the OHCA density distribution, the selected drone bases, and their coverage areas, is presented in Figure 2.

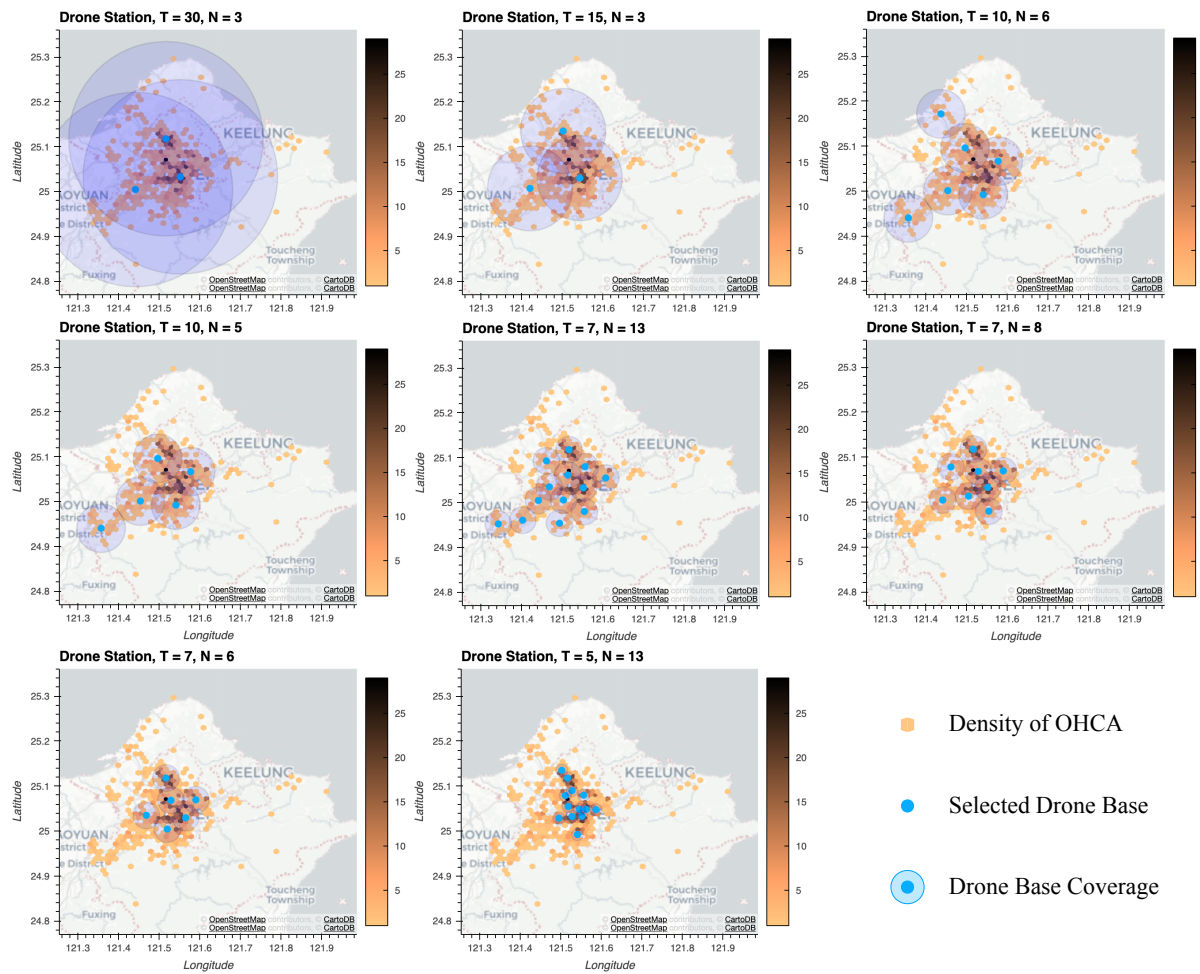


Figure 2. OHCA Density Distribution with Selected Drone Bases and Coverage Areas

4.3 System Performance

The primary outputs of the simulation will be the response time differences and the usage rates of the drones. The secondary outcomes include the response time differences and usage rates across different area types. The indices referenced in the analysis of system and area type comparisons are detailed in Table 3 and Table 4, respectively. The system performance outputs are presented in Table 5, while a comparison of different areas is provided in Table 6. The top three performances for each index will be highlighted in red.

Table 3. System Performance Index Description

Index	Description
System U.R.	Total system drone utilization rate.
RT > 5min U.R.	Drone utilization rate for OHCA cases with response times longer than 5 minutes.
# Drone served	Number of OHCA cases served by drones.
System RT Diff	The difference between historical EMS response time and simulated system response time for the total system.
Drone served RT Diff	The difference between historical EMS response time and simulated system response time for OHCA cases utilizing drones.

Table 4. Area Type Comparison Index Description

Index	Description
Urban U.R.	Urban OHCA drone utilization rate
Urban RT Diff	The difference between historical EMS response time and simulated response time for urban OHCA cases utilizing drones.
Suburban U.R.	Suburban OHCA drone utilization rate
Suburban RT Diff	The difference between historical EMS response time and simulated response time for suburban OHCA cases utilizing drones.
Rural U.R.	Rural OHCA drone utilization rate
Rural RT Diff	The difference between historical EMS response time and simulated response time for rural OHCA cases utilizing drones.
Rural-Urban Diff	The simulated response time difference between rural and urban regions.

Table 5. System with Drone's Performance

Test	t	f	N	System U.R.	RT > 5min U.R.	# Drone served	System RT Diff	Drone served RT Diff
1'	30	100%	3	9.09%	16.94%	222	22.235	245.770
9'	15	95%	3	8.59%	16.00%	205	21.023	246.044
16'	10	95%	6	11.35%	21.16%	271	29.609	261.741
17'	10	90%	5	10.64%	19.83%	254	28.511	268.945
24'	7	90%	13	18.77%	34.97%	448	53.397	285.078
25'	7	80%	8	15.25%	28.42%	364	39.760	261.439
26'	7	70%	6	14.08%	26.23%	336	35.733	254.618
35'	5	50%	13	20.78%	38.72%	496	53.042	255.782

Table 6. Area Type Comparison

Test	t	f	N	Urban U. R.	Urban RT Diff	Suburban U. R.	Suburban RT Diff	Rural U. R.	Rural RT Diff	Rural-Urban Diff
1'	30	100%	3	8.85%	223.633	10.76%	372.186	10.53%	322.906	99.273
9'	15	95%	3	8.41%	226.716	10.07%	353.051	5.26%	525.370	298.654
16'	10	95%	6	10.29%	240.594	19.10%	306.018	10.53%	1306.764	1066.170
17'	10	90%	5	9.86%	246.661	16.32%	321.982	10.53%	1306.764	1060.103
24'	7	90%	13	17.74%	262.941	25.35%	371.484	31.58%	595.254	332.313
25'	7	80%	8	15.00%	248.770	17.01%	342.993	15.79%	246.955	-1.816
26'	7	70%	6	13.94%	238.853	15.28%	354.550	10.53%	341.979	103.126
35'	5	50%	13	20.91%	242.070	20.49%	353.231	10.53%	363.445	121.375

5. DISCUSSION

In the drone base location selection process, it can be observed that as the maximum flight time t decreases, the number of required drone bases increases, as shown in Table 2. Due to population density, which is strongly correlated with OHCA density, a smaller t ensures shorter response times but causes the drones to cluster in the central urban areas, making it difficult to cover all OHCA cases. However, dense areas usually have more ambulance bases. As a result, the drone bases are not that beneficial to the system if placed in dense urban areas.

According to the simulation results, Test 24' in Table 5 and Table 6, with a maximum flight time of 7 minutes, 90% OHCA case coverage, and 13 drone bases, achieved the best overall performance. In this scenario, the drones could serve 448 cases, representing 18.77% of all OHCA cases and 34.97% of cases with a historical response time longer than 5 minutes. The system response time difference is 53.397 seconds, indicating that the addition of drones can reduce the response time by an average of 53.397 seconds per case. For the cases served by drones, the response time was reduced by 285.078 seconds per case.

Focusing on the area-type comparison, Test 24' achieved the best performance across urban, suburban, and rural areas, serving 31.58% of rural OHCA cases and reducing the response time by an average of 595.254 seconds for these cases. Since most fire stations are located in relatively urban areas, if the primary goal is to improve efficiency in rural areas with minimal cost, Test 17' would be a good choice, as it can reduce the rural response time by up to 1306.764 seconds per case. On the other hand, if the objective is to maximize the number of cases covered regardless of location, Test 35' would be a viable option.

6. CONCLUSION AND FUTURE WORK

This work has demonstrated the assessment of adding drone into an EMS for AED delivery in a cross city geospatial area. Through the analysis of data collected from Taipei City and New Taipei City, the drones improve OHCA response time. The results demonstrate that the maximum flight time and coverage rate influence the selection of drone base locations, which in turn impacts the response time difference. Among the tested scenarios, Test 24 provides the most balanced performance both system-wide and across all area types, improving rural OHCA coverage while maintaining a reasonable number of drone bases.

The data used in this research covers only three months. Using a larger dataset or data from other cities could provide more robust insights. Additionally, in urban areas, the presence of tall buildings may affect the flight paths of drones, making it essential to account for these factors to better reflect real-world conditions. Furthermore, the selection of drone base locations should not be limited to fire stations; instead, potential locations could be optimized based on the distribution of OHCA cases.

ACKNOWLEDGMENTS

We sincerely thank the Taipei and New Taipei City Fire Department for their invaluable assistance in collecting the data essential to this study, which formed the foundation of our research. We are also grateful to the support from the National Science and Technology Council of Taiwan for grant NSTC 113-2628-E-002-026-MY3, which made this study possible. This support was crucial to the success of our work, and we deeply appreciate it.

REFERENCES

- Baumgarten, M. C., Röper, J., Hahnenkamp, K., & Thies, K.-C. (2022). Drones delivering automated external defibrillators—Integrating unmanned aerial systems into the chain of survival: A simulation study in rural Germany. *Resuscitation*, 172, 139–145.
- Boutillier, J. J., Brooks, S. C., Janmohamed, A., Byers, A., Buick, J. E., Zhan, C., Schoellig, A. P., Cheskes, S., Morrison, L. J., & Chan, T. C. Y. (2017). Optimizing a drone network to deliver automated external defibrillators. *Circulation*, 135(25), 2454–2465.
- Claesson, A., Fredman, D., Svensson, L., Ringh, M., Hollenberg, J., Nordberg, P., Rosenqvist, M., Djarv, T., Österberg, S., Lennartsson, J., & others. (2016a). Unmanned aerial vehicles (drones) in out-of-hospital-cardiac-arrest. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, 24, 1–9.
- Derkenne, C., Jost, D., De L'Espinay, A. M., Corpet, P., Frattini, B., Hong, V., Lemoine, F., Jouffroy, R., Roquet, F., Marijon, E., & others. (2021). Automatic external defibrillator provided by unmanned aerial vehicle (drone) in Greater Paris: a real world-based simulation. *Resuscitation*, 162, 259–265.
- Myat, A., Song, K.-J., & Rea, T. (2018). Out-of-hospital cardiac arrest: current concepts. *The Lancet*, 391(10124), 970–979.
- O'Keeffe, C., Nicholl, J., Turner, J., & Goodacre, S. (2011). Role of ambulance response times in the survival of patients with out-of-hospital cardiac arrest. *Emergency Medicine Journal*, 28(8), 703–706.
- Scholz, S. S., Wähnert, D., Jansen, G., Sauzet, O., Latka, E., Rehberg, S., & Thies, K.-C. (2023). AED delivery at night—Can drones do the Job? A feasibility study of unmanned aerial systems to transport automated external defibrillators during night-time. *Resuscitation*, 185, 109734.
- Starks, M. A., Chu, J., Leung, K. H. B., Blewer, A. L., Simmons, D., Hansen, C. M., Joiner, A., Cabañas, J. G., Harmody, M. R., Nelson, R. D., & others. (2024). Combinations of first responder and drone delivery to achieve 5-minute AED deployment in OHCA. *JACC: Advances*, 3(7_Part_2), 101033.

Evaluating the Impact of Problem-Based Learning and Digital Models on Civil Engineering Education

Hamidreza Alavi¹, Brian Sheil² and Ioannis Brilakis³

1) Ph.D., Research Assoc. Department of Engineering, University of Cambridge, UK. Email: sa2194@cam.ac.uk

2) Ph.D., Assoc Prof., Department of Engineering, University of Cambridge, UK. Email: bbs24@cam.ac.uk

3) Ph.D., Prof., Department of Engineering, University of Cambridge, UK. Email: ib340@cam.ac.uk

Abstract: To address the increasing complexity and sustainability challenges facing civil engineering, an educational approach that integrates theory with practical experience is needed. Traditional methods of teaching often fail to equip students with the skills needed by the industry which later on hampers their job readiness and the development of infrastructure. This study investigates the use of Problem-Based Learning as an integrative feature of BIM in a civil engineering module for the purpose of increasing engagement, problem-solving skills, and comprehension of complex concepts. The research included both pre- and post-class surveys that were directed toward the students and the professor. The pre-class surveys were used for the examination of the anticipated expectations, perceived difficulties, and awareness of PBL and BIM, while the post-class surveys were designed for the estimation of the shift in perceptions, confidence levels, and contentedness. The gathered data were also examined through tools like descriptive statistics and thematic coding. The two approaches, PBL and BIM, were the most positively rated, where instead of discussing the effectiveness of these methods, 62.5% of students mentioned that BIM was the most beneficial tool for deepening structural concepts. The instructor particularly made mention of successful alliances with industry and sustainable direction but pointed to delays in permits and logistics maintenance as big threats. The integration of PBL and BIM in the civil engineering curriculum is found to be a significant aim of improved teaching as students are more motivated to complete course objectives and are more confident when they are exposed to real-life problems of the industry. It is proposed that PBL sessions be held in advance and that new digital resources be designed. Further research should be conducted to find out the effect of such methods on career readiness in the long term.

Keywords: Civil Engineering Education, Problem-Based Learning (PBL), Building Information Modeling (BIM), Student Engagement, Curriculum Development, Professional Practice, Sustainability.

1. INTRODUCTION

Civil engineering education faces unique challenges, as students must acquire a blend of theoretical knowledge and practical skills to succeed in real-world applications. Traditionally, lecture-based approaches have formed the core of engineering programs, focusing on conceptual knowledge and formulaic problem-solving. However, these methods often fall short of equipping students for the complex, unpredictable conditions of actual engineering projects (Das et al., 2023). In recent years, the field has witnessed increasing demand for an education that not only covers foundational knowledge but also fosters the ability to navigate real-world challenges such as sustainability, project constraints, and collaboration with various stakeholders.

Addressing this gap, Problem-Based Learning (PBL) has emerged as a transformative approach, gaining traction in civil engineering programs worldwide. PBL situates students in the midst of complex, real-world problems that require interdisciplinary thinking and collaborative effort (Azam et al., 2024). Rather than presenting a fixed set of facts, PBL engages students in problem-solving, where they must analyze situations, explore potential solutions, and evaluate outcomes. This approach not only deepens technical understanding but also cultivates critical thinking, adaptability, and teamwork—skills essential to civil engineering practice. Recent studies have shown that students in PBL-based courses demonstrate increased engagement, improved problem-solving abilities, and a greater capacity for handling multifaceted engineering tasks (Čubela et al., 2023; Farooq & Riaz, 2024).

Building Information Modeling (BIM) has become a cornerstone in the integration of digital models within civil engineering education, enhancing the learning experience by offering interactive, visual representations of structural elements. BIM's dynamic, multi-dimensional modeling allows students to visualize and simulate engineering processes, transforming abstract concepts into concrete experiences (Zou et al., 2021; Alavi, 2022). By interacting with digital models, students gain insights into the structural and environmental implications of their decisions. Studies have found that combining PBL with BIM fosters a learning environment that closely mirrors professional practice, offering students a hands-on, realistic experience (Ullmann et al., 2023). The addition of digital tools in PBL not only reinforces technical knowledge but also prepares students to work with the technologies and methodologies they will encounter in their careers.

This study evaluates the effectiveness of integrating PBL and BIM in a civil engineering module to enhance student engagement, confidence, and understanding of complex engineering concepts. Pre-class and post-class surveys were administered to gauge student expectations, perceived challenges, familiarity with PBL and BIM, and the impact of these methods on learning outcomes. The professor's perspective was also gathered,

providing insights into the instructional and logistical challenges involved in delivering a module centered on PBL and digital models. Analyzing this data provides a comprehensive view of how integrating PBL and BIM can impact civil engineering education, bridging the gap between theoretical knowledge and practical application, and fostering skills relevant to today's engineering challenges.

2. METHOD

This study aimed to understand how using PBL and BIM in a civil engineering module impacted students' learning experiences. To do this, we collected feedback from both the students and the professor involved in the course, using a straightforward pre- and post-module survey process for the students, and a mix of surveys and interviews with the professor. This approach allowed us to capture a range of perspectives on the module's design, delivery, and effectiveness.

2.1 Survey and Interview Approach

For the students, we created an accessible survey using Google Forms, making it easy for them to participate. To simplify the process, a QR code linked directly to the survey form was provided before the course began. This meant students could scan the code and complete the survey on their phones, right before the first class, giving us a snapshot of their expectations, initial confidence, and thoughts on potential challenges. After the module ended, we used the same approach, sharing the QR code again so they could provide follow-up feedback. Using the same questions in both surveys let us track any changes in their perceptions over the course.

To gather the professor's insights, we took a slightly different approach. Besides completing a structured Google Form similar to the students', the professor also participated in informal interviews both before and after the module. These conversations allowed for a deeper dive into their experience delivering the course, discussing everything from anticipated challenges to observations about student engagement. In the pre-course interview, the professor shared their teaching goals and thoughts on potential logistical issues, while the post-course interview provided a chance to reflect on what worked well and what could be improved. These interviews added a layer of insight beyond what a standard survey could capture.

2.2 Participants

The students in this study were all undergraduates enrolled in the part of the Civil Engineering program at the University of Cambridge. This group came from varied academic backgrounds but was united in their focus on sustainable construction—a key theme of the course. Including the professor's perspective as both an instructor and observer allowed us to understand the teaching dynamics and challenges involved in introducing PBL and BIM in a foundational civil engineering module. This combined view of student and instructor experiences helped to capture a fuller picture of the module's impact.

2.3 Data Collection and Analysis

We gathered all survey responses through Google Forms, which helped keep everything organized. Thanks to the ease of access provided by the QR code, most students completed both surveys, giving us solid pre- and post-module data to work with. For the quantitative parts, we used descriptive statistics to analyze shifts in student confidence, perceived difficulty, and effectiveness ratings of PBL and BIM. This gave us a structured view of changes across the group.

For the open-ended survey responses and the professor's interviews, we used thematic analysis to pull out recurring themes. Student responses highlighted aspects like "building confidence," "real-world application," and "collaborative learning." The professor's interview responses were coded separately, bringing up themes such as "logistical hurdles," "effective teaching strategies," and "observed student growth." By combining these different forms of data, we were able to gain a clear understanding of how students and the professor experienced the course, with both structured statistics and in-depth reflections.

3. RESULTS

The results of this study present a comparative analysis of student expectations and outcomes before and after completing the module, alongside the professor's perspective on its implementation. This section summarizes the key findings in three areas: initial expectations and perceived challenges, effectiveness of the PBL and BIM methods, and shifts in confidence and learning engagement.

3.1 Initial Expectations and Perceived Challenges (Pre-Class Survey)

Before the module began, students expressed a mixture of enthusiasm and apprehension regarding the course's anticipated challenges. According to the pre-class survey, 57.1% of students expected a moderate to high level of difficulty, with particular concerns about complex topics such as environmental risk management and structural analysis. These areas were seen as potentially challenging, with students highlighting the need for additional support in applying these concepts practically. However, students demonstrated a moderate level of confidence in their ability to tackle these challenges, with an average confidence score of 4.29 out of 5, indicating a belief in their potential to succeed despite anticipated difficulties.

The pre-class survey also captured students' perceptions of which subjects were most crucial for their success in the module. Structural analysis, dynamics, and mathematics emerged as key subjects, with high

importance ratings indicating students' recognition of their foundational role in solving engineering problems. In contrast, subjects such as cost planning and health and safety were rated as less critical, suggesting these areas may not engage students as strongly as the core technical subjects. This prioritization indicates potential areas where instructional strategies could be tailored to increase student interest and understanding in less popular topics.

The professor's pre-class survey responses underscored the module's objective of providing students with practical, real-world experience through industry collaborations and sustainability topics. The professor rated the clarity of these objectives at 5/5, emphasizing that an essential component was engaging students through authentic industry scenarios, including introductions to industry practices and collaborative group projects. Sustainability was also highlighted as a focus area, aligning the module with broader societal needs and academic research goals.

In terms of instructional challenges, the professor anticipated high-risk areas in logistical issues, permitting delays, and project management (all rated 5/5), reflecting a realistic outlook on potential obstacles. Although the professor expressed moderate confidence (3/5) in addressing these challenges fully, they felt confident (4/5) in guiding students through them, underscoring a commitment to student support despite external constraints.

3.2 Effectiveness of the Module (Post-Class Survey)

The post-class survey revealed a notable shift in student perceptions regarding the module's difficulty. Initially, many students anticipated significant challenges; however, by the end of the module, 50% rated the difficulty level as moderate. This shift suggests that the module's structured support and PBL approach effectively mitigated students' initial concerns, allowing them to navigate complex topics with greater ease.

Student evaluations of PBL and BIM were particularly positive, with a half of students indicating that PBL significantly enhanced their problem-solving abilities and 62.5% reporting that BIM was highly effective in visualizing and understanding structural concepts. BIM's interactive capabilities enabled students to explore real-world simulations, enhancing their grasp of abstract engineering principles by offering practical, hands-on experience. Open-ended responses in the post-class survey highlighted students' appreciation for these methods, describing PBL as instrumental in fostering collaborative problem-solving skills and BIM as a valuable tool for visualizing the technical aspects of structural design.

Student confidence also saw an increase from pre-class levels. Post-module, confidence levels rose to an average of 4.5, with 62.5% of students expressing high confidence in recalling and applying the key concepts learned throughout the course. This increase underscores the success of the module's approach in equipping students with not only technical skills but also the confidence needed to apply them effectively.

The professor's post-class feedback indicated that the module met its primary objectives, achieving an overall rating of 4/5 for its effectiveness in providing students with hands-on, practical experience. While several challenges anticipated in the pre-class survey did arise—including logistical delays and weather-related disruptions—these were managed through adaptive strategies. The professor reported a particular instance where adverse weather (high winds) impacted crane operations, temporarily halting student activities. Such challenges underscored the realistic nature of the project-based module, exposing students to the unpredictability inherent in engineering projects.

The professor rated the PBL approach as highly effective (5/5) for engaging students in real-world problem-solving, confirming its role in enhancing student involvement and collaborative skills. However, digital tools such as BIM were rated at 3/5, suggesting that while beneficial, their potential may not have been fully realized. The professor noted that with additional training and support, BIM could be better integrated to support student learning in future iterations of the module.

In terms of resource availability, the professor initially rated digital support low (2/5) but increased this rating to 4/5 as resources were adjusted throughout the course. This adaptive response highlights the importance of real-time adjustments in educational delivery to ensure a supportive learning environment.

4. DISCUSSION

The findings of this study indicate a meaningful impact of PBL and BIM on enhancing student engagement, confidence, and understanding within a civil engineering module. This section discusses the effectiveness of these methods and their implications for curriculum development, as well as instructional challenges identified through the professor's feedback.

4.1 Impact of Problem-Based Learning on Student Engagement and Problem-Solving Skills

The study's results demonstrate that the PBL approach significantly enhanced student engagement and practical problem-solving skills. Initially, students expressed concerns regarding complex topics, particularly in areas like environmental risk management. By the end of the module, a half of students had moderated their perception of the difficulty level, suggesting that PBL's collaborative and hands-on framework alleviated some of their initial concerns. Through PBL, students were encouraged to work in teams, think critically, and address open-ended problems, which are all critical skills for real-world engineering projects.

The professor's post-class observations confirmed the efficacy of PBL, rating it highly effective for engaging students in real-world scenarios. The industry interactions and collaborative group projects fostered

through PBL provided students with practical insights into civil engineering, simulating the professional environments they will encounter after graduation. The alignment between student and professor feedback supports the notion that PBL, when applied effectively, offers an immersive experience that can bridge the gap between theoretical knowledge and professional skills.

4.2 Effectiveness of Digital Models in Enhancing Understanding

The integration of BIM significantly contributed to students' comprehension of complex engineering concepts, with 62.5% of students rating it as effective in enhancing their understanding of structural analysis. BIM provided a dynamic, interactive platform for students to visualize structural elements and simulate project processes, reinforcing theoretical knowledge through practical application.

However, the professor rated BIM's effectiveness at 3/5, noting that while beneficial, its full potential was not entirely realized due to limitations in technical support and training. This suggests that future implementations of BIM in similar modules could benefit from additional resources and support, ensuring that students are comfortable using the software and can fully engage with its capabilities. Incorporating introductory BIM tutorials or increased support could help students maximize the tool's educational potential, further bridging the gap between conceptual understanding and technical skill development.

4.3 Comparison of Pre- and Post-Class Perceived Challenges

The comparison between pre-class and post-class survey data reflects a shift in students' perceived challenge levels, underscoring the effectiveness of PBL and BIM in preparing students to tackle complex topics. Initially, over half the students expected significant difficulty; however, by the end of the course, only 50% maintained this perception, indicating that the module's support structures and practical applications helped students feel more prepared. This shift demonstrates the module's success in addressing initial concerns, as students grew more comfortable applying theoretical concepts to real-world problems through PBL and BIM.

The professor's perspective provided additional context, highlighting how logistical issues and permitting delays impacted the course's delivery. Despite these obstacles, the professor's commitment to adaptive teaching strategies helped sustain student engagement and fostered a supportive learning environment. This reflects the importance of instructor flexibility and responsiveness in project-based courses, where external constraints can influence the learning experience.

4.4 Correlation Between Familiarity with PBL and Learning Outcomes

The data suggests a positive correlation between students' familiarity with PBL and their learning outcomes. Students who had moderate familiarity with PBL before the course rated its effectiveness higher than those who were new to the approach. This finding supports the recommendation for preparatory sessions introducing students to PBL frameworks early in the curriculum, equipping them with foundational skills and helping align expectations. By offering introductory workshops or early PBL exposure, educators can further enhance students' engagement and effectiveness in PBL-centered modules.

4.5 Resource Utilization and Accessibility

Both the student and professor feedback underscored the importance of adequate resources and real-time support for optimizing the learning experience. The professor's initial low rating of digital support (2/5) improved as resources were adjusted throughout the module, eventually reaching a 4/5 rating. This improvement underscores the necessity of adaptive resource allocation in educational delivery, particularly in modules that heavily rely on digital tools like BIM. Ensuring access to technical support and relevant resources not only enhances the effectiveness of digital tools but also empowers students to experiment with complex simulations and refine their solutions.

Furthermore, the professor emphasized broader evaluation criteria such as sustainability and communication skills as essential learning outcomes. By incorporating these criteria, the module aligns with current professional expectations and equips students with holistic competencies that extend beyond technical skills. This comprehensive approach to student evaluation, including both technical and soft skills, ensures that civil engineering graduates are prepared to meet the diverse challenges of their field.

5. CONCLUSIONS

This study demonstrates the positive impact of integrating PBL and BIM within a civil engineering module, particularly in enhancing student engagement, confidence, and practical understanding. Traditional civil engineering education, often rooted in lecture-based approaches, can struggle to prepare students for the complexities of professional practice. By immersing students in real-world scenarios and collaborative problem-solving, PBL effectively addresses these gaps, fostering skills essential for handling multifaceted engineering challenges. BIM's role as an interactive tool further supports learning by providing visual, hands-on experiences that deepen comprehension of structural elements and project processes.

The pre-class and post-class surveys revealed several insights. Initially, 57.1% of students expected moderate to high difficulty, particularly in complex areas like environmental risk management. However, with the support of PBL and BIM, students' confidence levels improved significantly, from an average of 4.29 to 4.5, with 50% of students rating the module's difficulty as moderate by the end. Moreover, 62.5% of students found BIM

effective in enhancing their understanding, affirming the value of digital tools in bridging theoretical and practical knowledge. The professor's perspective corroborated these findings, noting that while logistical and permitting delays presented challenges, the real-world orientation of the module aligned well with professional engineering demands.

The findings of this study support the inclusion of PBL and BIM in civil engineering education as effective tools for fostering real-world skills. To optimize the benefits of PBL, we recommend introducing students to PBL frameworks early in their curriculum through preparatory workshops or introductory sessions. This preparation can help align expectations and build foundational skills, leading to more effective engagement in PBL modules. In addition, enhancing technical support and resources for BIM will enable students to fully leverage digital models, reinforcing their ability to apply complex concepts and refining their design capabilities.

Further research should explore the long-term effects of PBL and BIM on students' knowledge retention and career preparedness. Longitudinal studies tracking graduates' performance in professional settings could provide valuable insights into the practical benefits of these educational methods. Studies could also investigate how continuous resource adaptation, including technical support and real-time feedback, impacts student learning and engagement in digitally supported PBL modules. Future studies may consider examining the integration of sustainability and communication skills into student evaluations, providing a more holistic approach to skill development in civil engineering education.

All in all, integrating PBL and BIM within civil engineering curricula bridges the gap between theoretical learning and practical application, equipping students with critical technical and collaborative skills. By fostering these competencies, civil engineering programs can better prepare graduates to meet the evolving challenges of the industry and contribute effectively to sustainable and resilient infrastructure development.

ACKNOWLEDGMENTS

This work was supported by an EU Commission-funded project, the AEGIR project [grant number 101079961], and the OMICRON project, EU H2020 [grant number 955269].

REFERENCES

- Alavi, H., Building information modeling for facility managers, Universitat Politècnica de Catalunya, 2022.
- Azam, R., Farooq, M. U., & Riaz, M. R. (2024). A Case Study of Problem-Based Learning from a Civil Engineering Structural Analysis Course. *Journal of Civil Engineering Education*, 150(3).
- Čubela, D., Rossner, A., & Neis, P. (2023). Using Problem-Based Learning and Gamification as a Catalyst for Student Engagement in Data-Driven Engineering Education. *Education Sciences*, 13(12), 1223.
- Das, D., Ariffin, K., & Abu Hashim, A. H. (2023). Employing the Problem-Based Learning Approach in Civil Engineering Education: The Highway Engineering Experience. *SpringerLink*.
- Farooq, M. U., & Riaz, M. R. (2024). Analysis of Problem-Based Learning vs. Traditional Teaching among Students in Civil Engineering. *Journal of Civil Engineering Education*, 150(3).
- Ullmann, T., Wild, F., & Scott, P. (2023). Learning through reflection: The role of BIM in engineering education. *International Journal of Engineering Education*, 35(2).
- Zou, Y., Kiviniemi, A., & Jones, S. W. (2021). BIM and the impact on building regulations. *Automation in Construction*, 83, 134-144.

Enhancing Tower Crane Operation Skills in Aerial Construction Factories through Virtual Reality Training: A Controlled Study

Jingfeng Yuan¹, Lingxiao Wang², Feixiang Hu³, Yang Su⁴, Chen Li⁵, Hao Tian⁶, Weidong Lin⁷

1) Professor, Dept. of Construction and Real Estate, School of Civil Engineering, Southeast Univ., Nanjing, China. Email: jingfeng-yuan@seu.edu.cn

2) Ph.D. Candidate, Dept. of Construction and Real Estate, School of Civil Engineering, Southeast Univ., Nanjing, China. E-mail: lingxiao@seu.edu.cn

3) Ph.D. student, Dept. of Construction and Real Estate, School of Civil Engineering, Southeast Univ., Nanjing, China. E-mail: 230238389@seu.edu.cn

4) Student, Dept. of Construction and Real Estate, School of Civil Engineering, Southeast Univ., Nanjing, China. Email: 213223669@seu.edu.cn

5) Student, Dept. of Construction and Real Estate, School of Civil Engineering, Southeast Univ., Nanjing, China. E-mail: 213221666@seu.edu.cn

6) Student, Dept. of Construction and Real Estate, School of Civil Engineering, Southeast Univ., Nanjing, China. E-mail: 3119454955@qq.com

7) Student, Dept. of Construction and Real Estate, School of Civil Engineering, Southeast Univ., Nanjing, China. Email: 1442174242@qq.com

Abstract: This study examines the impact of virtual reality (VR) training on enhancing the skills of tower crane operators in Aerial Construction Factories (ACFs). As ACF operations become increasingly complex with higher safety demands, traditional training methods struggle to simulate high-risk scenarios and provide sufficient hands-on experience. By leveraging VR technology, this study offers an immersive and interactive training environment where operators can practice in risk-free, controlled virtual scenarios. VR-based experiments were conducted to measure performance in task completion time, operational accuracy, self-efficacy, and the frequency of collision warnings during simulated tasks. The results demonstrated significant improvements in both task accuracy and operator confidence, along with reductions in task completion time and collision warnings. These findings suggest that VR training addresses the limitations of conventional methods, offering a more cost-effective, scalable, and safe solution for improving crane operator skills within ACFs. Ultimately, the study highlights the transformative potential of VR technology in enhancing operator proficiency, safety awareness, and accident prevention in the challenging and hazardous environment of ACFs.

Keywords: Aerial construction factory, Virtual reality, Tower crane operators, Skill enhancement

1. INTRODUCTION

Aerial Construction Factories (ACFs) represent a transformative advancement in modern construction systems, integrating tower cranes, concrete pumps, and other large-scale machinery to streamline operations in high-rise building projects (Melenbrink et al., 2020). Their ability to enhance efficiency, ensure construction quality, and improve safety in complex environments highlights their critical role in meeting the demands of modern urban development (Ma et al., 2024). These systems are characterized by their capacity to execute complex, large-scale construction tasks efficiently, often under challenging environmental conditions (Ma et al., 2024). The operational environment of ACFs is inherently complex due to factors such as high-altitude work, limited spatial maneuverability, and dynamic interactions among multiple equipment systems (Cai et al., 2019). Within this context, tower crane operators play a pivotal role, not only managing routine tasks such as lifting and transporting heavy materials but also performing precise and synchronized operations under constrained and elevated conditions (Ali et al., 2024). These heightened operational demands necessitate advanced control strategies and exceptional skill levels to ensure safety and productivity (Ali et al., 2024). The complicated nature of these tasks amplifies the risks of human error, underscoring the critical significance of enhancing operator competencies through targeted training and technological support (Shringi et al., 2023).

Enhancing the control skills of tower crane operators is crucial for ensuring both safety and efficiency in construction projects, particularly within ACFs, where the complexity and risks associated with crane operations are significantly improved (Hammad et al., 2023). Tower cranes in ACFs not only face the usual challenges of lifting and moving heavy materials but also require precise coordination in constrained, high-altitude environments (Chen et al., 2022). These conditions demand greater precision and control to prevent accidents and minimize material handling errors (Ni et al., 2024). Prior studies have highlighted that proficient crane operators can significantly reduce construction time and costs, contributing to the overall success of ACF projects (Sadeghi et al., 2023). As ACFs and other construction projects become more automated and complex, the requirements for continuous skill development and advanced training programs for crane operators become even more critical (Ali et al., 2024).

The skill training of tower crane operators is crucial for realizing safety and efficiency during construction

work(Song et al., 2021). However, traditional training methods, such as classroom instruction and on-the-job training, have limitations in addressing complex construction environments and emergencies (Moon et al., 2019). In recent years, VR technology has gained significant attention as an innovative industrial training tool due to its immersive, interactive, and safe characteristics (Nair et al., 2024). The origins of VR can be traced back to the 1960s, when Ivan Sutherland and his team developed the "Sword of Damocles," the first head-mounted display system (Sutherland, 1968). This groundbreaking invention integrated real-time graphics and motion tracking, laying the foundation for modern VR technologies (Steuer, 1992). Over the decades, studies have demonstrated that VR offers significant advantages in operator training (Adami et al., 2021), simulating hazardous scenarios (Cross et al., 2022), providing collision warnings (Li et al., 2018), enhancing self-efficacy (Shringi et al., 2023), replicating complex tasks (Xie et al., 2021), and creating high-altitude work environments (Melenbrink et al., 2020).

The high-risk nature of tower crane operations within ACF makes improving operator abilities very necessary. Inadequate training can lead to heavy accidents, property damage, and even fatalities (Wiethorn, 2018). Human error, often stemming from insufficient training and lack of experience, is a leading cause of crane-related accidents, especially in the challenging environments of ACFs (Zhang et al. 2020). Therefore, effective training programs that not only enhance technical skills but also improve decision-making and boost confidence in high-pressure environments are essential (Z. Zhang et al., 2022). However, despite the significant advantages of VR technology in tower crane operation training, its application in the emerging context of ACFs remains limited. The ACF environment is characterized by high-altitude operations, spatial constraints, and multi-equipment coordination, which impose higher demands on operators' skills and safety awareness (Ma et al., 2024). Currently, studies on VR training systems tailored to the specific operational conditions of ACFs are scarce, highlighting the urgent need for further exploration to address the unique requirements of this domain.

This study explores the effectiveness of VR training for tower crane operators within ACFs by examining key performance metrics such as task completion times, accuracy, self-efficacy, and the frequency of collision warnings. By providing empirical evidence of VR's benefits, this study contributes to the growing body of literature on VR applications in industrial training and aims to inform the development of more effective training programs for improving crane operator skills, particularly in the demanding environments of ACFs.

2. METHOD

2.1 Procedure

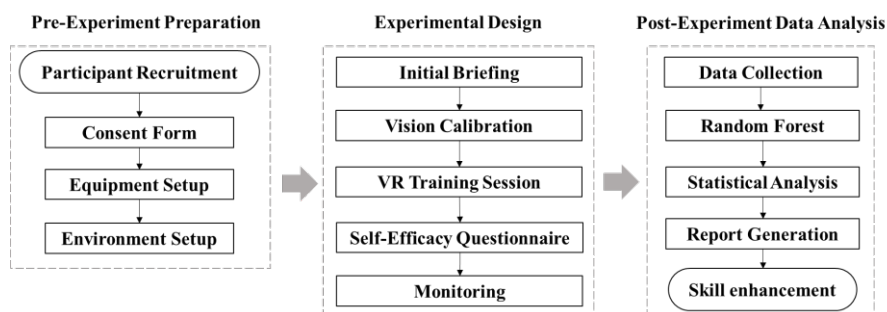


Figure 1 Experimental workflow for VR training in tower crane skills

According to Moon et al. (2019) and Adami et al. (2021), this framework diagram (Figure 1) outlines the comprehensive process of the experimental study conducted within the context of ACFs, divided into three main phases: Pre-Experiment Preparation, Experimental Design, and Post-Experiment Data Analysis, as shown in followed.

In the Pre-Experiment Preparation phase, participants are recruited and provided with consent forms (Shringi et al., 2023). The necessary VR equipment is set up, and the ACF simulation environment is prepared to reflect the complex, high-altitude conditions typical of these environments (Moon et al., 2019). This ensures that the experiment replicates the real-world challenges faced by tower crane operators in ACFs.

In the Experimental Design phase, participants receive an initial briefing, followed by vision calibration to ensure accurate data collection (Xie et al., 2021). They then engage in a VR training session simulating tower crane operations in an ACF setting. After completing the training, participants are asked to fill out a self-efficacy questionnaire (Ye et al., 2022). Continuous monitoring throughout this phase ensures the integrity and accuracy of the collected data, particularly in simulating high-risk ACF scenarios (Adami et al., 2021).

The Post-Experiment Data Analysis phase involves collecting all performance metrics, such as task completion times, accuracy, and collision warnings, followed by statistical analysis to assess the impact of VR training in an ACF context (Song et al., 2021). Reports are generated, and a feedback session is conducted to

present the findings. This structured approach, tailored to the unique demands of ACFs, ensures the reliability and validity of the experimental outcomes while addressing the specific complexities of training tower crane operators in such environments (Shringi et al., 2022).

2.2 Pre-Test

(1) Participant Selection

This study recruited 20 individuals from civil engineering-related majors, representing a range of experience levels in tower crane operations, from novices to more experienced professionals. Figure 2 illustrates part of the experimental scene where participants are equipped with VR headsets and controllers (Xie et al., 2021). The selection aimed to include a diverse group of participants to comprehensively evaluate the effectiveness of VR training across different skill sets. Each participant was briefed on the study's objectives, procedures, and potential risks before signing informed consent forms, ensuring voluntary participation and understanding of the experiment (Shringi et al., 2022). The VR technology provided participants with an immersive, realistic environment to practice tower crane operations under various simulated conditions (Song et al., 2021), as depicted in Figure 2.



Figure 2 Participants engaged in VR-based tower crane operation

(2) VR Model Development

The VR model was carefully designed to simulate real-world construction environments, with a focus on the operational challenges within ACFs (Adami et al., 2021). In these environments, complex machine interactions present significant challenges for crane operators (W. Zhou et al., 2018). Figure 3 illustrates the setup of the VR equipment, which includes a VR headset connected to a high-performance computer capable of generating a highly immersive virtual environment (Yang et al., 2022). This environment replicates the spatial dynamics of the ACF, including the coordination between tower cranes and other machinery such as concrete placement machines.

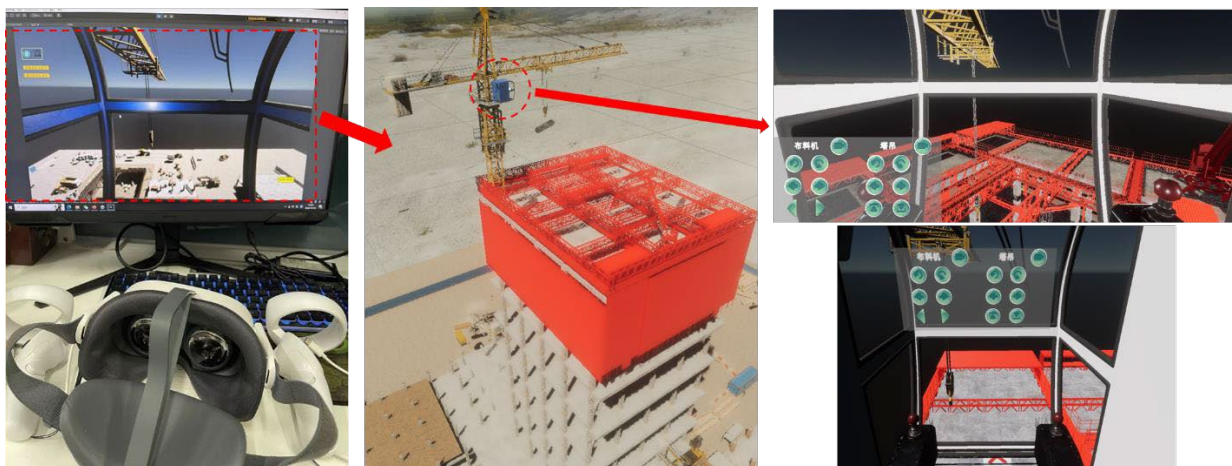


Figure 3 VR equipment and environment setup

The VR model provides detailed visuals from the operator's perspective, simulating crane operations with a high degree of realism (Dhalmahapatra et al., 2021). Participants engage in lifelike training sessions, allowing them to practice operations in a safe, controlled setting that enhances their familiarity with the equipment and the operational complexities of ACF environments (Yang et al., 2022).

(3) Establishing a Safety Zone

To ensure the physical safety of participants during the VR training sessions, a designated safety zone was

created (Tak et al., 2021). This zone, visually marked in the testing area, provided a secure space where participants could operate the VR equipment without the risk of physical collisions or injury. The boundaries of the safety zone served as a constant reminder for participants to remain within the designated area during the simulation (Sudiarno et al., 2024).

Before starting the experiment, participants were required to sign informed consent forms after being thoroughly briefed on the study's objectives, procedures, potential risks, and their rights as participants (Shringi et al., 2022). Establishing the safety zone and obtaining informed consent ensured the experiment followed ethical guidelines and offered participants confidence in the safety measures. These precautions were crucial for maintaining both the reliability and validity of the experiment while prioritizing participant well-being (Shringi et al., 2023).

2.3 Experimental Design

(1) Experiments Goal

In the VR experiment, three target points labeled A, B, and C were designated for the crane operator to place the load. These points were chosen to evaluate the operator's precision and control during operation (Weichert et al., 2013). The Euclidean distance between the actual drop points and the target points during the VR training sessions was analyzed as a quantitative measure of the operator's handling skills (Oagaz et al., 2021). A smaller Euclidean distance indicated greater accuracy and precision in the operator's actions, reflecting improved operational capability (Kaplan et al., 2021). This approach not only validated the effectiveness of the training program but also provided valuable insights into areas requiring further improvement, offering a clear and actionable metric for performance evaluation (T. Zhou et al., 2022).

(2) Training Sessions

To further enhance the control abilities of crane operators, an intermediate training tutorial was incorporated between the two main experimental sessions. This tutorial, developed using Python, was designed to simulate a controlled environment where operators maneuver a load around virtual representations of workers and materials to reach a designated target point. The tutorial provided a safe, risk-free setting for operators to focus on improving their precision, dexterity, and operational control (Xie et al., 2021). Through this interactive exercise, operators were challenged to navigate obstacles while maintaining accuracy in load placement, offering a hands-on opportunity to refine their spatial awareness and coordination skills (Yang et al., 2022).

The tutorial was not merely a repetition of the VR training but served as a targeted supplement, allowing operators to practice critical techniques in a more focused and interactive manner. By emphasizing tasks that required careful maneuvering and precise execution, it reinforced the foundational concepts introduced during the initial VR simulations (Masiello et al., 2022). Furthermore, the tutorial facilitated gradual skill development by presenting operators with scenarios that mimicked real-world complexities, thereby bridging the gap between theoretical knowledge and practical application (Abduppattayevich et al., 2023). This additional training session proved to be a crucial component in fostering sustained improvement in crane operation proficiency and ensuring that operators were better prepared for the challenges of real-world tasks (Adami et al., 2022).

2.4 Post-Test

(1) Recording Experiment Time

In the VR experiment, task completion time serves as a key indicator of crane operator proficiency (Adami et al., 2021). By tracking the time taken to complete transportation tasks in both VR sessions, we can assess improvements in the operators' control abilities. Shorter completion times typically reflect greater familiarity with the controls and higher skill levels in operating the crane (Fang et al., 2018). This temporal analysis provides a clear, quantifiable measure of operator efficiency. Comparing the task completion times between the two experimental sessions enables us to evaluate the effectiveness of the training program. A significant reduction in time from the first to the second session suggests enhanced operational competence, demonstrating the operator's increased dexterity and confidence in handling the crane. This method offers a comprehensive evaluation of the training's impact on improving crane operation efficiency.

(2) Recording Operational Accuracy

$$d = \frac{\sqrt{(p_{Ax}^a - p_{Ax}^t)^2 + (p_{Ay}^a - p_{Ay}^t)^2} + \sqrt{(p_{Bx}^a - p_{Bx}^t)^2 + (p_{By}^a - p_{By}^t)^2} + \sqrt{(p_{Cx}^a - p_{Cx}^t)^2 + (p_{Cy}^a - p_{Cy}^t)^2}}{3} \quad (1)$$

In the VR experiment, the accuracy of crane operators was evaluated by measuring the average Euclidean distance between the actual drop points and the target points A, B, and C, as defined by Equation (1) (Schneider et al., 2020). This metric was used to assess the precision of each operator's performance across two experimental sessions. By calculating the mean Euclidean distance, we quantitatively determined how closely the operators positioned the load relative to the intended targets under varying conditions. This method provided a robust measure of accuracy, allowing us to compare the performance of each operator between the two sessions. A decrease in the average Euclidean distance in the second session would indicate improved control and precision, which would suggest the effectiveness of the training and VR simulation in enhancing crane operation skills (Wang et al., 2024). This assessment approach ensures a clear and objective evaluation of each operator's potential

progress and overall capability.

(3) Recording Self-Efficacy

Appendix 1 presents a self-efficacy scale, developed based on self-efficacy theory (Bandura et al., 1999; Ye et al., 2022), to assess tower crane operators' perceptions of their ability to control and operate cranes. The scale includes ten statements addressing various aspects of problem-solving, handling unforeseen situations, and managing unexpected events in the context of crane operations. Respondents rate their agreement with each statement on a scale from "Completely Disagree" to "Completely Agree" (Ye et al., 2022). In the VR experiment, this self-efficacy scale is administered to operators before and after participating in the VR training to measure changes in their confidence regarding crane operation. By comparing the pre- and post-experiment scores, researchers can evaluate whether VR training effectively enhances the operators' self-efficacy. An increase in self-efficacy scores after the training would indicate that operators feel more capable and confident in handling crane operations, reflecting improved understanding and control of the tower crane (Song et al., 2021). This suggests that VR training is a valuable tool for enhancing both the practical skills and confidence of tower crane operators (Shringi et al., 2023).

(4) Recording Collision Alert Frequency

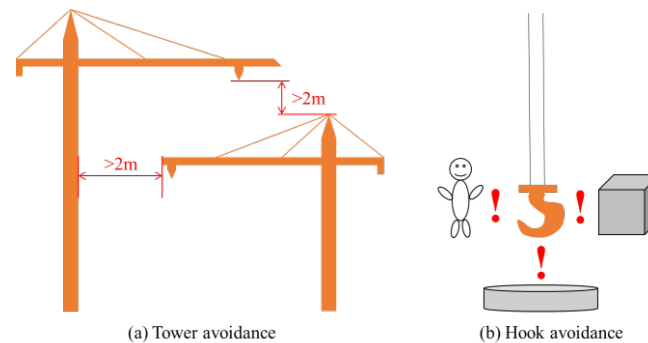


Figure 4 For towers and hooks, the scope of collision warning

Figure 4 illustrates the collision alert mechanism used in a VR experiment for tower crane operators. In Figure 4(a), the diagram outlines the collision alert range between different tower cranes, specifying a safe distance of over 2 meters (Hwang, 2012). Figure 4(b) shows the collision alert range for the crane's hook with humans and materials, highlighting the importance of maintaining a safe operating distance (Hwang, 2012). These visual cues within the VR environment are designed to warn operators of potentially hazardous operations, promoting safer crane operation practices. The primary objective of the experiment is to assess the operational proficiency of tower crane operators by recording the frequency of collision alerts triggered during VR simulations. By analyzing these alerts, researchers can identify areas where operators may need additional training or practice.

3. RESULTS

3.1 Performance Improvements Across Key Metrics

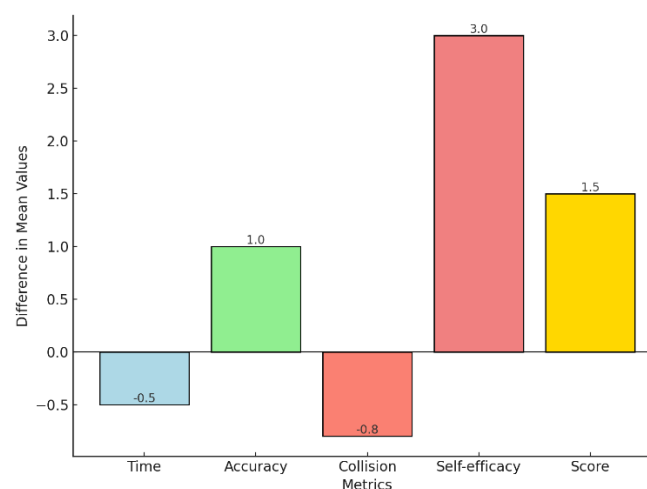


Figure 5 Comparison of performance improvements across key metrics

Figure 5 highlights the improvements observed in key performance metrics during the second experiment. After normalization and averaging, task completion time decreased by approximately 0.5 units, indicating that participants became more efficient in handling the crane, completing tasks in less time. The accuracy of crane operations showed a notable improvement, with a 1.0 unit increase, reflecting better precision in positioning the crane's load. Additionally, the number of collision alerts dropped by around 0.8 units, signifying enhanced control and fewer errors in avoiding hazards.

The most significant improvement was in self-efficacy, with an increase of over 3.0 units, indicating a considerable boost in participants' confidence and ability to handle crane operations. The overall score, which combines various performance factors, also improved by 1.5 units, showcasing the general enhancement in operational proficiency. These improvements demonstrate the effectiveness of the VR-based training program, particularly in boosting confidence, accuracy, and safe crane operation.

3.2 Relationship Between Performance Metrics and Operational Skill Enhancement

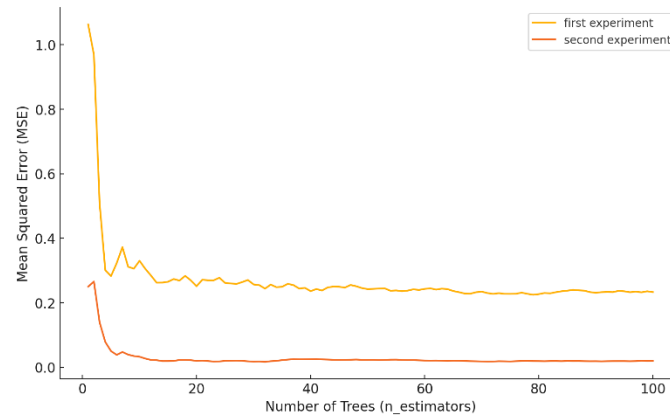


Figure 6 Random forest training iterations vs MSE

As shown in Figure 6, a Random Forest model was employed to evaluate the relationship between key performance metrics and operational performance in both the first and second experiments (Fang et al., 2021). The model's mean squared error (MSE) decreased significantly with the increase in the number of trees ($n_{\text{estimators}}$), particularly in the second experiment, which showed a much lower and more stable MSE compared to the first experiment. This suggests that participants in the second session exhibited more consistent improvements, with enhanced predictive accuracy, which corresponds to better overall crane operation proficiency.

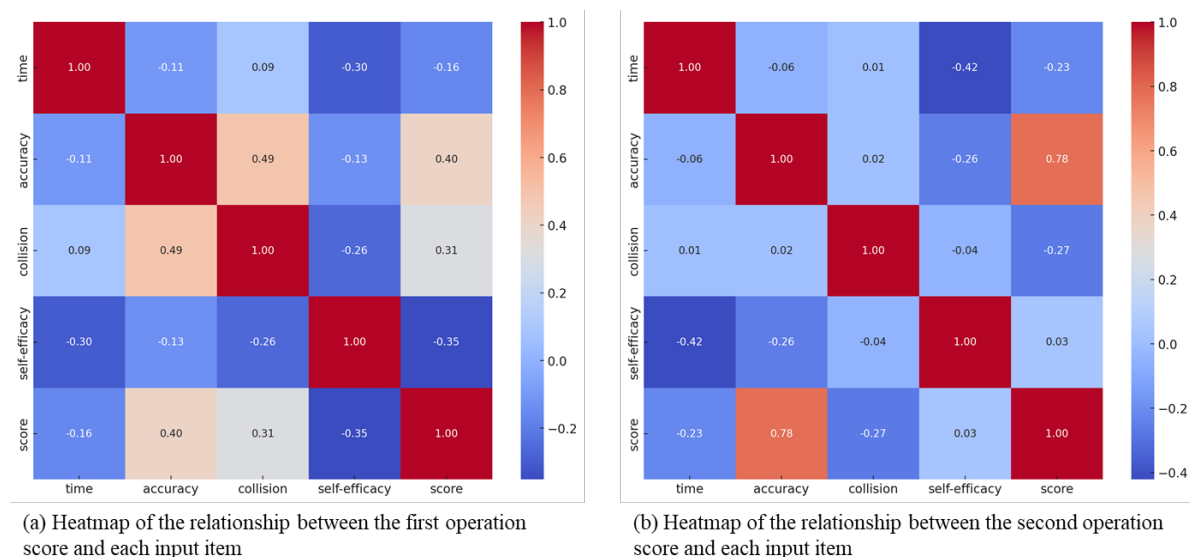


Figure 7 Heatmaps of the relationship between the first and second operation score and each input item

Figure 7 illustrates heatmaps depicting the correlation between the operational scores and each input feature for both experiments. In the first operation (Figure 7a), accuracy and score are moderately correlated (0.31),

while the relationship between accuracy and collision is weak (0.26). However, in the second experiment (Figure 7b), the correlation between accuracy and score strengthened significantly (0.78), underscoring the critical role that improved precision plays in enhancing performance. Additionally, the heatmaps reveal that self-efficacy exhibited a strong correlation with scores in both experiments, but the second experiment (0.82) shows a clear improvement. This indicates that the participants' increased confidence was strongly tied to their overall performance improvements, making self-efficacy a key factor in achieving proficiency.

4. DISCUSSION

The findings from this study demonstrate that VR training significantly enhances the operational proficiency of tower crane operators in ACF environments. Key performance metrics, including task completion time, accuracy, collision alerts, and self-efficacy, showed substantial improvements across two experimental sessions, highlighting the training's impact on crane operation skills. The observed reduction in task completion time, paired with increased accuracy, underscores the effectiveness of VR-based training in enhancing both operational speed and precision. Literature supports that VR simulations provide a risk-free, immersive training environment, allowing operators to refine skills that contribute to greater efficiency and safety (Shringi et al., 2022). As illustrated in Figures 6 and 7, the Random Forest model used in this study indicates a significant increase in the predictive power of these metrics from the first to the second experiment, with the correlation between accuracy and overall score particularly strengthened in the second session. This outcome aligns with studies showing that precision plays a critical role in improving operational performance in complex environments (Cross et al., 2022).

The concept of self-efficacy emerged as a pivotal factor in enhancing operational performance. Participants who experienced an increase in confidence demonstrated a marked improvement in efficiency and were more adept at preventing accidents. Literature suggests that self-efficacy in high-risk operations, such as tower crane handling, is linked to better task execution and fewer errors, as seen in studies where increased confidence correlated with higher accuracy and operational safety (Adami et al., 2021). The strong association between self-efficacy and operational scores observed in the second experiment (Figure 7b) underscores the critical role of psychological preparedness, particularly confidence and decision-making abilities, in crane operations. Furthermore, the observed decrease in collision alerts aligns with findings that VR-based simulations, by enhancing situational awareness, yield safer operational outcomes (Shringi et al., 2023).

Table 1. Solutions for different severity levels of operational issues in tower crane operations

Types of issues	Mild	Moderate	Severe
slow operation	Increase practice frequency to improve familiarity with operational details, gradually reducing task completion time without major errors (Gallagher et al., 2005).	Participate in specialized training to gain a deeper understanding of operational methods and optimize workflows, minimizing delays caused by repeated actions (Aziz et al., 2013).	Implement smart systems to simplify operational processes, addressing significant delays caused by unfamiliarity and high error rates (Maurer et al., 2010).
low operational accuracy	Focus on mastering basic skills through consistent practice and learning from experience to improve precision and avoid errors (Aziz et al., 2013).	Engage in professional training to understand operational requirements better and overcome challenges through frequent practice (Maurer et al., 2010).	Introduce smart technologies to reduce operational complexity, thereby minimizing errors and preventing accidents during execution (Moniri et al., 2016).
low self-efficacy	Set achievable goals to build confidence gradually, improve familiarity with procedures, and reduce errors through consistent practice (Gallagher et al., 2005).	Strengthen personal capabilities through regular practice and feedback, improving confidence in task execution (Kopp et al., 2020).	Provide additional support and collaborative opportunities to rebuild confidence, while minimizing errors through guided assistance (Adams-White et al., 2018).
frequent collision alerts	Enhance communication to ensure operators fully understand task requirements; encourage slow, deliberate actions to improve safety awareness (Aziz et al., 2013).	Engage in scenario-based simulations to identify and address safety-critical tasks, fostering clear communication and feedback (Gallagher et al., 2005).	Deploy intelligent error-prevention systems to reduce the likelihood of human mistakes, ensuring operators understand tasks and safety concepts clearly (Maurer et al., 2010).

Table 1 outlines solutions for addressing various levels of operational challenges, including slow operation, low accuracy, low self-efficacy, and frequent collision alerts. For minor issues, increasing practice time and providing opportunities for task repetition are effective. However, for more severe issues, such as frequent collisions or significant delays in task completion, specialized training or implementing smart systems becomes essential. These systems aid in minimizing human errors and simplifying complex operations. Studies suggest that smart systems can significantly reduce error rates in complex operational contexts (Li et al., 2018). Additionally,

VR training, by simulating high-risk scenarios and enabling repeated practice, enhances operators' situational awareness and confidence, thereby improving overall operational safety and efficiency (Shringi et al., 2023). The strategies presented in Table 1 offer clear guidance for addressing operational inefficiencies and risks across various severity levels, providing valuable insights for improving operator performance and safety in high-risk environments like ACFs.

5. CONCLUSIONS

This study has provided clear evidence that VR training significantly improves the operational skills of tower crane operators, especially within the challenging environment of ACFs. The experimental results revealed that after VR training, task completion time was reduced by approximately 0.5 units, while operational accuracy increased by 1.0 unit. These findings demonstrate that VR training effectively enhances both the speed and precision of crane operations, which is particularly critical in ACFs, where high-altitude work demands precise control and quick execution to prevent costly delays or accidents. Additionally, the frequency of collision alerts decreased by around 0.5 units, further underscoring the role of VR in improving operators' safety awareness and operational efficiency.

The study also highlighted the significance of self-efficacy in determining operational performance. Self-efficacy increased by over 3.0 units in the second experiment, and operators with higher confidence performed better in terms of accuracy and were more capable of handling complex and unexpected situations. Specifically, the correlation between self-efficacy and operational score reached 0.82 in the second experiment, demonstrating a strong link between psychological readiness and operational success. The VR training not only improved technical skills but also boosted operators' confidence by providing repeated exposure to high-risk simulated scenarios, reducing the likelihood of human error and enhancing overall safety.

Furthermore, the analysis of different operational issues in Table 1 emphasized the need for tailored solutions based on the severity of the problem. For mild issues such as slow operation or occasional errors, increased practice time and feedback can lead to significant improvements. However, for more severe challenges like low operational accuracy or frequent collision alerts, more advanced interventions such as smart systems to reduce human error are necessary. The solutions outlined in Table 1 suggest that a combination of VR-based training and targeted technological support can effectively address operational challenges, improving performance and safety in high-risk environments like ACFs.

This study has limitations that should be addressed. The experiment involved students, which may not fully reflect industry conditions, and the small sample size limits the reliability of the results. Additionally, challenges in industry adoption, such as high initial costs, the need for customized content, and resistance to new methods, must be considered. Future research should focus on larger, more diverse samples and demonstrate the cost-effectiveness and long-term benefits of VR training to promote its broader acceptance in the construction industry.

ACKNOWLEDGMENTS

The work described in this paper was supported by National Key R&D Program of China (grant no. 2022YFC3802201) and Postgraduate Research & Practice Innovation Program of Jiangsu Province (grant no. KYCX23_0334).

REFERENCES

- Abdupattayevich, P. A., & G'ayratovich, S. S. (2023). Improving the effectiveness of the training process of students. *International Journal of Social Science & Interdisciplinary ISSN: 2277-3630 Impact Factor: 8.036*, 12(09), 50–52.
- Adami, P., Rodrigues, P. B., Woods, P. J., Becerik-Gerber, B., Soibelman, L., Copur-Gencturk, Y., & Lucas, G. (2021). Effectiveness of VR-based training on improving construction workers' knowledge, skills, and safety behavior in robotic teleoperation. *Advanced Engineering Informatics*, 50, 101431. doi: <https://doi.org/10.1016/j.aei.2021.101431>
- Adami, P., Rodrigues, P. B., Woods, P. J., Becerik-Gerber, B., Soibelman, L., Copur-Gencturk, Y., & Lucas, G. (2022). Impact of VR-based training on human–robot interaction for remote operating construction robots. *Journal of Computing in Civil Engineering*, 36(3), 4022006.
- Adams-White, J. E., Wheatcroft, J. M., & Jump, M. (2018). Measuring decision accuracy and confidence of mock air defence operators. *Journal of Applied Research in Memory and Cognition*, 7(1), 60–69.
- Ali, A. H., Zayed, T., Wang, R. D., & Kit, M. Y. S. (2024). Tower crane safety technologies: A synthesis of academic research and industry insights. *Automation in Construction*, 163, 105429.
- Aziz, R. F., & Hafez, S. M. (2013). Applying lean thinking in construction and performance improvement. *Alexandria Engineering Journal*, 52(4), 679–695.
- Bandura, A., Freeman, W. H., & Lightsey, R. (1999). *Self-efficacy: The exercise of control*. Springer.

- Cai, S., Ma, Z., Skibniewski, M. J., & Bao, S. (2019). Construction automation and robotics for high-rise buildings over the past decades: A comprehensive review. *Advanced Engineering Informatics*, 42, 100989.
- Chen, Y., Zeng, Q., Zheng, X., Shao, B., & Jin, L. (2022). Safety supervision of tower crane operation on construction sites: An evolutionary game analysis. *Safety Science*, 152, 105578.
- Cross, J., Boag-Hodgson, C., Ryley, T., Mavin, T. J., & Potter, L. E. (2022). Using extended reality in flight simulators: a literature review. *IEEE Transactions on Visualization and Computer Graphics*, 29(9), 3961–3975.
- Dhalmahapatra, K., Maiti, J., & Krishna, O. B. (2021). Assessment of virtual reality based safety training simulator for electric overhead crane operations. *Safety Science*, 139, 105241.
- Fang, Yihai, Cho, Y. K., Durso, F., & Seo, J. (2018). Assessment of operator's situation awareness for smart operation of mobile cranes. *Automation in Construction*, 85, 65–75.
- Fang, Yuan, Lu, X., & Li, H. (2021). A random forest-based model for the prediction of construction-stage carbon emissions at the early design stage. *Journal of Cleaner Production*, 328, 129657.
- Gallagher, A. G., Ritter, E. M., Champion, H., Higgins, G., Fried, M. P., Moses, G., Smith, C. D., & Satava, R. M. (2005). Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training. *Annals of Surgery*, 241(2), 364–372.
- Hammad, A. W. A., Rey, D., Akbarnezhad, A., & Haddad, A. (2023). Integrated mathematical optimisation approach for the tower crane hook routing problem to satisfy material demand requests on-site. *Advanced Engineering Informatics*.
- Hwang, S. (2012). Ultra-wide band technology experiments for real-time prevention of tower crane collisions. *Automation in Construction*, 22, 545–553.
- Kaplan, A. D., Cruick, J., Endsley, M., Beers, S. M., Sawyer, B. D., & Hancock, P. A. (2021). The effects of virtual reality, augmented reality, and mixed reality as training enhancement methods: A meta-analysis. *Human Factors*, 63(4), 706–726.
- Kopp, P. M., Senner, V., & Gröpel, P. (2020). Regular exercise participation and volitional competencies. *Sport, Exercise, and Performance Psychology*, 9(2), 232.
- Li, X., Yi, W., Chi, H.-L., Wang, X., & Chan, A. P. C. (2018). A critical review of virtual and augmented reality (VR/AR) applications in construction safety. *Automation in Construction*, 86, 150–162.
- Ma, S., Li, Z., Li, L., Zhang, S., & Zheng, R. (2024). Deciphering the Key Characteristics of On-Site Industrialized Construction: Inspiration from China. *Journal of Construction Engineering and Management*, 150(4), 5024003.
- Masiello, I., Herault, R., Mansfeld, M., & Skogqvist, M. (2022). Simulation-based VR training for the nuclear sector—a pilot study. *Sustainability*, 14(13), 7984.
- Maurer, T. J., & Weiss, E. M. (2010). Continuous learning skill demands: Associations with managerial job content, age, and experience. *Journal of Business and Psychology*, 25, 1–13.
- Melenbrink, N., Werfel, J., & Menges, A. (2020). On-site autonomous construction robots: Towards unsupervised building. *Automation in Construction*, 119, 103312.
- Moniri, M. M., Valcarcel, F. A. E., Merkel, D., & Sonntag, D. (2016). Human gaze and focus-of-attention in dual reality human-robot collaboration. *2016 12th International Conference on Intelligent Environments (IE)*, 238–241.
- Moon, S., Becerik-Gerber, B., & Soibelman, L. (2019). Virtual learning for workers in robot deployed construction sites. *Advances in Informatics and Computing in Civil and Construction Engineering: Proceedings of the 35th CIB W78 2018 Conference: IT in Design, Construction, and Management*, 889–895.
- Nair, A. J., Manohar, S., Mittal, A., & Chaudhry, R. (2024). Unleashing Digital Frontiers: Bridging Realities of Augmented Reality, Virtual Reality, and the Metaverse. In *The Metaverse Dilemma: Challenges and Opportunities for Business and Society* (pp. 85–112). Emerald Publishing Limited.
- Ni, Z., Cai, S., & Ni, C. (2024). Construction Safety Risk Assessment and Cause Analysis for High-Cable Tower Cranes. *Engineering Proceedings*, 55(1), 96.
- Oagaz, H., Schoun, B., & Choi, M.-H. (2021). Performance improvement and skill transfer in table tennis through training in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 28(12), 4332–4343.
- Sadeghi, H., Zhang, X., & Mohandes, S. R. (2023). Developing an ensemble risk analysis framework for improving the safety of tower crane operations under coupled Fuzzy-based environment. *Safety Science*, 158, 105957.
- Schneider, D., Otte, A., Kublin, A. S., Martschenko, A., Kristensson, P. O., Ofek, E., Pahud, M., & Grubert, J. (2020). Accuracy of commodity finger tracking systems for virtual reality head-mounted displays. *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 804–805.
- Shringi, A., Arashpour, M., Dwyer, T., Prouzeau, A., & Li, H. (2023). Safety in off-site construction: simulation of crane-lifting operations using VR and BIM. *Journal of Architectural Engineering*, 29(1), 4022035.

- Shringi, A., Arashpour, M., Golafshani, E. M., Rajabifard, A., Dwyer, T., & Li, H. (2022). Efficiency of VR-based safety training for construction equipment: Hazard recognition in heavy machinery operations. *Buildings*, 12(12), 2084.
- Song, H., Kim, T., Kim, J., Ahn, D., & Kang, Y. (2021). Effectiveness of VR crane training with head-mounted display: Double mediation of presence and perceived usefulness. *Automation in Construction*, 122, 103506.
- Steuer, J. (1992). Defining Virtual Reality: Dimensions Determining Telepresence. *Journal of Communication*, 42(4), 73–93. doi: <https://doi.org/10.1111/j.1460-2466.1992.tb00812.x>
- Sudiarno, A., Dewi, R. S., Widyaningrum, R., Ma'arij, A. M. D., & Supriatna, A. Y. (2024). Investigating the Future Study Area on VR Technology Implementation in Safety Training: A Systematic Literature Review. *Journal of Safety Science and Resilience*.
- Sutherland, I. E. (1968). A head-mounted three dimensional display. *Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I*, 757–764.
- Tak, A. N., Taghaddos, H., Mousaei, A., Bolourani, A., & Hermann, U. (2021). BIM-based 4D mobile crane simulation and onsite operation management. *Automation in Construction*, 128, 103766.
- Wang, Z., Huang, C., Yao, B., & Li, X. (2024). Integrated reinforcement and imitation learning for tower crane lift path planning. *Automation in Construction*, 165, 105568.
- Weichert, F., Bachmann, D., Rudak, B., & Fisseler, D. (2013). Analysis of the accuracy and robustness of the leap motion controller. *Sensors*, 13(5), 6380–6393.
- Wiethorn, J. D. (2018). *An analytical study of critical factors of lift planning to improve crane safety based on forensic case studies of crane accidents*. The University of Texas at Austin.
- Xie, B., Liu, H., Alghofaili, R., Zhang, Y., Jiang, Y., Lobo, F. D., Li, C., Li, W., Huang, H., & Akdere, M. (2021). A review on virtual reality skill training applications. *Frontiers in Virtual Reality*, 2, 645153.
- Yang, F., & Goh, Y. M. (2022). VR and MR technology for safety management education: An authentic learning approach. *Safety Science*, 148.
- Ye, G., Xiang, Q., Yang, L., Yang, J., Xia, N., Liu, Y., & He, T. (2022). Safety stressors and construction workers' safety performance: The mediating role of ego depletion and self-efficacy. *Frontiers in Psychology*, 12, 818955.
- Zhang, W., Zhu, S., Zhang, X., & Zhao, T. (2020). Identification of critical causes of construction accidents in China using a model based on system thinking and case analysis. *Safety Science*, 121, 606–618.
- Zhang, Z., Pan, W., & Pan, M. (2022). Critical considerations on tower crane layout planning for high-rise modular integrated construction. *Engineering, Construction and Architectural Management*, 29(7), 2615–2634.
- Zhou, T., Zhu, Q., Shi, Y., & Du, J. (2022). Construction robot teleoperation safeguard based on real-time human hand motion prediction. *Journal of Construction Engineering and Management*, 148(7), 4022040.
- Zhou, W., Zhao, T., Liu, W., & Tang, J. (2018). Tower crane safety on construction sites: A complex sociotechnical system perspective. *Safety Science*, 109, 95–108.

Appendix 1

Table Self-efficacy scale

The following 10 sentences pertain to your general perceptions of your ability to operate robots. Please mark the appropriate box on the right with a "✓" based on your actual situation (how you truly feel). There are no right or wrong answers, and you do not need to spend much time on each sentence.

	Completely Disagree	Somewhat Disagree	Somewh at Agree	Completely Agree
1. If I make an effort, I can always solve problems.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Even if others oppose me, I can still manage to complete robot operation tasks.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. For me, controlling the robot and achieving construction goals is easy.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. I am confident that I can effectively handle any unexpected events.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. With my intelligence, I can definitely cope with unforeseen situations.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. If I put in the necessary effort, I can solve most problems.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. I can face difficulties calmly because I trust in my ability to handle problems.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. When facing a problem, I can usually find several solutions.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. When trouble arises, I can usually think of some ways to cope with it.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. No matter what happens to me, I can handle it with ease.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

CYBERSECURITY-AWARE DECENTRALIZED MACHINE LEARNING FRAMEWORK FOR CONSTRUCTION EQUIPMENT MOTION RECOGNITION USING BLOCKCHAIN

Chengliang Zheng¹, Xingyu Tao², Jiarui Lin³, Moumita Das⁴, Wenchi Shou⁵, Jack C.P. Cheng^{6*}

1) Ph.D. Candidate, Key Laboratory of Aerospace Information Security and Trusted Computing, Ministry of Education, School of Cyber Science and Engineering, Wuhan University, Wuhan, China. E-mail: chengliang@ust.hk

2) Research Assistant Professor, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR. Email: xtaoab@connect.ust.hk

3) Associate Professor, Department of Civil Engineering, Tsinghua University, Beijing, China. lin611@tsinghua.edu.cn

4) Postdoctoral Fellow, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong. E-mail: moumitadas@ust.hk

5) Senior Lecturer, School of Engineering, Design & Built Environment, Western Sydney University, Sydney, Australia. E-mail: W.Shou@westernsydney.edu.au

6) Professor, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong. E-mail: cejcheng@ust.hk

Abstract: Artificial intelligence (AI) is playing an increasing role in the construction industry to enhance productivity, reduce safety accidents, and optimize collaboration efficiency. However, attacks on AI systems also introduce cybersecurity threats that could lead to severe consequences, such as equipment damage, financial loss, operational downtime, safety accidents, and potential loss of life. Motivated by the construction industry's limited efforts to defend against AI cybersecurity vulnerabilities—a result of a lack of awareness and IT resources—this paper aims to propose a cybersecurity-aware decentralized machine learning (CADML) framework to protect the life cycle cybersecurity of machine learning (ML) models leveraging blockchain. First, the workflow of the CADML framework will be introduced to illustrate the logic of blockchain-ML integration. Second, a new blockchain smart contract algorithm, ML-embed smart contract (MLSC), will be developed to train and apply AI in a decentralized manner. The primary innovation framework extends current "partially" blockchain-ML integration methods to enable the ML's "lifecycle" (from raw data storage, training, implementation, to model update) to operate in a decentralized and secure blockchain environment. The framework is tested to recognize construction equipment motions. Results show that (1) the ML model could be successfully trained and implemented within a blockchain and (2) the ML performance (accuracy, precision, and recall) is acceptable.

Keywords: Machine Learning, Blockchain, Smart Contract, Construction Equipment, Motion Recognition

1. INTRODUCTION

Artificial intelligence (AI) (e.g., machine learning (ML)) is transforming the construction sector, which is benefiting AI applications from areas like on-site safety management (Baker et al., 2020), schedule management (Pan et al., 2021), supply chain optimization (Baduge et al., 2022), and the control of construction robotics and drones (Bademosi et al., 2021). However, most AI models in this sector rely on centralized servers controlled by certain project members or third-party vendors for training and implementation, which introduces severe cybersecurity risks (Yazdinejad et al., 2023). These vulnerabilities include the potential for AI model manipulation—where attackers might alter training data or model parameters—and single-point failures that could compromise entire systems. An example of such a manipulation is a tape attack, where small pieces of tape are strategically placed on signs in a way that causes an AI-driven construction vehicle to misread the sign, leading to potential accidents or misrouting (Bansal et al., 2023). Moreover, the opaque nature of AI decision-making processes makes it difficult to detect when a system has been compromised or is repeating errors, thereby increasing the demand for explainable AI.

Integrating blockchain technology with AI systems presents a promising solution to the cybersecurity vulnerabilities inherent in centralized AI architectures (Han et al., 2023). Blockchain is a decentralized digital ledger that records transactions across multiple computers in such a way that the registered transactions cannot be altered retroactively (Tao et al., 2021). This technology is known for its robust security features, which include immutability, transparency, and the elimination of a single point of failure, making it highly resistant to tampering and cyber-attacks (Xu et al., 2023). Successful blockchain implementations in the construction industry include automating payments and enforcing contract terms, reducing disputes and delays (Hamledari et al., 2021). Blockchain also improves supply chain management and collaborations by ensuring the traceability and quality of materials from origin to installation (Tao et al., 2023). Additionally, integration with IoT devices allows for real-time infrastructure monitoring, ensuring compliance and facilitating proactive maintenance (L. Wu et al., 2023). By leveraging blockchain, AI systems in construction can also achieve enhanced security in several ways. First, blockchain can provide a secure platform for sharing AI training data and AI models among multiple stakeholders, reducing the risk of data tampering by ensuring that data alterations are traceable and transparent (Salah et al., 2019). Additionally, blockchain's decentralized nature prevents single-point failures, as the AI system's operational

data and decisions are recorded on a blockchain distributed across many nodes (Adel et al., 2022).

Existing research primarily examines blockchain and AI separately, with most studies still theoretical. There has been limited exploration into integration solutions within construction. H. Wu et al. (2023) tried to improve quality traceability on construction sites using blockchain and computer vision, though only image hashes were stored on-chain, with the AI model and raw data kept off-chain. Similarly, Adel et al. (2022) introduced a decentralized AI system that leverages blockchain technology to calculate construction costs. The study demonstrated that the blockchain safeguarded the integrity and authenticity of the AI models. Moreover, it ensured that all predictions were synchronized across the blockchain, creating a transparent and auditable environment for implementation (Yang et al., 2022). However, the model's training occurred off-chain, indicating only partial integration. This is because current AI-blockchain integrations rely on smart contracts, a blockchain element that can self-execute agreements with predefined rules and conditions written in code (Lu et al., 2021). In the context of construction, smart contracts support basic functions like information exchange, payment processes, and version management (Ciotto et al., 2021). However, integrating computationally demanding construction AI algorithms, such as machine learning or deep learning, poses challenges due to the limited computational capacity of smart contracts, arising from their decentralized setup and consensus mechanisms (Singh et al., 2020). Developing smart contracts that accommodate complex AI algorithms is technically demanding.

Thus, this paper aims to answer how to protect the lifecycle cybersecurity of construction AI in a decentralized blockchain network. Therefore, this paper proposes a Cybersecurity-Aware Decentralized Machine Learning (CADML) framework to enhance cybersecurity across the lifecycle of machine learning (ML) models. Initially, it will introduce the CADML framework's workflow to demonstrate the integration logic between blockchain and ML. Subsequently, a novel blockchain smart contract algorithm will be designed for decentralized training and AI application. This framework advances beyond current partial integrations by facilitating the complete lifecycle of ML models—from data storage and model training to implementation and updates—within a secure, decentralized blockchain environment. The framework's efficacy is tested by recognizing construction equipment movements.

2. CYBERSECURITY-AWARE DECENTRALIZED MACHINE LEARNING (CADML) FRAMEWORK

2.1 Workflow of the CADML framework

Figure 1 shows the workflow of the CADML framework. Figure 1 (a) shows the CADML architecture, in which the raw data are securely placed in the blockchain database, and the smart contract hold the ML algorithm. The training and implementation are operated within the smart contract so that all authorized project members can maintain the lifecycle security of AI.

The detailed ML workflow in the smart contract is shown in Figure (b) with six steps: (1) Data Acquisition and Preprocessing: Sensor-generated data is continuously transmitted to a processing platform at regular intervals, known as time windows (TW). This data is preprocessed to reduce noise, lower dimensionality, and filter irrelevant information. It is then stored on the blockchain to guarantee its immutability. (2) AI Model Training: Smart contracts access the training data stored on the blockchain and perform security checks before formatting the data to suit model training requirements. If the AI model is actively being trained, it uses this data to update itself from version n to version $n+1$. (3) Model State Evaluation: Post training, the updated model and key metrics, such as accuracy, are reviewed by a controller to determine if the model is ready for deployment. It transitions from the "training" to the "application" phase if it meets the required accuracy. (4) Model Application: When new data, such as excavator poses, is received, it undergoes security checks and formatting by smart contracts before being fed into the AI model. (5) Continuous Model Monitoring: Smart contracts continuously monitor the model's performance during application. Changes in the site conditions or sensor alterations that affect data characteristics may degrade the model's effectiveness, necessitating a return to the training phase if the pose recognition accuracy falls below acceptable levels. (6) Result Documentation: If the model remains accurate and functional, the application results, including model hash, timestamp, and recognition outcomes (e.g., excavator poses), are recorded on the blockchain for transparency and traceability.

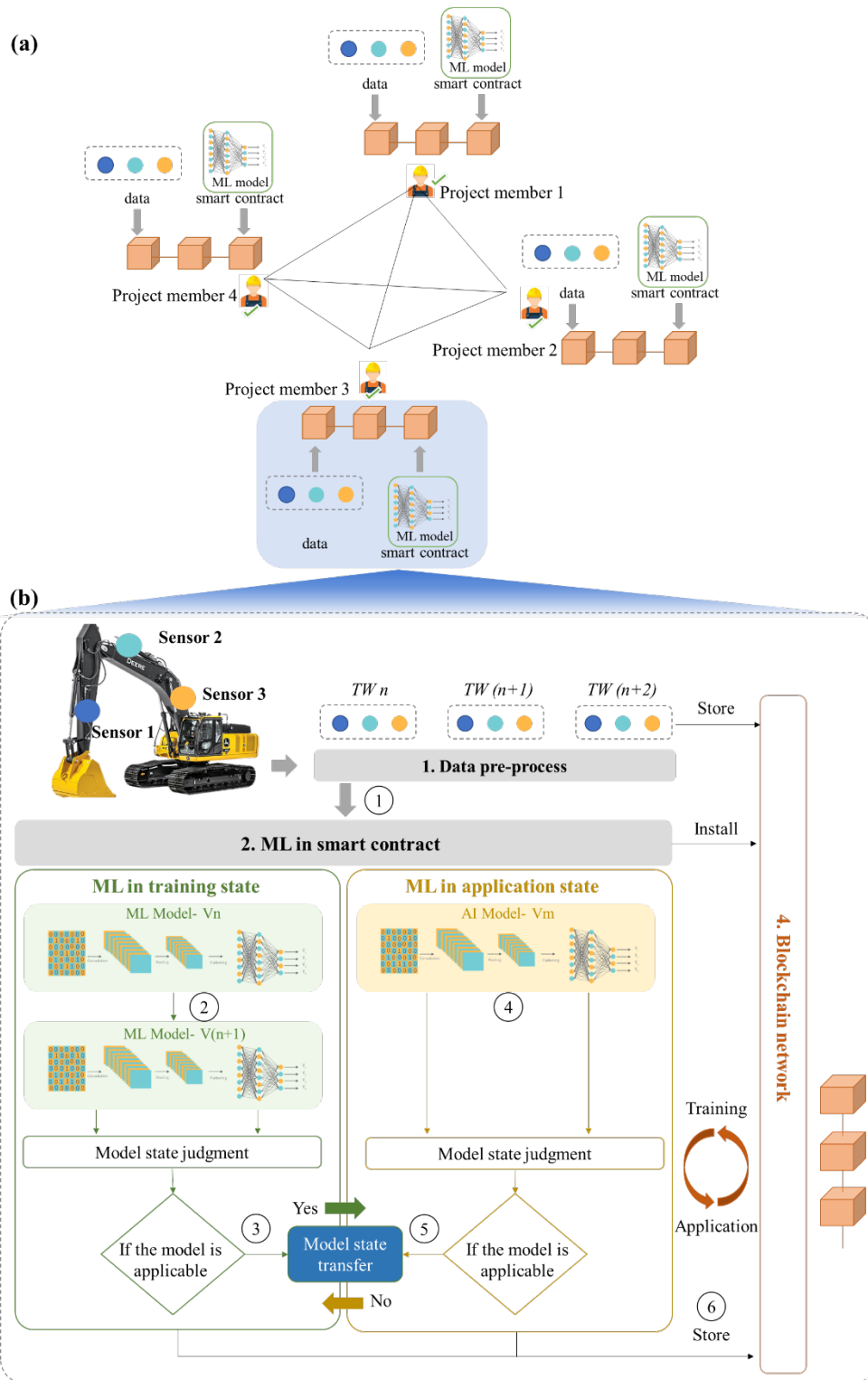


Figure 1. CADML framework

2.2 ML-embed smart contract (MLSC) algorithm

Figure 1 shows the MLESC algorithm. Unlike existing smart that can only process simple computing, the MLSC includes kernel and user spaces. The kernel space supports developers with essential frameworks and common algorithms for AI model development, integrating two key AI/ML frameworks: Massive Online Analysis (MOA) (Li et al., 2018) and Waikato Environment for Knowledge Analysis (WEKA) (Daw et al., 2020). MOA processes large data streams for real-time analysis, while WEKA offers robust data processing and model conversion tools. The user space in smart contracts is designated for training and utilizing AI models, structured into six essential layers: (1) Security Check Layer, which ensures data integrity by hashing to block tampered data, (2) Data Adaptation Layer, tasked with modifying data to suit specific algorithm requirements, (3) Algorithm

Model Layer, the central component where training and application codes reside, (4) Feedback Layer, which communicates model state information either externally or to the Controller Layer, (5) Output Layer, which logs model data onto the blockchain, and (6) Controller Layer, which switches the AI model between training and application phases. These layers collectively facilitate AI's secure and efficient operation within smart contracts.

In construction pose recognition, sensor data is initially processed by the MLSC's Security Check Layer. The data then moves to the Data Adaptation Layer, where it's converted from string to Instance type, aligning with model requirements. The algorithm model layer uses this data for training or pose classification, which the controller layer controls. Model performance metrics like accuracy are relayed to project managers via the Feedback Layer in real-time. Finally, essential data and results are recorded on the blockchain by the Output Layer. The source code is available at <https://github.com/iEricZHENG/BlockchainAI> (Zheng et al., 2024).

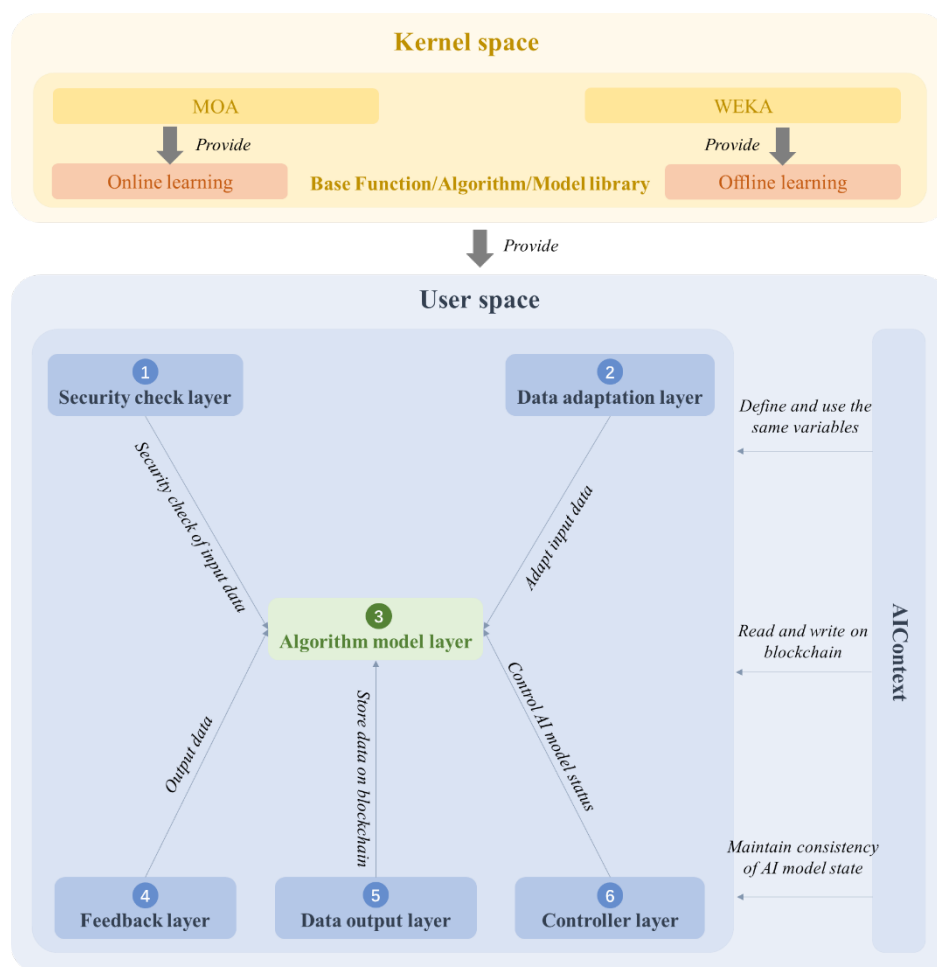


Figure 2. MLSC algorithm

3. RESULTS

3.1 Validation preparation

This paper validates the CADML framework by recognizing the motion of an excavator, followed by the validation environment and preparation.

Blockchain platform. Hyperledger Fabric (HF) was selected for the CADML framework because of its strong privacy protections, customizable development features, and proven feasibility in construction scenarios (Androulaki et al., 2018). The blockchain experiment utilizes HF version 2.3.0 with two organizations.

Data collection. The inertial measurement unit (IMU), consisting of an accelerometer and a gyroscope, was used in this experiment to track the spatial movements of independent components for excavators. The time-series data collected by the sensors includes acceleration and angular rate.

Machine learning selection. ROSE (Robust Online Self-Adjusting Ensemble) (Cano et al., 2022) is a dynamic ensemble learning algorithm suitable for classifying tasks in data stream environments like excavator pose recognition. It effectively handles changing, imbalanced data and concept drift, adapting continuously to new inputs. This robustness addresses class imbalance, drift, and noise, ensuring high accuracy in real-time applications.

3.2 Results of CADML framework validation

Figure 3 (a) is the sensor installation. IMU sensors were affixed to the boom, arm, bucket, and cabin of the equipment. The appropriateness and effectiveness of the sensor placement and type were confirmed by Tang et al. (2023). Figure 3 (b) illustrates the training outcomes of the Machine Learning-Enabled Smart Contract (MLES). Specifically, Fig. 8(a) details the iterations of algorithm training recorded on the blockchain, while Fig. 8(b) displays a training record for version 2.0.0 of the ML algorithm. This record includes three components: (1) ML hash, where the blockchain calculates and logs the hash value to ensure each algorithm version is verifiable and secure. (2) Input data, documenting all raw training data, such as IMU sensor readings for x, y, and z-axis accelerations, along with the corresponding excavator posture. Three poses are identified: "0" for stationary, "1" for engine vibrations, and "2" for boom motion. (3) Timestamp, marking the completion of each training session, which helps in assessing the training efficiency of the ML algorithm. Figure 3 (c) demonstrates that the MLSC algorithm effectively identifies the excavator's boom motion, classified as state "2."

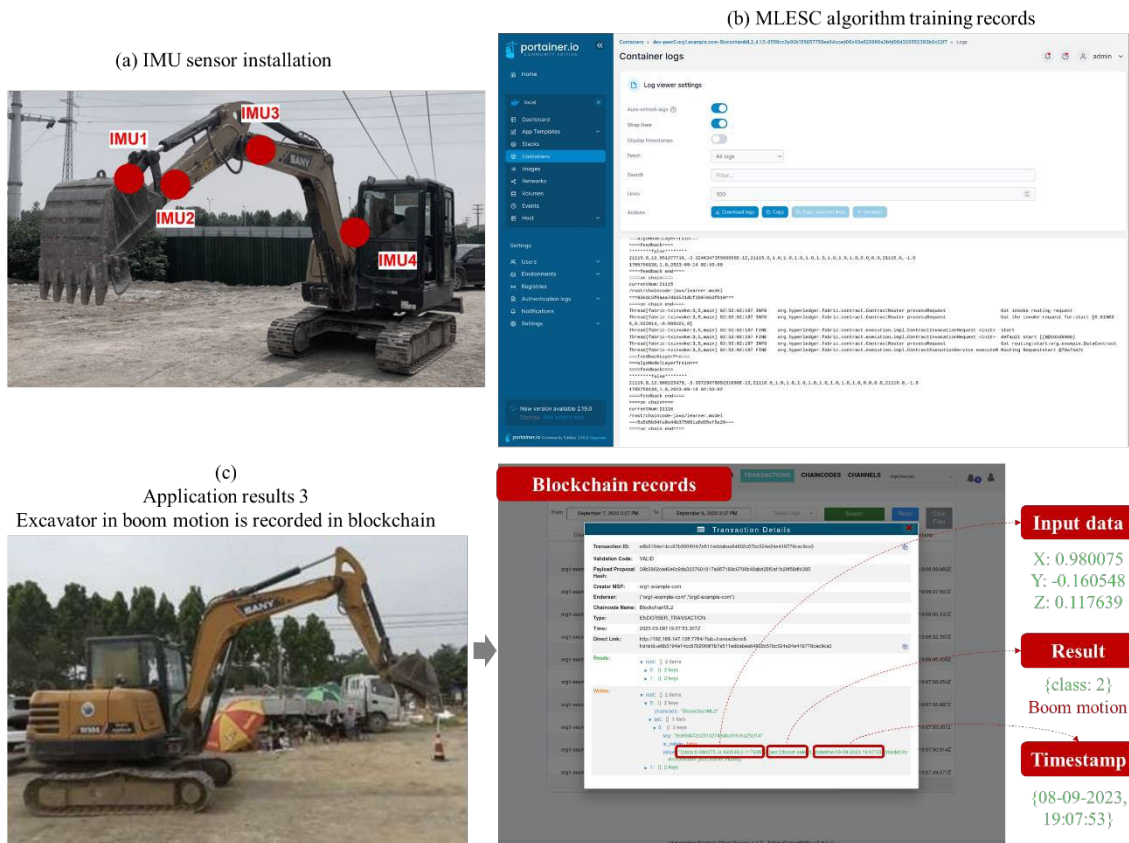


Figure 3. The CADML framework is successfully validated in recognizing excavator motion

This study evaluates the excavator pose recognition model using accuracy, precision, and recall. Accuracy, a common metric for classification models, measures the likelihood that the model accurately identifies the excavator's state. However, its utility diminishes with imbalanced sample classes, potentially skewing the perceived effectiveness of the model. This issue arises when one class predominates, causing the model to excel at recognizing that specific pose but falter with others. Hence, precision and recall are also employed to provide a more rounded assessment. Precision represents the ratio of correctly identified positive predictions (e.g., truly stationary excavators classified as stationary), minimizing the risk of false positives. Meanwhile, recall quantifies the correct positive predictions against all actual positives, which is crucial for ensuring the model reliably detects stationary states without misclassifying them. Together, these metrics offer a comprehensive view of the model's performance, addressing both its strengths and limitations in differentiating excavator poses.

Figure 4 displays the performance of the three metrics, starting from zero and rising with the increase in training data. As the data approaches 3,000 entries, accuracy achieves approximately 95%, precision hits around 94%, and recall reaches about 96%, after which each metric stabilizes. According to (Slaton et al., 2020; Tang et al., 2022), the acceptable thresholds for machine pose recognition are set at 85% for average accuracy, 88% for precision, and 87.6% for recall. Given these benchmarks in our results, which exceed these thresholds, are deemed satisfactory.

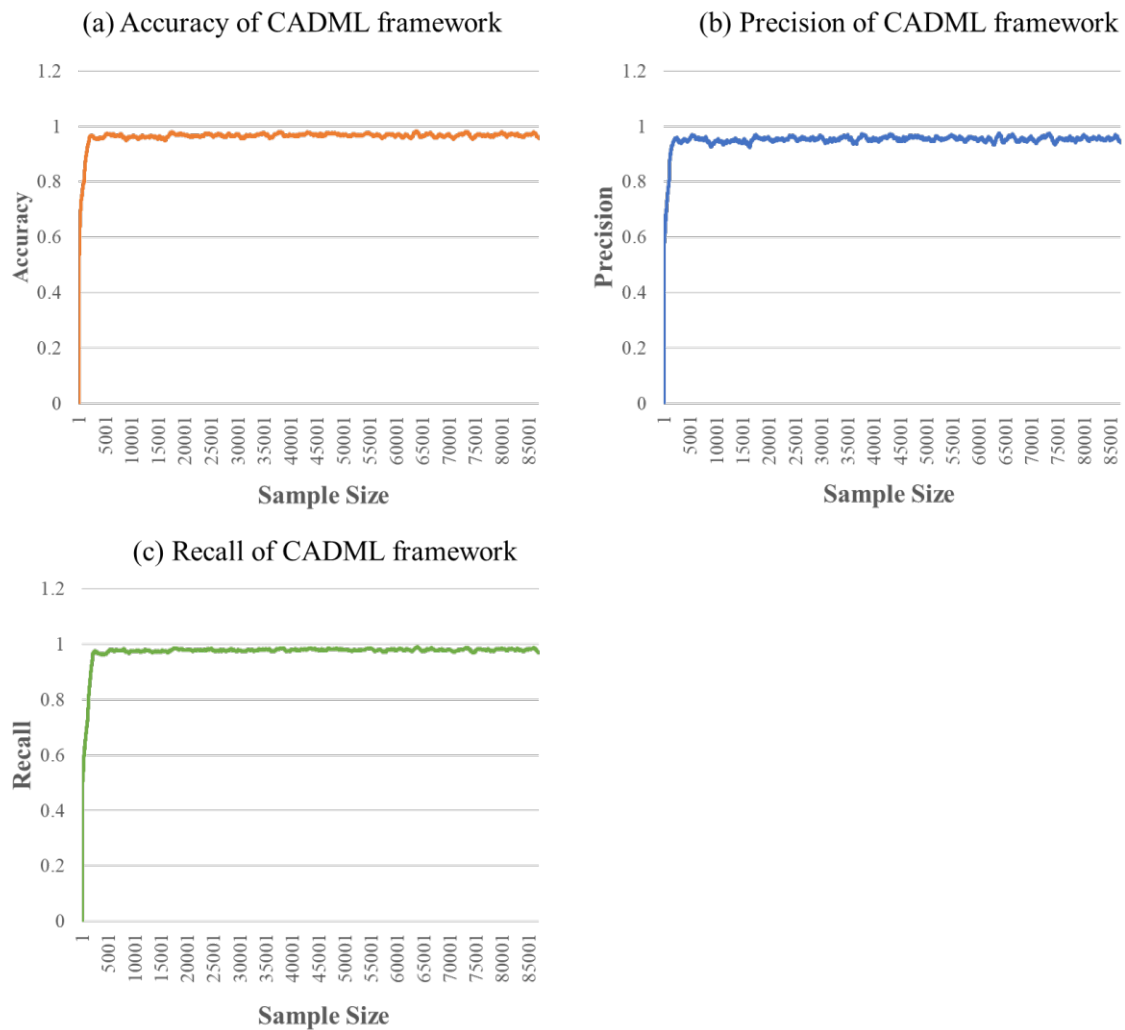


Figure 4. Computing performance of the CADML framework

4. CONCLUSIONS

Integrating AI technologies in the construction industry brings cybersecurity concerns such as data tampering, opacity in AI operations, and vulnerabilities from centralized control points. Although some initial studies have converged blockchain technology, a decentralized database technology that secures data immutability, traceability, and robustness, with AI, they only guaranteed "partial" security, as they stored part of AI data (e.g., raw data or trained AI models) to blockchain. This paper advances existing works and introduces a cybersecurity-aware decentralized machine learning (CADML) framework to bolster the "lifecycle" cybersecurity of AI data in construction by leveraging blockchain's. Two research objectives have been achieved. First, the mechanism of how an AI model is trained and implemented within a decentralized blockchain is illustrated via a CADML framework workflow. Second, an MLSC algorithm is developed within the framework to enable the lifecycle AI model management. The framework is successfully validated in the scenarios of excavator motion recognition. Moreover, computing performance, including accuracy, precision, and recall, is acceptable. This CADML framework addresses the cybersecurity challenges faced by construction AI and enhances the robustness and reliability of AI applications in project management contexts, opening the door to encourage more stakeholders to contribute AI data and computing sources.

This work is an initial exploration of combining construction AI with blockchain. Two limitations still exist. The first challenge is slower AI training speeds due to blockchain's processing limitations, which hinders timely deployment. Future research could optimize blockchain architectures or develop efficient consensus algorithms to quicken AI training while maintaining security. Additionally, the complexity of configuring blockchain systems poses a barrier to widespread adoption. Simplifying blockchain interfaces and increasing educational efforts to boost blockchain comprehension among construction professionals could mitigate this issue.

5. DISCUSSION

- **DAI Contribution to Construction and Building Operations:** The CADA architecture can significantly

enhance construction management and building operations by integrating machine learning models directly into smart contracts. The unique aspect of CADAi development in construction management is its ability to create tailored models that reflect specific project needs, allowing for continuous learning and adaptation.

- **Collaborative AI Development for LLMs in Construction:** The collaborative AI development method enabled by the CADAi architecture contributes to the development of large language models (LLMs) tailored for the construction industry. By allowing multiple stakeholders, such as architects, engineers, and construction workers, to contribute their unique datasets through smart contracts, this architecture fosters a rich and diverse training environment. For instance, an LLM could be trained on data from various project phases, providing insights into best practices, safety protocols, and regulatory compliance. This collaborative approach ensures that LLMs are not only robust but also reflective of the industry's diverse needs, ultimately leading to more effective AI applications in construction.
- **Enhancing Data Sharing in the Construction Industry:** The CADAi architecture has significant potential to enhance data sharing related to AI in the construction industry. By utilizing smart contracts, stakeholders can securely share data while maintaining control over their contributions. For example, subcontractors can share their performance metrics, while suppliers can provide data on material usage and quality. This data can be aggregated to train machine learning models that improve project outcomes, such as predicting equipment failures or optimizing supply chain logistics. The decentralized nature of this architecture encourages a culture of transparency and collaboration, allowing for a more interconnected construction ecosystem where insights derived from shared data lead to innovation and efficiency.

REFERENCES

- Adel, K., Elhakeem, A., & Marzouk, M. (2022). Decentralizing construction AI applications using blockchain technology. *Expert Systems with Applications*, 194, 116548. doi:<https://doi.org/10.1016/j.eswa.2022.116548>
- Androulaki, E., Barger, A., Bortnikov, V., Cachin, C., Christidis, K., Caro, A. D., . . . Yellick, J. (2018). *Hyperledger fabric: A distributed operating system for permissioned blockchains*. Paper presented at the The 13th EuroSys Conference, Porto, Portugal.
- Bademosi, F., & Issa Raja, R. A. (2021). Factors Influencing Adoption and Integration of Construction Robotics and Automation Technology in the US. *Journal of Construction Engineering and Management*, 147(8), 04021075. doi:[https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0002103](https://doi.org/10.1061/(ASCE)CO.1943-7862.0002103)
- Baduge, S. K., Thilakarathna, S., Perera, J. S., Arashpour, M., Sharafi, P., Teodosio, B., . . . Mendis, P. (2022). Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications. *Automation in Construction*, 141, 104440. doi:<https://doi.org/10.1016/j.autcon.2022.104440>
- Baker, H., Hallowell, M. R., & Tixier, A. J. P. (2020). AI-based prediction of independent construction safety outcomes from universal attributes. *Automation in Construction*, 118, 103146. doi:<https://doi.org/10.1016/j.autcon.2020.103146>
- Bansal, V., Bhardwaj, A., Singh, J., Verma, D., Tiwari, M., & Siddi, S. (2023, 12-13 May 2023). *Using Artificial Intelligence to Integrate Machine Learning, Fuzzy Logic, and The IOT as A Cybersecurity System*. Paper presented at the 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE).
- Cano, A., & Krawczyk, B. (2022). ROSE: robust online self-adjusting ensemble for continual learning on imbalanced drifting data streams. *Machine Learning*, 111(7), 2561-2599. doi:<https://doi.org/10.1007/s10994-022-06168-x>
- Ciotta, V., Mariniello, G., Asprone, D., Botta, A., & Manfredi, G. (2021). Integration of blockchains and smart contracts into construction information flows: Proof-of-concept. *Automation in Construction*, 132, 103925. doi:<https://doi.org/10.1016/j.autcon.2021.103925>
- Daw, S., & Basak, R. (2020, 11-13 March 2020). *Machine Learning Applications Using Waikato Environment for Knowledge Analysis*. Paper presented at the 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC).
- Hamledari, H., & Fischer, M. (2021). Construction payment automation using blockchain-enabled smart contracts and robotic reality capture technologies. *Automation in Construction*, 132, 103926. doi:<https://doi.org/10.1016/j.autcon.2021.103926>
- Han, H., Shiwakoti, R. K., Jarvis, R., Mordi, C., & Botchie, D. (2023). Accounting and auditing with blockchain technology and artificial Intelligence: A literature review. *International Journal of Accounting Information Systems*, 48, 100598. doi:<https://doi.org/10.1016/j.accinf.2022.100598>
- Li, C., & Zhou, H. (2018). Enhancing the Efficiency of Massive Online Learning by Integrating Intelligent Analysis into MOOCs with an Application to Education of Sustainability. 10(2), 468.
- Lu, W., Li, X., Xue, F., Zhao, R., Wu, L., & Yeh, A. G. O. (2021). Exploring smart construction objects as

- blockchain oracles in construction supply chain management. *Automation in Construction*, 129, 103816. doi:<https://doi.org/10.1016/j.autcon.2021.103816>
- Pan, Y., & Zhang, L. (2021). Roles of artificial intelligence in construction engineering and management: A critical review and future trends. *Automation in Construction*, 122, 103517. doi:<https://doi.org/10.1016/j.autcon.2020.103517>
- Salah, K., Rehman, M. H. U., Nizamuddin, N., & Al-Fuqaha, A. (2019). Blockchain for AI: Review and Open Research Challenges. *IEEE Access*, 7, 10127-10149. doi:<https://doi.org/10.1109/ACCESS.2018.2890507>
- Singh, S., Sharma, P. K., Yoon, B., Shojafar, M., Cho, G. H., & Ra, I.-H. (2020). Convergence of blockchain and artificial intelligence in IoT network for the sustainable smart city. *Sustainable Cities and Society*, 63, 102364. doi:<https://doi.org/10.1016/j.scs.2020.102364>
- Slaton, T., Hernandez, C., & Akhavian, R. (2020). Construction activity recognition with convolutional recurrent networks. *Automation in Construction*, 113, 103138. doi:<https://doi.org/10.1016/j.autcon.2020.103138>
- Tang, J., Luo, H., Chen, W., Wong, P. K.-Y., & Cheng, J. C. P. (2022). IMU-based full-body pose estimation for construction machines using kinematics modeling. *Automation in Construction*, 138, 104217. doi:<https://doi.org/10.1016/j.autcon.2022.104217>
- Tang, J., Wang, M., Luo, H., Wong, P. K.-Y., Zhang, X., Chen, W., & Cheng, J. C. J. A. i. C. (2023). Full-body pose estimation for excavators based on data fusion of multiple onboard sensors. 147, 104694.
- Tao, X., Das, M., Liu, Y., & Cheng, J. C. P. (2021). Distributed common data environment using blockchain and Interplanetary File System for secure BIM-based collaborative design. *Automation in Construction*, 130, 103851. doi:<https://doi.org/10.1016/j.autcon.2021.103851>
- Tao, X., Das, M., Zheng, C., Liu, Y., Wong, P. K.-Y., Xu, Y., . . . Cheng, J. C. P. (2023). Enhancing BIM security in emergency construction projects using lightweight blockchain-as-a-service. *Automation in Construction*, 150, 104846. doi:<https://doi.org/10.1016/j.autcon.2023.104846>
- Wu, H., Li, H., Luo, X., & Jiang, S. (2023). Blockchain-Based Onsite Activity Management for Smart Construction Process Quality Traceability. *IEEE Internet of Things Journal*, 10(24), 21554-21565. doi:<https://doi.org/10.1109/JIOT.2023.3300076>
- Wu, L., Lu, W., & Chen, C. (2023). Resolving power imbalances in construction payment using blockchain smart contracts. *Engineering, Construction and Architectural Management*(ahead-of-print). doi:<https://doi.org/10.1108/ECAM-03-2023-0194>
- Xu, Y., Tao, X., Das, M., Kwok, H. H. L., Liu, H., Wang, G., & Cheng, J. C. P. (2023). Suitability analysis of consensus protocols for blockchain-based applications in the construction industry. *Automation in Construction*, 145, 104638. doi:<https://doi.org/10.1016/j.autcon.2022.104638>
- Yang, Q., Zhao, Y., Huang, H., Xiong, Z., Kang, J., & Zheng, Z. (2022). Fusing Blockchain and AI With Metaverse: A Survey. *IEEE Open Journal of the Computer Society*, 3, 122-136. doi:<https://doi.org/10.1109/OJCS.2022.3188249>
- Yazdinejad, A., Dehghantanha, A., Parizi, R. M., Srivastava, G., & Karimipour, H. (2023). Secure Intelligent Fuzzy Blockchain Framework: Effective Threat Detection in IoT Networks. *Computers in Industry*, 144, 103801. doi:<https://doi.org/10.1016/j.compind.2022.103801>
- Zheng, C., Tao, X., Dong, L., Zukaib, U., Tang, J., Zhou, H., . . . Shen, Z. (2024). Decentralized artificial intelligence in construction using blockchain. *Automation in Construction*, 166, 105669. doi:<https://doi.org/10.1016/j.autcon.2024.105669>

BUILDING DECAY IN HONG KONG: ASSESSING AND MAPPING BUILDING CONDITION INDEX USING CHATGPT AND AIRBORNE POINT CLOUDS

Wai Hung LEE¹, and Fan XUE²

1) B.Sc. Student, Department of Real Estate and Construction, The University of Hong Kong, Pokfulam, Hong Kong SAR.
Email: sammylee@connect.hku.hk

2) Ph.D., Assoc. Professor., Department of Real Estate and Construction, The University of Hong Kong, Pokfulam, Hong Kong.
Email: xuef@hku.hk

Abstract: Hong Kong faces critical challenges in the maintenance and redevelopment of aged buildings. Recently, advancements in multi-modal generative AI (GenAI) and high-definition urban geospatial data, such as point clouds, have offered new opportunities to the architectural, engineering, and construction industry. This paper defines, assesses, and maps a Building Condition Index (BCI) as the condition of aged building fabrics using GenAI and high-definition geospatial data. First, a BCI is defined as a numerical scale of multi-dimensional factors, including floor area, building age, management quality, and the presence of unauthorized building works. Then, multiple data sources, including building exterior photos, airborne point clouds, and government building datasets, are processed and trained for the BCI using multiple regression and image embedding with ChatGPT4. Finally, a comprehensive BCI map and focused BCI hot spots can be visualized for an urban area. Experiments with over 1,200 building data points in Kowloon City, Hong Kong, indicated the robustness of the BCI in explaining the exogenous factors causing decayed buildings while accurately reflecting the building condition of buildings.

Keywords: Building decay, Building condition index, Assessment and mapping, Airborne point clouds, Generative AI

1. INTRODUCTION

Hong Kong faces a critical impasse in addressing the challenges of aging buildings and their redevelopment, intensifying urban decay, a process characterized by the gradual decline of urban areas due to neglect and insufficient investment in the maintenance of old buildings. As of 2021, around 22% of Hong Kong's buildings, approximately 9,600 structures, are over 50 years old and increasingly unfit for habitation (BD, 2023a). Many of these are "three-nil" buildings, which lack owners' corporations, residents' organizations, or property management companies, leading to inadequate maintenance and rapid deterioration (Lee & Chan, 2008). Redevelopment efforts under Cap. 545 are further hampered by fragmented ownership structures that require developers to secure over 70% of undivided shares, a challenging and often prohibitive task in an uncertain economic environment (Wang et al., 2022). Even when shares are acquired, economic uncertainties often cause developers to delay redevelopment, leaving many buildings in disrepair.

Building maintenance is crucial for preserving the structural integrity, functionality, and aesthetic appeal of buildings, thereby ensuring safety and regulatory compliance (Seeley, 1987). This involves addressing various challenges such as wear and tear of building fabrics and building services issues like water seepage or structural issues. In Hong Kong's context of aging infrastructure, a range of maintenance strategies are employed to mitigate deterioration. Preventive maintenance includes regular inspections and timely repairs to prevent major failures, while ad-hoc maintenance responds reactively to defects as they arise (Chanter, 2008). Additionally, planned maintenance schedules activities based on thorough assessments of building conditions, ensuring a proactive approach to upkeep (Jasiulewicz-Kaczmarek, 2016). Policies such as Hong Kong's Mandatory Building Inspection Scheme (MBIS) underscore the importance of routine inspections in maintaining older structures, where buildings aged over 30 years old are required to be inspected every 10 years (BD, 2023b).

However, challenges persist in the maintenance of aged buildings in Hong Kong. First, fragmented ownership in these buildings complicates coordination among multiple stakeholders, making it difficult to implement consistent maintenance practices. Financial constraints are significant, as the high costs associated with repairing and upgrading older buildings often discourage property owners, particularly in Hong Kong, where property management companies assist in procurement processes. The complexities of managing multiple ownerships can lead to conflicts among owners, complicating decision-making for necessary repairs and renovations. This situation is exacerbated by instances of bid-rigging, where contractors collude to inflate prices or manipulate the bidding process. For example, in the Garden Vista case, a contractor was jailed for offering bribes to facilitate a bid-rigging scheme that resulted in a renovation contract worth HK\$260 million (Leung et al., 2022). Worse still, the prevalence of "three-nil" buildings leads to inadequate maintenance and accelerated decay due to the absence of coordinated upkeep efforts (Ho et al., 2011; Hui et al., 2008). Moreover, the aging population residing in these buildings often lacks the resources or knowledge to advocate for proper maintenance, compounding the neglect issue (Hui et al., 2008). These multifaceted challenges create a complex environment that hinders the effective maintenance of Hong Kong's aging building stock, perpetuating urban decay and posing

risks to public safety and the overall livability of the city.

AI technological advancements have significantly enhanced building maintenance strategies. For instance, the integration of 2D imaging embedding with deep machine learning into building inspection has gained traction in the industry for detecting building defects. Studies have shown that image embedding techniques can effectively identify issues such as cracks and spalling by analyzing visual data through deep learning algorithms (Jiang et al., 2021). This growing emphasis on structured maintenance is essential in addressing the challenges posed by aging buildings, thereby supporting efforts to combat urban decay and improve overall urban livability. With image embedding and other sensing and processing technologies, building surveyors are able to deliver more systematic and efficient approaches to building upkeep.

Recently, advancements in multi-modal generative AI (GenAI), exemplified by the latest iterations of ChatGPT, offer new opportunities across various sectors by enhancing content creation, automating routine tasks, and facilitating more natural human-computer interactions (García-Peñalvo & Vázquez-Ingelmo, 2023). These GenAI systems can analyze and generate text, images, and sometimes audio. Hence, it seems that this provides a new opportunity for building maintenance, not only to conduct defect diagnosis but also to automatically provide feedback on their presence, provide herewith scheduled maintenance activities, predict potential failures, and provide real-time updates to residents, thus increasing efficiency and responsiveness. However, despite their potential, GenAI models exhibit notable limitations. They often produce hallucinations, generating information that appears plausible but is factually incorrect or nonsensical, which can lead to misinformation in maintenance reports (Salvagno et al., 2023). Additionally, GenAI struggles with understanding complex contexts and nuances, resulting in responses that may not align with user intentions or the specific requirements of building maintenance tasks (Chiu, 2023).

Another opportunity comes from the emergence of high-definition geospatial data, which presents significant opportunities for enhancing building maintenance and management practices. One of the most promising advancements in this field is the use of point cloud technology, which captures detailed three-dimensional data of buildings through methods such as laser scanning and photogrammetry. This technology generates a dense collection of points that represent the physical features of a structure, allowing for precise modeling and analysis of building fabrics on a larger scale. By segmenting point clouds, in-depth building characteristics can be identified, which presents new opportunities for aiding building inspections, monitoring structural integrity, and identifying maintenance needs with greater accuracy and efficiency. Furthermore, geospatial data can facilitate planned maintenance by analyzing spatial patterns under City Information Modelling (CIM) (Xue et al. 2021). For instance, integrating point cloud data with geographic information systems (GIS) enables the visualization of building conditions in relation to environmental factors, helping to prioritize maintenance efforts effectively.

In response to the aforementioned opportunity to promote the effectiveness of building maintenance, this paper introduces, assesses, and maps a Building Condition Index (BCI) as the condition of aged building fabrics using GenAI and high-definition airborne point clouds.

2. RESEARCH METHODS

2.1 Definition of Building Condition Index

The BCI in this paper is a numerical scale employed to assess the overall condition of a building's fabrics, typically ranging from 20 (very poor) to 100 (excellent). Equation (1) defines BCI on the impact of property management on building quality:

$$\begin{aligned} \ln(\text{BCI}) = & C + \beta_1(\text{BUILDING_AGE}) + \beta_2(\text{BUILDING_AGE}^2) + \beta_3(\text{GFA}) + \beta_4(\text{GFA}^2) \\ & + \beta_5(\text{FLOOR}) + \beta_6(\text{FLOOR}^2) \\ & + \beta_7(\text{MANAGEMENT_COMPANY}) + \beta_8(\text{MANAGEMENT_COMPANY} \times \text{BUILDING_AGE}) \\ & + \beta_9(\text{THREE_NIL}) + \beta_{10}(\text{THREE_NIL} \times \text{BUILDING_AGE}) \\ & + \beta_{11}(\text{UBW}) + \epsilon \end{aligned} \quad (1)$$

Where the BCI is the index, C is a constant, and the descriptions of the rest variables are listed in Table 1. The variables in Table 1 cover the multiple facets of building conditions, such as physical appearance in aesthetics, quality of management in maintenance organization, and the presence of unauthorized building works (UBW) in related legal issues. The BCI can thus provide a quantitative, overall measure of building health based on various factors, according to Ho (2013).

Table 1. Descriptions of Variables

Variables	Descriptions
BUILDING_AGE	The age of the building (linear term)
BUILDING_AGE ²	The squared age of the building, capturing nonlinear effects
GFA	Gross Floor Area (linear term)

GFA ²	The squared Gross Floor Area
FLOOR	The number of floors (linear term)
FLOOR ²	Squared number of floors
MANAGEMENT_COMPANY	Dummy variable for the presence of a management company
MANAGEMENT_COMPANY × BUILDING_AGE	The interaction term between the dummy of the management company and the building age
THREE_NIL	Dummy variable for the “three-nil” status (buildings without proper management)
THREE_NIL × BUILDING_AGE	The interaction term between three-nil status and building age
UBW	Dummy variable of the presence of Unauthorized Building Works (UBW)
ε	Error term capturing unexplained variance in BCI

2.2 Case area and data source

As shown in Figure 1, the case area is selected as a part of the Kowloon City District in Hong Kong in this paper. The area was selected for encompassing typical aged and newly built buildings in the sub-districts of Kowloon City, Ma Tau Wai, and To Kwa Wan. A total of 1,211 buildings were selected and analyzed. As indicated by the colors in Figure 1, the majority of the buildings in the area were over 40 years old, while certain clustered blocks of buildings were over 60 years old (indicated as dark red).

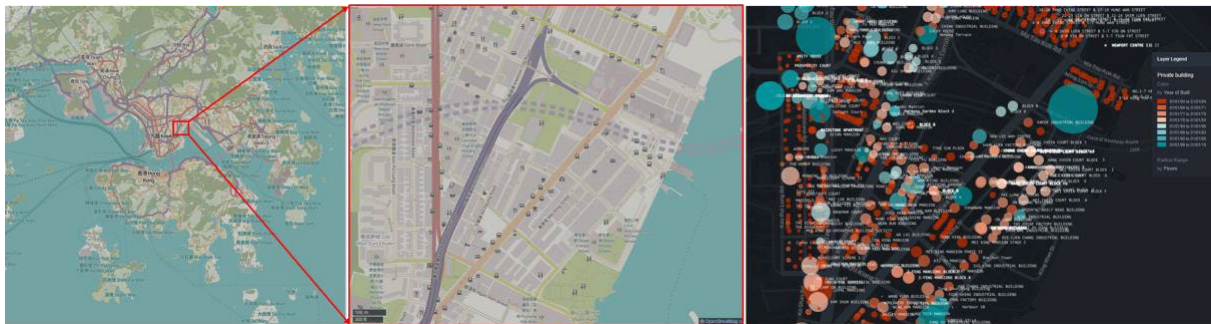


Figure 1. A case area of 1,211 buildings in Kowloon City District, Hong Kong (warmer color indicates older buildings, map sources: OpenStreetMap, MapBox)

Figure 2 stipulates the data sources and overall workflow to construct the BCI. Most of the independent variables marked in light green are available publicly online or via a site survey. For example, the source of building age, gross floor area, and number of floors is the Buildings Department’s building information and age records. Information regarding the presence of the management company and the three-nil status for each building is confirmed by searching the Land Registry records and verifying the registration of the Deed of Mutual Covenants for the building. Site visits to old buildings where the management company is absent were conducted to further confirm the absence of the owners’ corporation and owners’ committee, substantiating the three-nil status of these buildings.

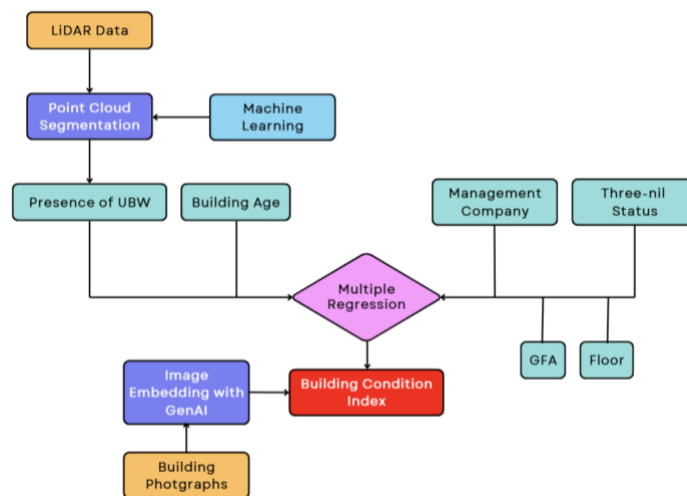


Figure 2. Flowchart of the construction of BCI

The processing for “presence of UBW” in Figure 2 begins with LiDAR data. The LiDAR data collected by CEDD (2023) was processed through point cloud segmentation and machine learning algorithms to identify potential UBW features, such as rooftop structures. Simultaneously, key variables, such as building age, Gross Floor Area (GFA), number of floors, the presence of a management company, and “three-nil” status (denoting buildings lacking proper management), are systematically collected. Additionally, building photographs undergo image embedding using Generative AI (GenAI), which predicts a preliminary BCI score by analyzing visual indicators of building conditions. These GenAI-generated BCI scores are then integrated into a multiple regression model to evaluate the relationships between the BCI and the exogenous factors influencing building conditions. This dual approach not only verifies the accuracy of the AI-predicted BCI but also enhances the explanatory power of the regression model, offering a robust framework for constructing the BCI in assessing and addressing building quality within old districts in Hong Kong.

2.3 Data processing for unauthorized building works and site image embedding

This paper triangulates two data sources to confirm the presence of UBW. First, searches of the Land Registry records for the disposal or removal orders regarding UBW were conducted as training data and references. The base point cloud data was obtained from an airborne LiDAR scan by the Civil Engineering Development Department in 2020 (CEDD 2023). Using the annotated UBW in the CEDD’s (2023) LiDAR data, we then trained machine learning to segment 3D point clouds to identify potential rooftop areas that constitute UBWs. This is a task-specific point cloud segmentation, in which one can discern various rooftop elements such as air conditioning units, water tanks, and other installations, which are typical provisions of rooftop structures.

Regarding the use of machine learning to segment point clouds of rooftop UBWs, the CANUPO plugin within CloudCompare v2 was used to handle multi-dimensional data and various point cloud feature sets, including intensity, number of returns, and return number, to name but a few, alongside random forests. These point cloud features serve as a foundation for training the machine learning model, especially the intensity feature set, denoting the reflectivity of bi-spectral surveyed surfaces, which helps classify the different materials of rooftop structures (Zhao et al., 2022). To be specific about the training process, first, we segment the rooftop point cloud, then analyze the pattern and distribution of the intensity dataset of the rooftop point cloud with random forest classification to identify the potential steel-sheeting roof covers eventually, which are ubiquitously recognized as UBWs in old buildings (BD, 2023b). The time of training was 31 hours, with an overall CANUPO confidence of 0.94.

Figure 3 shows the detected anomalies of possible UBWs on rooftop structures. The results are mostly the temporary steel sheet structures, which often constitute UBWs as deviating from the temporary structures. The automated anomaly identification process for UBW significantly enhances the efficiency and accuracy of UBW detection compared to traditional manual inspections.

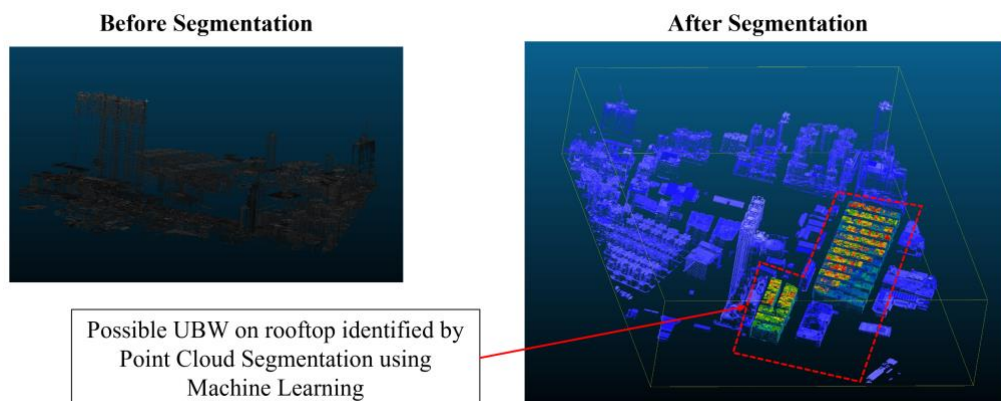


Figure 3. Illustration of point cloud segmentation to automatically identify UBW

ChatGPT (ver. 4o) and high-resolution images of building facades are employed to offer embedded visual features to BCI. Figure 4 shows an example. A high-resolution image of a case building was taken by smartphone and uploaded through an intuitive interface. Then, the process employs robust image embedding techniques using the ChatGPT-4o service’s enhanced photograph for stringent quality standards essential for precise analysis. Following this, the Image Preprocessing visual analysis capabilities extract critical visual features such as signs of wear, discoloration, cracks, and structural irregularities. These refined images are then subjected to Feature Analysis, where a pre-trained transformer neural network (TNN), enriched with photographic feature datasets curated by OpenAI, adeptly describes and captures intricate image details. This descriptive data is subsequently transformed into numerical embeddings that quantitatively represent the building’s appearance and

condition, enabling a nuanced evaluation of various structural aspects.

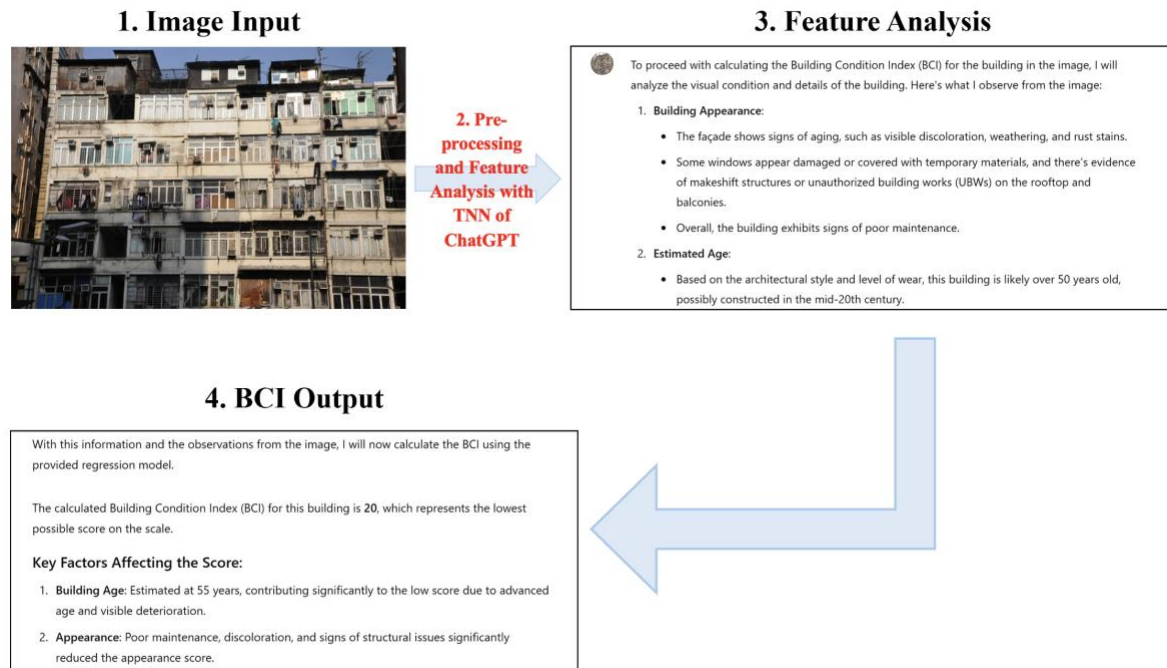


Figure 4. Example of image-based feature processing by TNN of ChatGPT for BCI assessment

In the final Output Generation stage, the system calculates the BCI score using a weighted algorithm that accounts for the severity and impact of each identified feature, normalizing the score for standardized interpretation. The output BCI is then recorded for every building in the study area as the dependent variable.

2.4 Overview of the dataset in the case area

Table 2 presents the descriptive data of all the variables. The variables in the dataset were in accordance with a multi-variable regression analysis. For example, the maximum (“best building condition”) of $\ln(\text{BCI})$ was $\ln(100) = 4.603$, while the minimum (or the “worst”) was $\ln(20) = 2.996$. The median of BCI was $\text{EXP}(3.88) = 48.4$, which was closer to the end of 20. Similarly, the median BUILDING_AGE was 44 years, and the median of GFA was 2,381.2 square feet, which was relatively aged and small for a high-density city like Hong Kong. The median number of floors was eight storeys, which was lower than the average of a high-rise urban form in Hong Kong. About 11% of buildings, on average, were identified as “There-nil” buildings.

Table 2. Descriptions Data

	$\ln(\text{BCI})$	BUILDING_AGE	BUILDING_AGE^2	GFA	GFA^2	FLOOR	FLOOR^2
Mean	3.898738	41.77666	2013.713	4613.769	98902886	10.21942	171.7536
Median	3.882826	44	1936	2381.223	5670223	8	64
Maximum	4.603086	77	5929	102683.8	1.05E+10	59	3481
Minimum	2.996134	1	1	38.90204	1513.369	0	0
Std. Dev.	0.441536	16.38824	1328.868	8812.469	5.84E+08	8.207008	313.3496
Skewness	-0.181719	-0.29059	0.51038	5.715663	11.18781	1.694284	5.252967
Kurtosis	1.98624	2.450962	2.602456	44.53008	154.2416	8.008287	39.9004
Jarque-Bera Probability	86.115 0	47.46157 0	89.09948 0	137765.1 0	1735570 0	2714.976 0	109297 0
Sum	6947.55	74446	3588436	8221736	1.76E+11	18211	306065
Sum Sq. Dev.	347.2125	478331.1	3.15E+09	1.38E+11	6.08E+20	119959.2	1.75E+08
Observations	1211	1211	1211	1211	1211	1211	1211

Table 2. Descriptions Data (Con't)

	MANAGEMENT_COMPANY	MANAGEMENT_COMPANY*BUILDING_AGE	THREE_NIL	THREE_NIL*BUILDING_AGE
Mean	0.838384	33.68462	0.11055	6.945006
Median	1	38	0	0
Maximum	1	75	1	77
Minimum	0	0	0	0
Std. Dev.	0.368202	20.09236	0.313662	19.85852
Skewness	-1.838551	-0.411676	0.483944	0.55713
Kurtosis	4.380271	2.042261	7.169976	7.690309
Jarque-Bera Probability	1145.398 0	118.4417 0	3123.593 0	3575.483 0
Sum	1494	60026	197	12376
Sum Sq. Dev.	241.4545	718994.8	175.2217	702356.6
Observations	1211	1211	1211	1211

Figure 5 stipulates the Pearson's correlation coefficient matrix between the variables. It can be seen that both the squared features, including BUILDING_AGE² and GFA², had strong positive correlations with the original feature. In addition, the last two variables, i.e., THREE_NIL and THREE_NIL*BUILDING_AGE, had an almost perfect ($r = 0.99$, $N = 1,211$) linear correlation. In contrast, the THREE_NIL and BUILDING_AGE had a moderate positive linear correlation ($r = 0.45$, $N = 1,211$). From the first row of Figure 5, it can be seen that ln(BCI) had the most strong linear correlations with BUILDING_AGE ($r = -0.47$), THREE_NIL ($r = -0.42$), and THREE_NIL*BUILDING_AGE ($r = -0.41$).

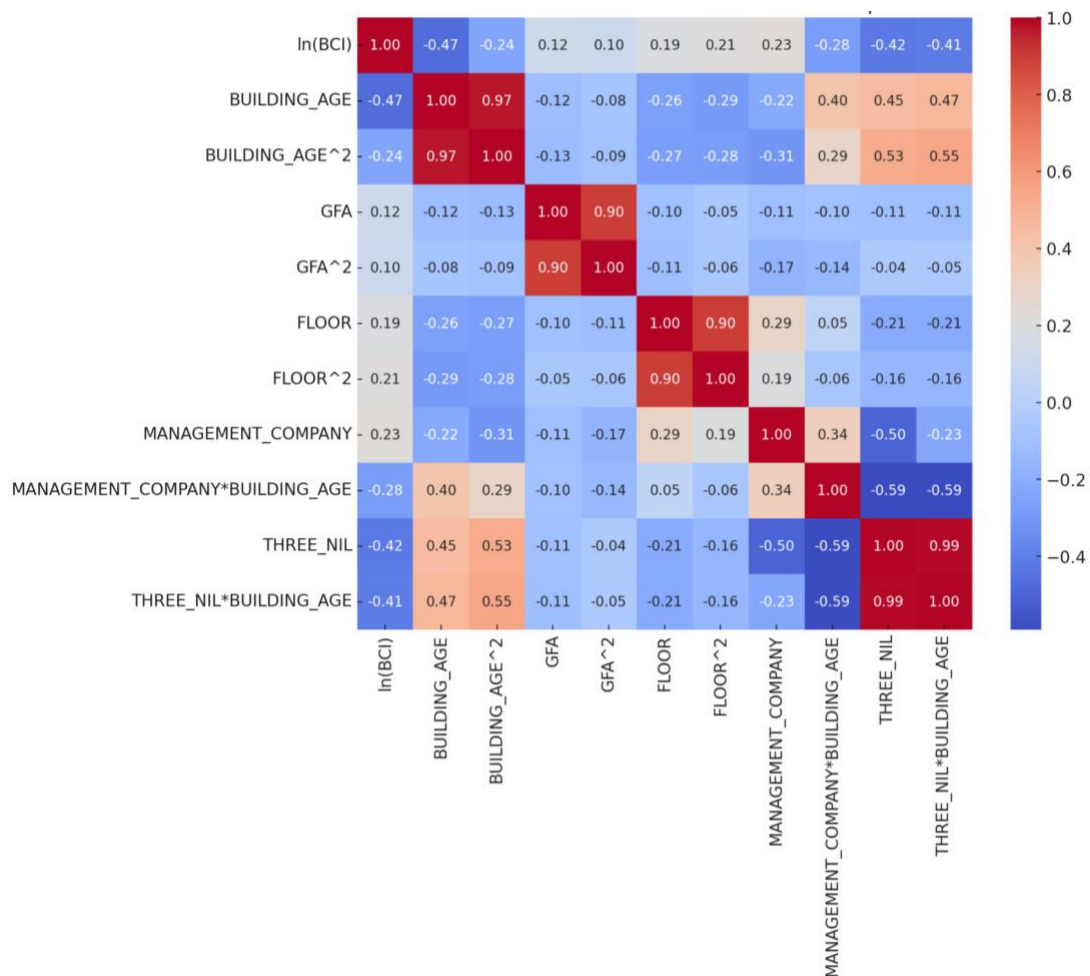


Figure 5. Correlation coefficient matrix heatmap

3. RESULTS

3.1 Regression results

The least squares regression analysis examined the determinants of the natural logarithm of BCI using a sample of 1,211 observations. Table 3 shows the empirical results of the regression.

Table 3. Multiple regression results (row in bold when a variable's p-value ≤ 0.001)

Variable	Coefficient	(Std. Error)	p-value
Constant	4.5477	(0.0439)	0.0000
BUILDING_AGE*	-0.0412	(0.00199)	0.0382
BUILDING_AGE^2**	0.000124	(0.0000164)	0.0043
GFA**	-2.52E-06	(9.46E-07)	0.0079
GFA^2*	3.42E-11	(1.38E-11)	0.0135
FLOOR	-0.00183	(0.00109)	0.0947
FLOOR^2	2.56E-05	(2.68E-05)	0.3385
MANAGEMENT_COMPANY**	0.0703	(0.0472)	0.0031
MANAGEMENT_COMPANY*BUILDING_AGE***	0.00707	(0.00196)	0.0003
THREE_NIL***	-0.8665	(0.1287)	0.0000
THREE_NIL*BUILDING_AGE**	-0.00768	(0.00285)	0.0072
UBW***	-0.0321	(7.32E-07)	0.0003

R-squared: 0.8716

No. of Observations: 1211

*: Significant at 5% level; **: Significant at 1% level; ***: Significant at 0.1% level

A negative relationship between building age and the Building Condition Index (BCI) is anticipated due to the increased likelihood of structural deterioration and higher maintenance costs associated with older buildings (Chanter & Swallow, 2008). Consequently, as buildings age, their condition is expected to decline, reducing their overall BCI. The regression results support this expectation, with the coefficient for BUILDING_AGE being -0.00412 ($p = 0.038$), indicating that each additional year of building age is associated with a 0.412% decrease in the BCI. Furthermore, the positive coefficient for BUILDING_AGE² (0.000124, $p = 0.004$) suggests a decelerating rate of decline in building conditions as buildings become increasingly old, possibly due to stabilization effects or periodic major renovations that temporarily improve building conditions.

Gross Floor Area is expected to have a complex relationship with the Building Condition Index. On the one hand, larger buildings may benefit from economies of scale in maintenance, while on the other hand, they may face higher absolute maintenance costs and greater challenges in upkeep (Ho et al., 2011). The analysis reveals a negative coefficient for GFA (-0.00000252, $p = 0.008$), indicating that larger floor areas are associated with a slight decrease in BCI. However, the positive coefficient for GFA² (0.0000000342, $p = 0.014$) implies that the negative impact of increasing floor area on building condition diminishes at higher levels of GFA, suggesting that very large buildings may implement more effective maintenance strategies that mitigate the adverse effects of size.

The number of floors in a building is expected to influence the BCI, potentially due to the complexities associated with maintaining taller structures (Seeley, 1987). The coefficient for FLOOR is -0.00182705 ($p = 0.095$), indicating a marginal negative effect on BCI, though it is not statistically significant at the conventional 5% level. The squared term for FLOOR (0.0000256, $p = 0.338$) is also not significant, suggesting that there is no clear nonlinear relationship between the number of floors and building conditions within the sampled data.

The presence of a management company is hypothesized to positively impact the BCI by ensuring regular maintenance and efficient management practices (Seeley, 1987). The coefficient for MANAGEMENT_COMPANY is 0.07035 ($p = 0.003$), confirming that buildings managed by professional companies tend to have higher BCIs. Additionally, the interaction term between MANAGEMENT_COMPANY and BUILDING_AGE (0.00707, $p < 0.001$) indicates that the positive effect of management companies on building condition buildings age, therefore, can be interpreted that aged building with management company can slightly enhance the BCI.

Three-nil buildings, which lack owners' corporations, residents' organizations, or property management companies, are expected to exhibit poorer maintenance and lower BCIs due to the absence of coordinated upkeep efforts (Ho et al., 2011). The regression results substantiate this expectation, with a significant negative coefficient for THREE_NIL (-0.86652, $p < 0.001$), indicating that three-nil buildings have substantially lower BCIs compared to those with proper management structures. The interaction term between THREE_NIL and BUILDING_AGE (-0.00768, $p = 0.007$) further reveals that the detrimental effect of being a three-nil building on the BCI intensifies as the building ages, highlighting the critical need for organized management in maintaining older buildings.

UBW is posited to negatively affect the BCI, as unauthorized modifications can lead to structural weaknesses and potential accelerated deterioration (Ho et al., 2011). The coefficient for UBW is -0.03215 ($p < 0.001$), indicating that higher instances of unauthorized building works are significantly associated with lower BCIs. This suggests that buildings with unauthorized modifications experience more rapid declines in condition, underscoring the importance of regulatory enforcement and regular inspections to prevent such works and maintain building integrity.

In short, the model demonstrated a high explanatory power, with an R-squared of 0.872 in Table 3 and an adjusted R-squared of 0.871, indicating that approximately 87.2% of the variability in $\text{LN}(\text{BCI})$ is accounted for by the independent variables included in the model, indicate the reliability of using ChatGPT analysis in quantifying the building condition of old buildings.

3.2 Visualization of BCI

Figure 6 visualizes the heatmap of BCI assessed with the regression model on ArcGIS Pro. The chart highlights the spatial distribution of structural conditions within the study area. Using a combination of detailed maps and georeferenced building data, the tool identifies clusters of aging or decayed buildings.

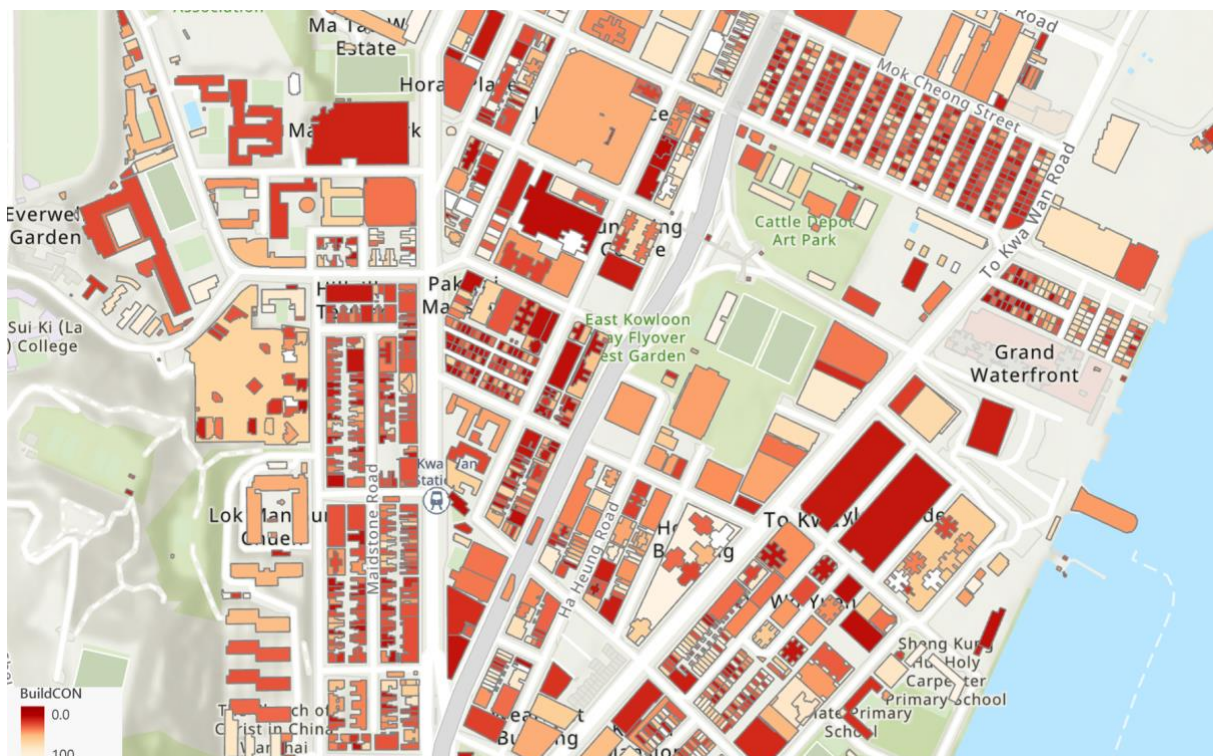


Figure 6. Visualization of BCI assessment results

Figure 7 shows a hotspot analysis of areas with significant concentrations of deteriorated structures. The hotspot areas pinpoint the central blocks that further refine the understanding of decayed buildings in the area. As shown in Figure 7, two prominent clusters of decayed buildings are identified: Cluster 1 along Ma Tau Wai Road and Lok Shan Road and Cluster 2 near Mok Cheong Street and To Kwa Wan Road. The clustering results provide new insights into urban decay patterns and facilitate data-driven urban planning decisions. For example, urban planners can prioritize redevelopment or maintenance efforts.

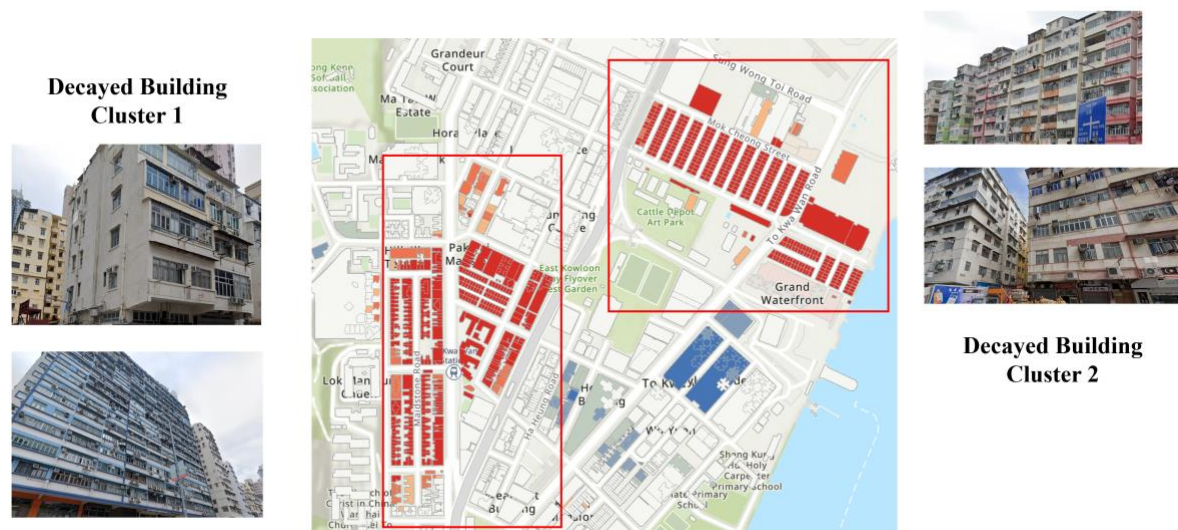


Figure 7. Hotspot Analysis showing the decayed building clusters

4. DISCUSSION

The BCI and assessment have several advantages. First, a multi-regression of the first part of BCI includes managerial conditions and UBWs, in addition to traditional physical building conditions. The significant coefficients (p -value < 0.001) of `MANAGEMENT_COMPANY*BUILDING_AGE`, `THREE_NIL`, and `UBW` showed the newly added variables impacted Hong Kong's aging buildings. Furthermore, the overall BCI integrates vector embedding of building exterior photos and the first part's regression result using the TNN of ChatGPT. This indicated that GenAI might have a huge potential in building and urban assessment. The assessment and mapping of BCI, as demonstrated in Sect. 3.2, visualizes and highlights the building decay in the target area.

There were also certain limitations in this paper. One limitation is the lack of urban environmental features, which might lead to confirmation bias. Another is the lack of a detailed maintenance record for each building, so the BCI might not represent the temporal relationship. Last but not least, the BCI is a localized regression indicator for typical building structures in the urban areas in Hong Kong, while the significance of variables `THREE_NIL` and `UBW` may not be reproduced in other cities.

5. CONCLUSIONS

Hong Kong has many aging buildings. This paper defines a novel building condition index (BCI) to represent the building decay conditions of existing buildings in Hong Kong. The BCI integrates big urban data of physical building conditions, managerial conditions, localized conditions such as `THREE_NIL` and `UBW`, and building exterior photos via multi-regression and ChatGPT. The assessed and mapped BCI can offer quantitative and visualized results of building conditions for building surveying, real estate, and urban planning. Future research directions are suggested as urban environmental features and maintenance record.

ACKNOWLEDGMENTS

The work in this paper was supported by the Undergraduate Research Fellowship Programme and the Teaching Development Grant from the University of Hong Kong.

REFERENCES

- BD. (2023a). Building information and age records. In *CSDI Portal*. Hong Kong: Buildings Department of the Government of Hong Kong SAR. <https://portal.csd.gov.hk/geoportal/#metadataInfoPanel>
- BD. (2023b). *Code of practice of Mandatory Building Inspection Scheme and Mandatory Window Inspection Scheme 2012 (2023 Edition)*. Hong Kong: Buildings Department of the Government of Hong Kong SAR. https://www.bd.gov.hk/doc/en/resources/codes-and-references/code-and-design-manuals/CoP_MBIS_MWISe.pdf
- CEDD. (2023). *The CEDD 2020 LiDAR Survey*. Hong Kong: Civil Engineering and Development Department of the Government of Hong Kong SAR. <https://sdportal.cedd.gov.hk/#/en/lidar>
- Chanter, B., & Swallow, P. (2008). *Building maintenance management*. John Wiley & Sons.
- Chiu, T. K. (2023). The impact of Generative AI (GenAI) on practices, policies and research direction in

- education: A case of ChatGPT and Midjourney. *Interactive Learning Environments*, 1-17. <https://doi.org/10.1080/10494820.2023.2253861>
- García-Peñalvo, F., & Vázquez-Ingelmo, A. (2023). What do we mean by GenAI? A systematic mapping of the evolution, trends, and techniques involved in Generative AI. <https://reunir.unir.net/handle/123456789/15134>
- Ho, D., Yau, Y., Wah Poon, S., Law, C. K., Chui, E., Wong, Y., ... & KH Wong, K. (2011). Urban renewal in Hong Kong: a community aspiration study. *Surveying and Built Environment*, 21, 63-77.
- Hui, E. C., Wong, J. T., & Wan, J. K. (2008). A review of the effectiveness of urban renewal in Hong Kong. *Property Management*, 26(1), 25-42. <https://doi.org/10.1108/02637470810848877>
- Jasiulewicz-Kaczmarek, M. (2016). SWOT analysis for Planned Maintenance strategy-a case study. *IFAC-PapersOnLine*, 49(12), 674-679.
- Jiang, Y., Han, S., & Bai, Y. (2021). Building and infrastructure defect detection and visualization using drone and deep learning technologies. *Journal of Performance of Constructed Facilities*, 35(6), 04021092. [https://doi.org/10.1061/\(ASCE\)CF.1943-5509.0001652](https://doi.org/10.1061/(ASCE)CF.1943-5509.0001652)
- Lee, G. K., & Chan, E. H. (2008). Factors affecting urban renewal in high-density city: Case study of Hong Kong. *Journal of Urban Planning and Development*, 134(3), 140-148. [https://doi.org/10.1061/\(ASCE\)0733-9488\(2008\)134:3\(140\)](https://doi.org/10.1061/(ASCE)0733-9488(2008)134:3(140))
- Leung, T. C., Tsang, K. P., & Tsui, K. K. (2022). Collusion In Private Procurement: How Does Mandatory Building Repair Increase Renovation Prices?. *The Singapore Economic Review*, 67(05), 1759-1779. <https://doi.org/10.1142/S0217590822430020>
- Salvagno, M., Taccone, F. S., & Gerli, A. G. (2023). Artificial intelligence hallucinations. *Critical Care*, 27(1), 180. <https://doi.org/10.1186/s13054-023-04473-y>
- Seeley, I. H. (1987). *Building maintenance*. Bloomsbury Publishing.
- Wang, Y., Fan, Y., & Yang, Z. (2022). Challenges, experience, and prospects of urban renewal in high-density cities: a review for Hong Kong. *Land*, 11(12), 2248. <https://doi.org/10.3390/land11122248>
- Xue, F., Wu, L., & Lu, W. (2021). Semantic enrichment of building and city information models: A ten-year review. *Advanced Engineering Informatics*, 47, 101245. <https://doi.org/10.1016/j.aei.2020.101245>
- Zhao, Y., Seo, H., & Chen, C. (2022). Displacement analysis of point cloud removed ground collapse effect in SMW by CANUPO machine learning algorithm. *Journal of Civil Structural Health Monitoring*, 12(2), 447-463.



Part 14

Digital Campus

A SCHEME AND APPLICATION OF A DIGITAL TWIN-BASED ROAD PAVEMENT MANAGEMENT SYSTEM

Koji Makanae¹, Yogo Kurokawa², Kiyotaka Suda³, Jevica⁴, and NurulNajwa Binti Khamis⁵

1) Ph.D., Professor, School of Project Design, Miyagi University, Japan, Email: makanae@myu.ac.jp

2) Ikee Co., Ltd., Matsuyama, Ehime, Japan

3) Kankyo Fudo Techno Co., Ltd, Tokyo, Japan

4) Construction IoT Research Lab, Aichi, Japan

5) Construction IoT Research Lab, Aichi, Japan

Abstract: It is necessary to conduct efficient repairs on road pavements by extending the service life of pavements and reducing life cycle costs (LCC). On the other hand, many small and medium-sized local governments and construction enterprises (SMEs) are responsible for road maintenance work, and the sophistication of ICT (information and communication technology)-based maintenance management technology has not yet progressed. The authors applied digital measurement technology and constructed a pavement management system based on digital twin technology, assuming that (SMEs) can use it. The maintenance cycle of road pavement can be expressed as follows: inspection, diagnosis, action, record, and next inspection. The authors use a digitalized virtual model (digital twin) as the core, acquire (sensing or monitoring) information and inspection results from the actual model existing in the physical space, and store and accumulate them in the virtual model, which becomes the digital twin. We defined a cycle model that accurately predicts the future by analyzing and simulating the data using a mathematical model and aimed to construct a system to realize this model. In this study, we developed a system based on a spatial information infrastructure. The system can display point cloud data generated from images taken by a simple in-vehicle stereo camera, road surface property data from simple laser measurement, and other data in conjunction with spatial information, and its usefulness has been evaluated. The system was developed using the GIS software "QGIS", which is freeware, to make it easy for SMEs to use. In the future, it is necessary to consider introducing a system that can be shared on the cloud for easier use.

Keywords: Road pavement management, Digital twin, Road maintenance, SMEs

1. INTRODUCTION

While infrastructure development is steadily progressing, countermeasures against the aging of the infrastructure that has been built up to date have also become a significant issue, and there is growing recognition of the need for appropriate maintenance and management. The "Pavement Inspection Guideline" (Road Bureau, Ministry of Land, Infrastructure, Transport and Tourism, Japan, 2016) was established in Japan in 2016, and inspections based on this guideline have been conducted since 2017. However, due to issues such as costs and burdens related to inspections, the Guidelines also set a guideline for inspection frequency of at least once every five years, even for roads with heavy traffic, and routine inspections are left to daily inspections. In addition, efforts are being made to introduce performance regulations and long-term warranty systems for pavement construction, which require an accurate and frequent understanding of pavement surface properties.

Against this background, this study discusses measures for the use of digital twin in road management, based on examples of the application of management technology using digital technology, assuming that it can be used by small and medium-sized local governments and construction enterprises (SMEs) that are responsible for road maintenance work, with the aim of advancing pavement maintenance technology.

2. DIGITAL TWIN AND SPATIOTEMPORAL INFORMATION SYSTEMS

With the evolution of information and communication technology in recent years, various sensors in physical space have been connected to digital networks, creating an environment in which information can be simultaneously obtained even in remote locations. By consolidating such information, it is possible to virtually reproduce the same environment in a virtual space as in a physical space. In other words, it can be regarded as a digital twin to the physical space, which has recently been called a "digital twin" (Saddik, 2018). In this study, we define "digital twin" as a virtual digital model that can reproduce a space-time equivalent to the real world by acquiring and storing various measurement information obtained from the real world.

With the rapid and dramatic progress of the Internet and mobile wireless communications, an environment in which various terminals are connected to the Internet was realized in the 2000s, which was described as "ubiquitous computing" at the time. The development of IPv6, which dramatically increased the number of terminal addresses on the Internet, and the spread of sensors and control devices combined with card-sized CPU boards and LPWA (low-power wireless communication) attracted attention to the concept of IoT (Internet of Things), in which all things are connected to the Internet. The IoT is a concept that connects all things to the Internet. The system that fuses cyberspace and physical space formed by these IoT devices connected to the

Internet and aggregating the digital data transmitted from them is called Cyber-Physical Systems (CPS) (National Institute of Standards and Technology, 2013), which is the core of Society 5.0 in Japan (Ministry of Internal Affairs and Communication, Japan, 2019), the keyword of the Japanese government's information policy.

In the CPS concept, digital data acquired from physical space is aggregated and reconstructed as a digital twin in a space similar to physical space. However, the representation of the aggregated information is not limited to a three-dimensional space identical to the physical space; only the necessary information can be extracted and represented as required, regardless of the dimension. For example, temperature information is obtained through sensors, but it is represented in visual forms such as maps (contour plots, etc.), tables, graphs, etc., and is not directly reproduced as gas temperatures that can be experienced.

An essential characteristic of the digital twin is the ability to store digital data acquired in time series. Events that occur in physical space are sequential and cannot be replayed. However, the past can be reproduced based on the data stored in the digital twin. In other words, past events that could not be reproduced in physical space can be reproduced in digital space. Furthermore, as time passes, digital data accumulates and forms big data. This big data records many of the events that occurred in the physical space in the past. Analysis of the vast amount of accumulated data improves the accuracy of predictions of future events. Furthermore, by comparing the results of predictions with sequentially acquired data and feeding the errors back into the prediction model, it is possible to improve the accuracy of predictions further. If such a system can be realized, a time machine can be constructed as a spatiotemporal information system that allows humans to freely move through time within the digital twin.

We consider the application of the concept of a spatiotemporal information system based on the digital twin to the infrastructure lifecycle. Planning and designing infrastructures in the initial stages are carried out in a virtual space, incorporating information from physical space (e.g., topography and geology) through surveys and investigations. Recently, this virtual space can be called an information space or cyberspace because it is constructed as digital information based on CAD and GIS. As the design process progresses, the virtual models (Building and Construction Information Modeling and Management: BIM/CIM) built in the virtual space become more detailed, and the amount of information increases. Once the design is completed, the design information is transferred to the construction contractor, who plans the construction process and more detailed construction design while construction proceeds in the physical space. When the construction of the target infrastructure is completed, the virtual model and the physical model are in perfect agreement, and the process then moves on to the facility operation and maintenance phase.

A physical model placed in the real world will deteriorate due to external influences such as vibration and weather caused by facility operation and may also be damaged due to disasters and accidents. On the other hand, the virtual model does not deteriorate; however, a mathematical prediction model can be applied to predict future deterioration conditions by understanding the deterioration and operational conditions obtained from the physical space. This enables the formulation of a systematic repair plan, which can be fed back as repair and improvement of facilities in the physical space. The realization of this process enables efficient maintenance management and reliable repair and improvement of infrastructure and is expected to extend the service life of the infrastructure itself.

3. SIGNIFICANCE OF DIGITAL TWIN APPLICATION TO ROAD MAINTENANCE AND MANAGEMENT

3.1 Objectives of Improvement on Road Maintenance and Management

In Japan, the development of modern infrastructure has been steadily progressing since the Meiji period (1868-1912), and especially in the postwar period, infrastructure was developed in line with rapid economic growth. On the other hand, problems with aging infrastructure have become apparent in Europe and the United States, where infrastructure had been developed ahead of Japan. In response, the Ministry of Land, Infrastructure, Transport and Tourism (MLIT) and others have been studying appropriate maintenance and management processes, and in 2016, the "Pavement Inspection Guidelines" were established for road pavements with the aim of carrying out efficient repairs to extend pavement life and reduce life cycle costs (LCC), and since the following year, 2017, inspections have been carried out based on these guidelines. The following year, in 2017, inspections based on this guideline began to be carried out.

On the other hand, due to the cost and burden of actual inspections, the inspection frequency in the same guidelines is set at once every five years or more for roads with rapid damage progression, and routine inspections are left to routine inspections. On the other hand, efforts are being made to introduce performance specifications and long-term warranty systems for pavement construction, and there is a need for accurate and frequent monitoring of pavement surface properties. Against this background, the objective of this project is to improve road maintenance and management by applying digital measurement techniques and utilizing digital twin technology.

Objective: To advance road maintenance and management technology by the application of measurement technology based on digital technology.

- Improve inspection frequency (real-time) and accuracy by various time-series data acquired by automatic

inspection systems.

- Accurately predict the future based on measurement data and support the development of management plans.
- Achieve information management within the time and space of the digital twin.
- It should be possible for SMEs involved in road management in rural areas to introduce the system.

In recent years, research has been conducted to apply the information stored in digital twins to maintain and manage infrastructures. In the field of road pavement, for example, research has been conducted mainly on the advancement of prediction technology, such as tunnel pavement maintenance by Yu et al. (2020), and the development of a cognitive digital twin for pavement infrastructure by Sierra et al. (2022). Talaghat et al. (2024) also described that road pavement DT research is still scarce and that there are only a few use cases in pavement reactive maintenance in their latest review paper about digital twin technology for road pavement. This research is unique in that it attempts to build a system that can be easily used by SMEs.

3.2 Application of Digital Twin in Road Maintenance Management

In road maintenance and management, the inspection includes not only the detection of abnormalities through daily inspections but also the identification of road surface conditions and underground cavities through periodic diagnoses. The pavement inspection guideline [1] mentioned above lists the following three indices as management criteria: (1) cracking rate, (2) rutting amount, and (3) International Roughness Index (IRI). The degree of progression of deterioration should be determined through soundness diagnosis based on the measured values of these indices, leading to systematic repair and maintenance.

The maintenance cycle can be expressed as the cycle of inspection, diagnosis, action, record, and next inspection. Through this cycle, it is necessary to create and enhance longevity plans and other measures to maintain the appropriate performance required of facilities for more extended periods of time and to promote efficient and effective maintenance management of structures. The process of applying the digital twin to this maintenance cycle is shown in Fig.1.

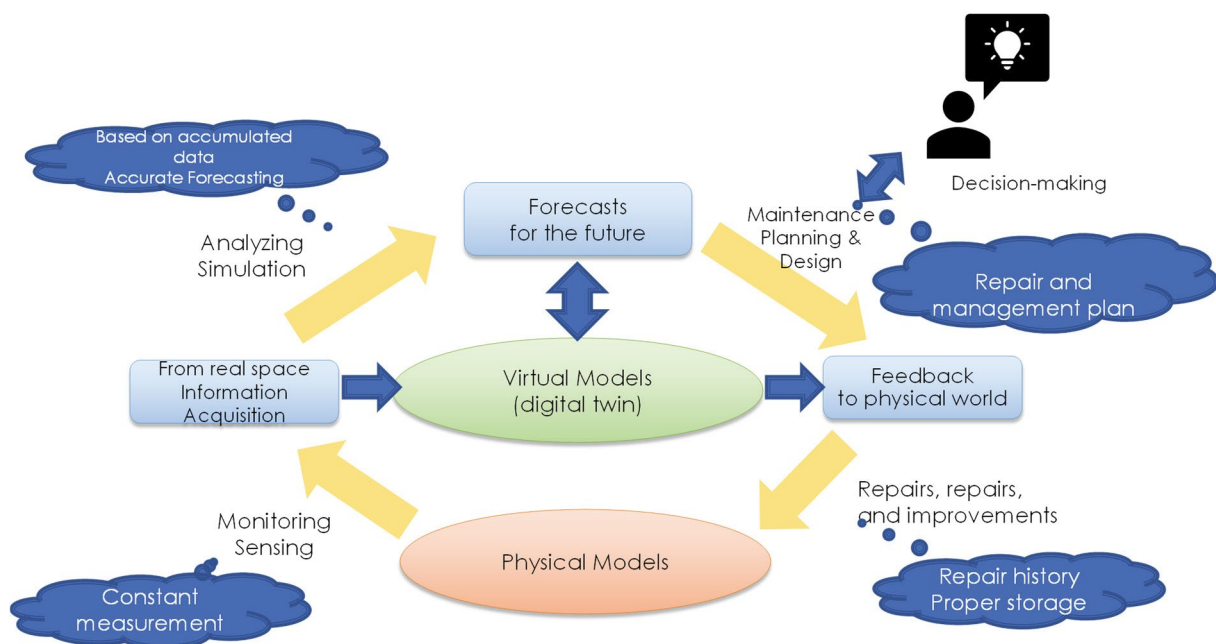


Figure 1. Road maintenance information cycle based on digital twin

Information such as sensors, various measurement results, and inspection results are acquired (sensing and monitoring) from the physical model in the physical space and stored and accumulated in the virtual model, which becomes the digital twin. Based on the accumulated data, accurate predictions of the future are made through analysis and simulation using a mathematical model. Based on these predictions, the road administrator can design a repair or management plan, and the information in the digital twin supports decision-making for that purpose. Based on the decision-making, repair, and improvement are carried out as feedback to reality, and the data obtained in the construction phase is stored in the digital twin as the repair history. In this way, the digital twin functions as the core of the maintenance cycle.

4. DEVELOPMENT OF A DIGITAL TWIN SYSTEM FOR ROAD MAINTENANCE AND MANAGEMENT

4.1 A Scheme of Digital Twin System Based on Spatial Information Infrastructure

A conceptual diagram of the maintenance management system to achieve road maintenance management is shown in Fig.2. Three-dimensional measurement data acquired during periodic inspections and maintenance work are stored as time series data in the management information database. In addition, IRI data, which is an index of longitudinal unevenness of the road surface, and human body measurement data (pulse, EEG, etc.), which evaluates driving comfort, are also stored as time-series data. Maintenance information, such as the discovery of abnormalities through daily inspections and associated repair work, is also stored in this data. Based on the information accumulated in this way, simulations using artificial intelligence (AI) and mathematical models are used to accurately predict the future and support the formulation of management plans. The information in the management information database can be expressed in various media (2D: drawings, 3D: perspective drawings, 4D: process expressions, reports, forms, etc.) as needed (see Fig. 2).

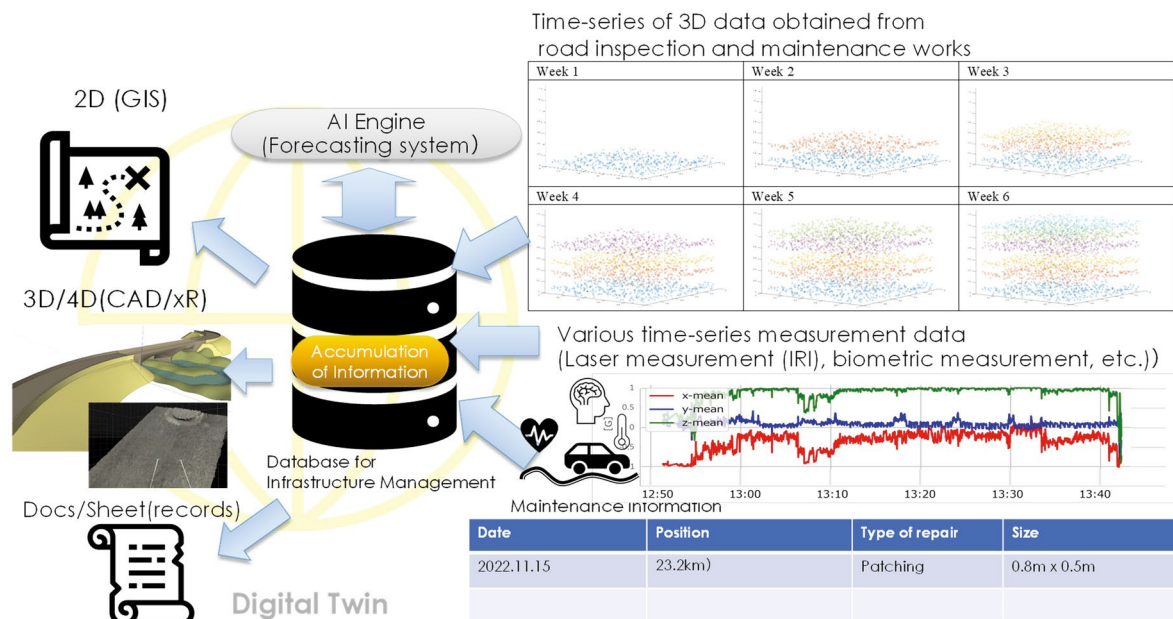


Figure 2. Conceptual diagram of road maintenance system

Spatial information infrastructure plays a significant role in managing large-scale spatial and infrastructure information that exists on national land. In actual system construction, spatial information systems (geographic information systems: GIS) that excel in handling spatial information infrastructures are used as a base. For example, the digital twin platform Plateau of the Urban Planning Bureau of the Ministry of Land, Infrastructure, Transport and Tourism is being developed based on Cesium, a 3D Web-GIS.

The scheme of system to realize the digital twin is shown in Fig. 3. The spatial information system exchanges information with the spatial information infrastructure (spatial data), and various data with location information that are input as medical records are linked and stored on the spatial information infrastructure through dedicated applications.

Through spatial information systems, information can be presented in various forms, such as drawings, perspectives (3D), VR, forms, and reports, depending on the situation, and in some cases, new information can be added by humans and stored on the spatial information infrastructure. In addition, data that can be deployed in 3D can be helpful for ICT construction and control of robots, and so on, at construction sites.

4.2 Advanced Forecasting Technology Through Linkage With BIM/CIM Models

Digitization is essential to improve the efficiency of maintenance management, and BIM/CIM models that show the ideal shape and target values of structures are indispensable to promote the advancement of information and communications systems (ICT). Constructing BIM/CIM models of existing structures is a significant issue. This project devised a simple method to build a 3D model that can be used as a substitute for BIM/CIM based on measurement data at construction time, without using 3D CAD, and is now being tested in practice (see Fig. 4).

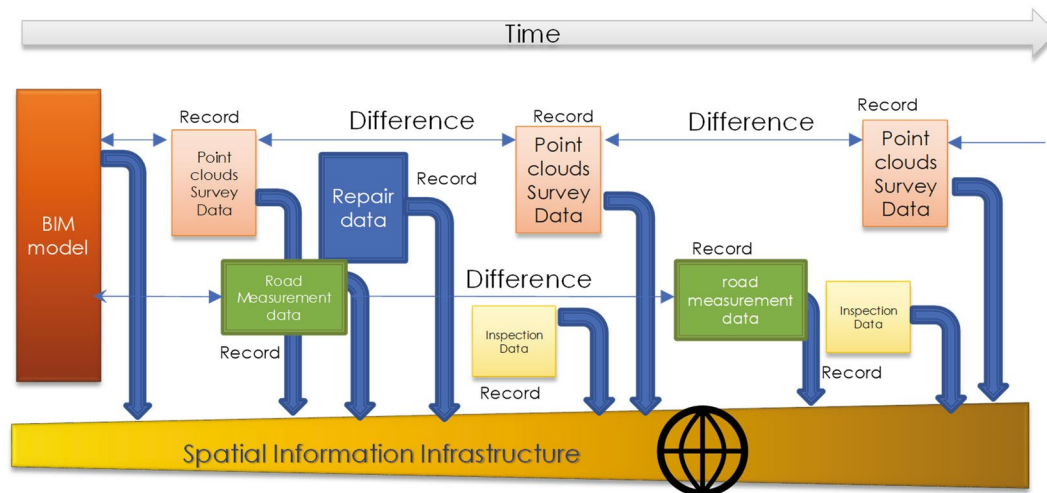
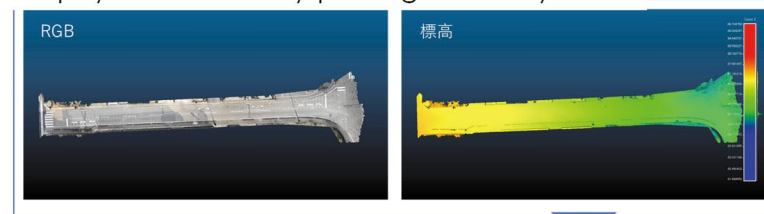


Figure 3. A scheme of an information system for a digital twin

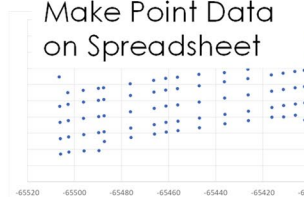
Milling Plan on Spreadsheet



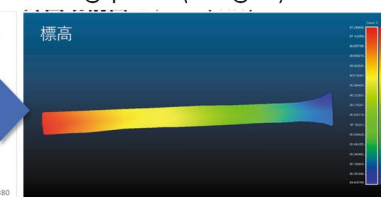
3D physical model by photogrammetry



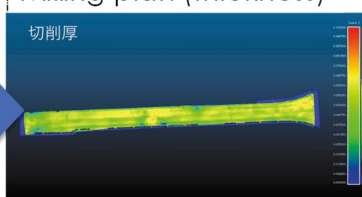
Make Point Data on Spreadsheet



Milling plan (height)



Milling plan (thickness)



従来 Build 3D model without 3D CAD

Figure 4. 3D pavement models by photogrammetry

The construction of a BIM/CIM model (or an alternative 3D model) will enable precise as-built shapes in 3D at the construction stage, which will lead to automated construction in the future, as well as the clear presentation of initial (ideal) values of geometry at the maintenance and management stage, which will facilitate comparison with the current status. Although this project's scope is limited to short-term 3D measurement, a long-term 3D measurement history can be accumulated through periodic measurement work in the future (see Fig. 5).

Accumulation of 3D measurement data and intercomparison of BIM/CIM models will improve the prediction accuracy of mathematical models for predicting deterioration and enable the formulation of management plans based on such models, thereby realizing appropriate maintenance management.

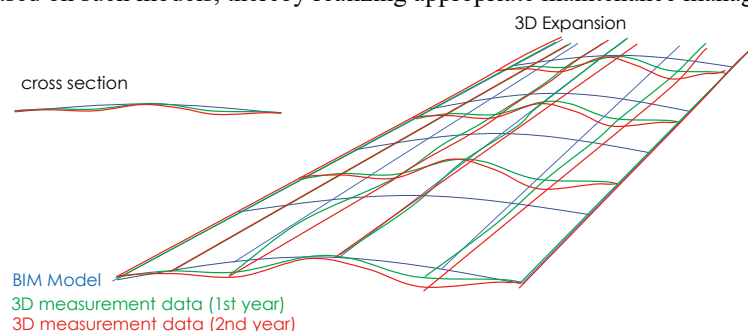


Figure 5. Comparison of CIM models and time-series 3D measurement data

5. DEVELOPMENT OF ROAD PAVEMENT MANAGEMENT SYSTEM ON DIGITAL TWIN

The previous sections have described the systems, information acquisition methods, and linkage with BIM/CIM models necessary for applying digital twins in road maintenance management. One of the challenges in this study is to be possible for SMEs involved in road management in rural areas to introduce the system. The GIS software “QGIS”, which is relatively easy to operate and freeware, is applied to the basis of the geographical information system (see Fig. 6).

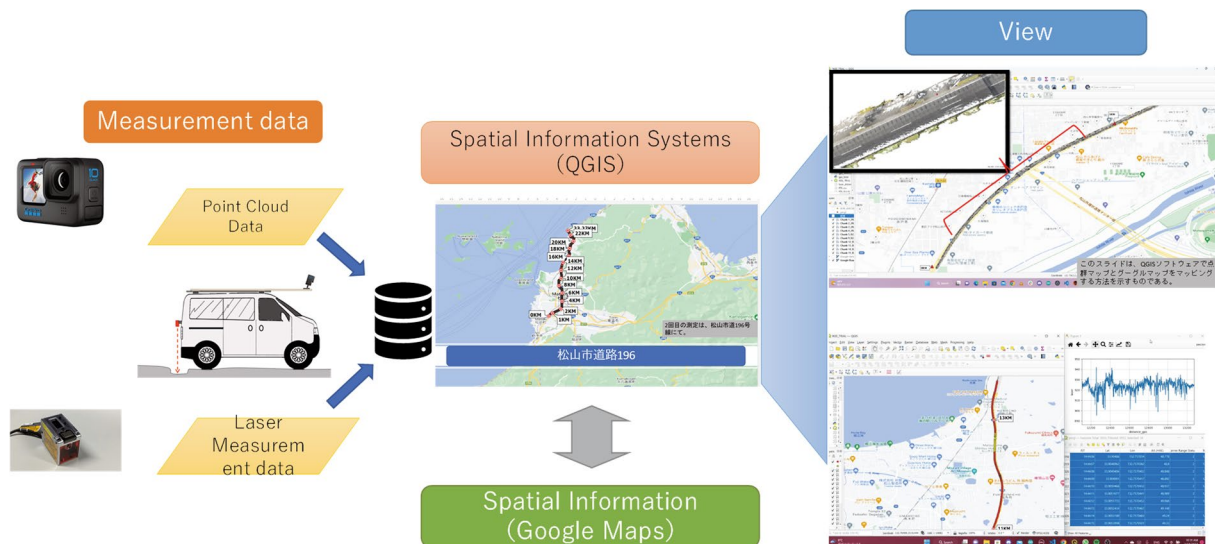


Figure 6. Measurement data management on QGIS

The 3D model acquired from the vehicle-mounted camera and laser and the road surface property data from the laser measurement is connected to the spatial information infrastructure and managed as a database, and it is shown that problem areas can be extracted by comparing both types of information. Although no comparison data is available at this time because of the short period of time required for the trial, we believe that the system will be able to support quick and reliable decision-making on road management by viewing this information once the data is accumulated and compared in the future.

The objective of this project was to advance pavement maintenance and management technology through the use of digital measurement technology and the digital twin and to construct a database that stores the three-dimensional shape of road surfaces and various measurement data acquired from physical space on a spatial information system, and to construct a system to support planned maintenance and management. The database is used to support the planned maintenance and management of roads. The issues to be addressed for future operation and development are as follows.

(1) Improvement of forecasting accuracy through data accumulation and measurement technology effects

The prototype system, which can store data from measurement, was constructed. It is expected that storing various measurement data and inspection/repair records through continuous operation will improve the prediction accuracy of deterioration and will be helpful for more accurate repair planning.

For more accurate prediction, further advancement of measurement technology, such as 3D point cloud data representing road surface geometry, creation of road surface profiles, crack detection, and so on, is desirable. It is desirable to create a system that improves the prediction accuracy of deterioration along the entire route by using these values, such as vehicle detection and shape detection using LiDAR and radar installed on the road surface, and constant measurement of deflection, actual load, and vibration using strain sensors and optical fibers embedded in the road surface (see Fig. 7).

(2) Increased versatility through cloud computing

The 3D model acquired from the vehicle-mounted camera and laser and the road surface property data from the laser measurement is connected to the spatial information infrastructure and managed as a database, and it is shown that problem areas can be extracted by comparing both types of information. Although no comparison data is available at this time because of the short period of time required for the trial, we believe that the system will be able to support quick and reliable decision-making on road management by viewing this information once

the data is accumulated and compared in the future.

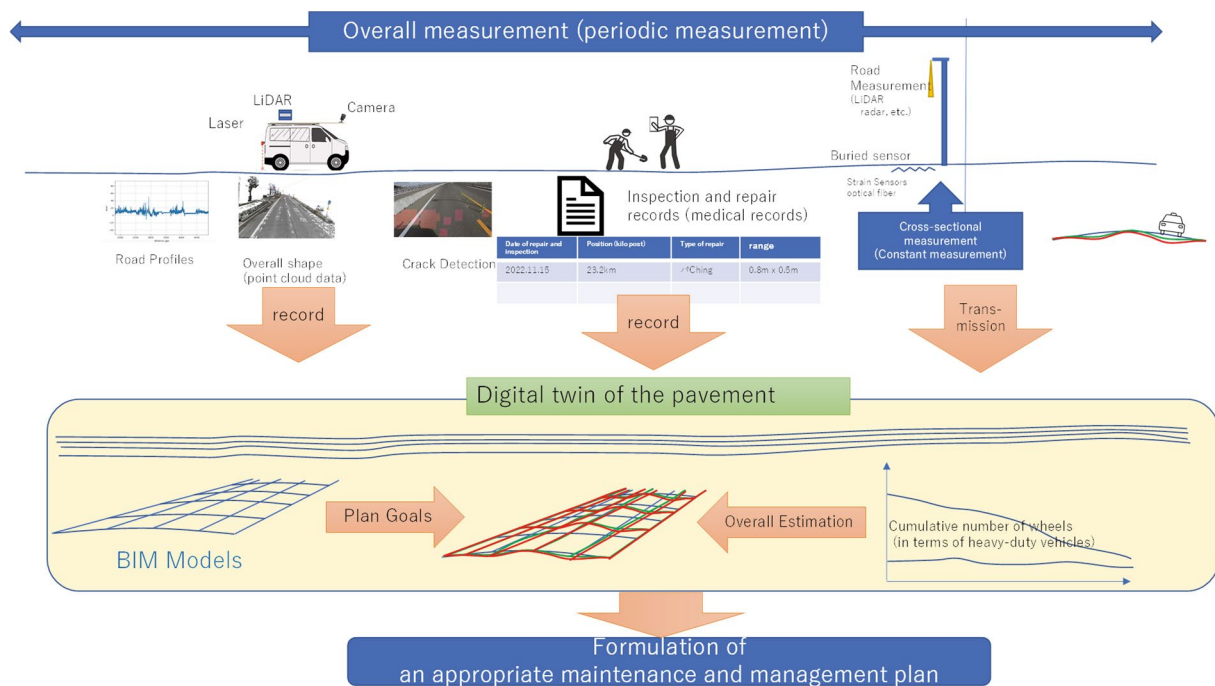


Figure 7. Maintenance management by combining overall measurement (periodic measurement) and cross-sectional measurement

6. CONCLUSIONS

In this study, for the purpose of applying digital twin to road maintenance projects, we developed a database that stores 3D shapes of road surfaces and various measurement data acquired from physical space on a spatial information system and constructed a system to support planned maintenance management using digital twin.

The database is expected to improve the accuracy of prediction of deterioration and other problems as various measurement data, inspection and repair records, etc., are accumulated through continuous operation in the future and to be helpful in the formulation of more accurate repair plans. The system was developed using QGIS, which is freeware, to make it easy for SMEs to use. However, there are some problems, such as preparing equipment individually. In the future, it is necessary to consider introducing a system that can be shared on the cloud for easier use.

ACKNOWLEDGMENTS

This research was supported by MLIT PRISM and the Construction Technology Research and Development Grant Project JPJ000094, and partially supported by JSPS KAKENHI Grant Number 24K00171.

REFERENCES

- National Institute of Standards and Technology (2013): Strategic Vision and Business Drivers for 21st Century Cyber-Physical Systems, Report from the Executive Roundtable on Cyber-Physical Systems(2013).
- Ministry of Internal Affairs and Communication, Japan (2019): Information and Communications, White paper 2019 , MIAC, Japan.
- Road Bureau, Ministry of Land, Infrastructure, Transport and Tourism (MLIT) (2016): Pavement Inspection Guideline, MLIT, Japan.
- Saddik, A. E. (2018): Digital Twins: The Convergence of Multimedia Technologies". IEEE MultiMedia 25 (2), 87–92.
- Sierra, C., Paul, S., Rahman, A., Kulkarni, A. (2022): Development of a Cognitive Digital Twin for Pavement Infrastructure Health Monitoring, Infrastructures, 7(9), 113.
- Talaghat, M. A., Golroo, A., Kharbouch, A., Rasti, M., Heikkilä, R. and Jurva, R. (2024): Digital Twin Technology for Road Pavement, Automation in Construction, 168, Part B, 105826.
- Yu, G., Zhang, S., Hu, M., Wang, Y.K. (2020): Prediction of Highway Tunnel Pavement Performance Based on Digital Twin and Multiple Time Series Stacking. Adv. Civ. Eng, 8824135, 21 pages.

DEVELOPMENT AND VALIDATION OF THE GAME ENGINE-BASED DIGITAL TWIN APP "CONNECTIA" FOR PRACTICAL USE IN THE CONSTRUCTION PHASE

Tomohide Yuasa¹, and Kodai Yamamoto²

- 1) M.S. degrees in Civil Eng. and M.S. degree in MBA, Advanced Technology Development Department, Obayashi Corporation, Tokyo, Email: yuasa.tomohide@obayashi.co.jp
- 2) Advanced Technology Development Department, Obayashi Corporation. Tokyo, Email: yamamoto.kodai@obayashi.co.jp

Abstract: CONNECTIA®, a "Digital Twin" application developed under the "Construction DX" initiative to improve productivity. It combines cloud-based data management with native app rendering using Unity, offering advanced visualization and user-friendly UI to simplify 3D workflows. It has shown potential time savings of over 10 hours per person per month.

Keywords: CONNECTIA, Digital Twin, 3D Model, BIM, CDE, Cloud, Game Engine, Unity, DX

1. BACKGROUND

The Japanese construction industry faces various challenges, including a labor shortage caused by a declining and aging workforce, long working hours, and harsh working conditions. Labor productivity in the sector has remained lower than in other industries for several decades. In response, the Ministry of Land, Infrastructure, Transport and Tourism has introduced the i-Construction initiative, aiming to improve productivity by 20% by 2025. However, to sustain productivity while observing the upper limit on overtime working hours introduced in FY 2024 under the revised Labor Standards Act, "work style reform" and further "productivity improvements" are critical challenges that the construction industry must address.

Viewing these challenges as opportunities for transformation, the construction industry has increasingly embraced the idea that digital transformation (DX), which brings innovation to construction production systems and revolutionizes production processes, offers an effective solution. Realizing DX in construction requires accurately assessing current operational processes and on-site conditions through sensing technologies and other tools. To achieve this, the authors have been developing an application named "CONNECTIA", which facilitates the creation of general-purpose digital twin models of construction sites. These models reproduce site conditions in real time by linking physical space with cyberspace. This technology, which applies various metaverse and game development technologies to the construction sector, was awarded the grand prize in the "Co-Creation Category" at CEATEC 2023, held in October 2023, in recognition of its technical excellence.

In order to solve the most important issues of "work style reform" and "productivity improvements," we call the overall efforts, including the development of new digital technologies, "construction DX" and consider them systematically. As shown in Fig. 1, we have defined the levels of construction DX, the basic level of digitalization, from DX level 1 to the final goal level 5, and are planning to achieve each level in stages. CONNECTIA, introduced in this paper, is an app for mainly achieving levels 2 and 3, which integrate and utilize digitized data on-site. Chapters 2 to 5 mainly discuss the challenges and solutions for achieving level 2, and Chapter 6 discusses the data integration platform required for level 3. The focus of this paper is mainly on level 2 in chapters 2 to 5.

In the future, we will return to level 1 and develop automatic generation of 3D data and related technologies, and we will aim to achieve level 4, which is robotics construction (construction sites where robots are the main players), such

as automating and autonomous construction machinery.

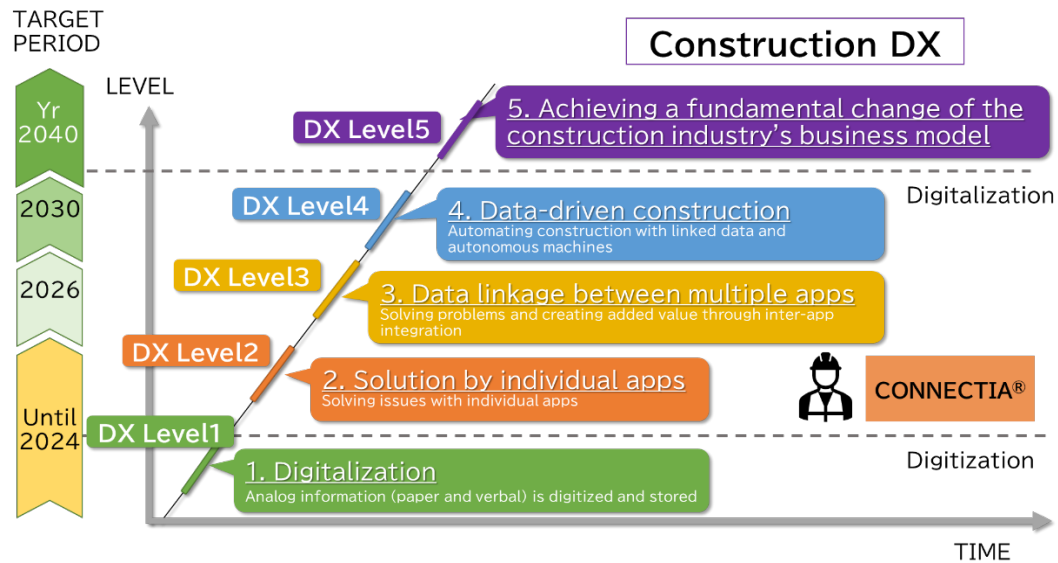


Fig. 1 Roadmap for Construction DX

2. CHALLENGES

In the construction industry, the benefits of 3D models, including BIM, likely require no detailed explanation. However, at least in the construction execution phase in Japan, these technologies have not yet been fully adopted or effectively utilized in daily operations. While there are numerous high-level factors, such as cultural and policy-related issues, two fundamental challenges stand out:

- 1) Hardware-related challenges (e.g., the need for high-spec PCs).
- 2) Software-related challenges (e.g., significant costs associated with learning and adopting new applications).

From 2022 to 2023, we conducted a study focusing on construction site offices that were reported to utilize 3D models. The findings revealed that only a few individuals used the relevant software or applications, while the majority of team members did not. Upon investigating the reasons, two key issues emerged.

First, the existing PCs at these sites lacked the specifications necessary to handle the software or applications effectively. Upgrading to PCs capable of processing 3D data would require significant time and financial investment. Second, transitioning from the traditional 2D-centric workflow to one centered on 3D models demands substantial time and resources for training and familiarization with the new software and applications. However, construction sites are generally unable to allocate the necessary time or budget for this transition. These challenges underline the difficulty of integrating 3D models into construction workflows, despite their potential benefits.

To overcome both hardware and software challenges, we developed CONNECTIA (Fig. 2), a digital twin application that allows anyone to easily manage construction work using 3D models. As there are software and applications available for handling 3D models at each stage, the scope of CONNECTIA is shown in Fig. 3. The system is designed to focus on functions specifically for the use of 3D models in construction work, making it easier to use than existing products. Additionally, it aims to create a common data environment (CDE) using various technologies that will be described later.

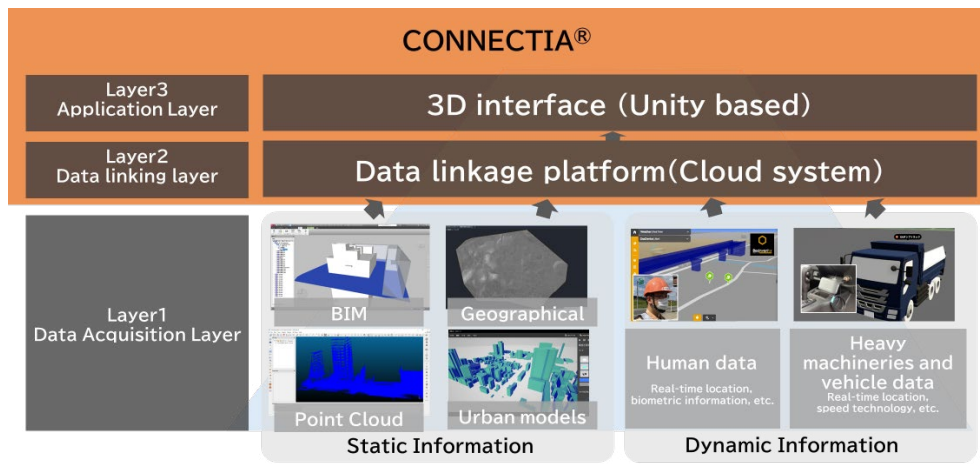


Fig. 2 Structure of CONNECTIA

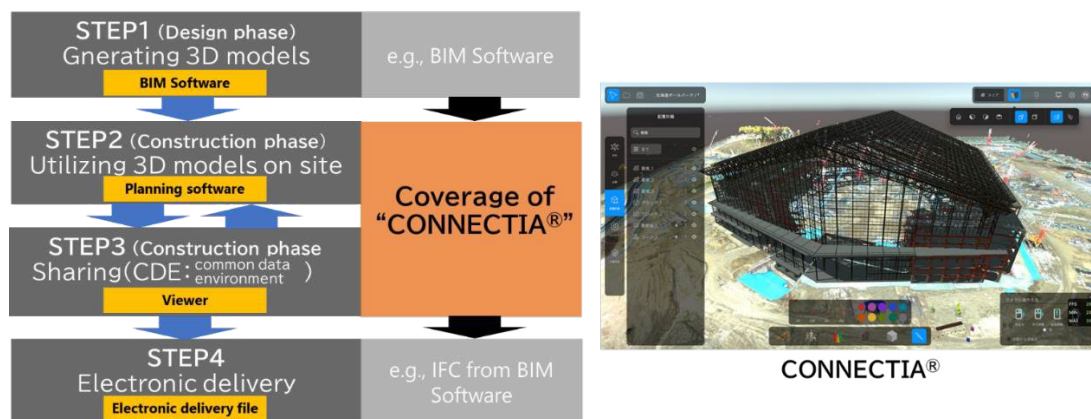


Fig. 3 Scope of CONNECTIA

3. CORE TECHNOLOGIES ADOPTED TO ADDRESS HARDWARE CHALLENGES

Recently, cloud applications that can be used directly in a web browser without installation have become mainstream, and several products that can handle 3D models have been introduced. However, for large-scale and high-volume 3D data typically used in construction, technical challenges remain, particularly in achieving high-speed rendering. This is evident in the fact that BIM software still predominantly relies on native applications. CONNECTIA combines the advantages of both approaches by adopting a hybrid system architecture: it uses cloud (AWS: Amazon Web Services) for data management and a native application for rendering (Fig. 4 - left). Since all data is stored in the cloud, a digital twin environment can be deployed anywhere as long as connection is established. The system is also designed using a microservices architecture that includes functions such as data utilization via API, authentication and authorization, and storage, ensuring the cloud's inherent availability and scalability. For 3D models, BIM data is processed for lightweight optimization in the dedicated cloud, while the native application uses a dedicated viewer, developed with the Unity game engine, for rendering. This allows the system to run smoothly even on standard PCs without a GPU (Fig. 4 - right). Details of the cloud system will be discussed in Chapter 6.

As a result, these efforts have largely addressed hardware challenges, with performance tests showing that project load times have been reduced by approximately 87%, delivering high performance compared to conventional BIM software (Fig. 5).

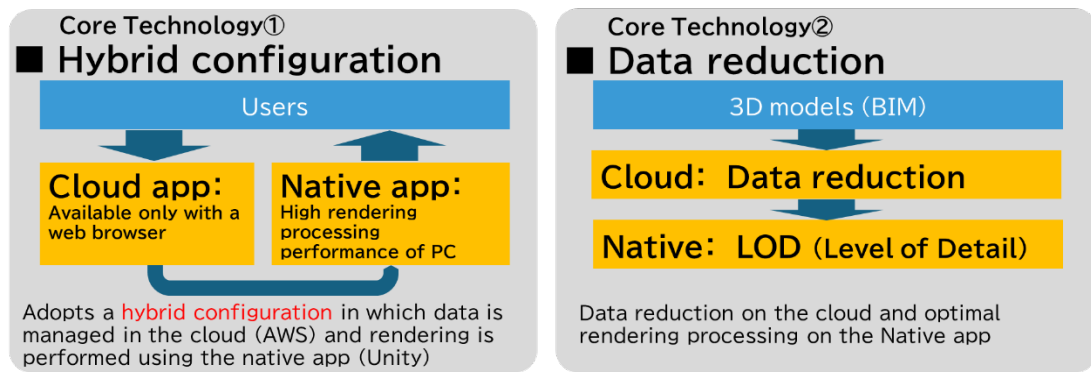


Fig. 4 Hybrid System (left) and Model Optimization Technologies (right)

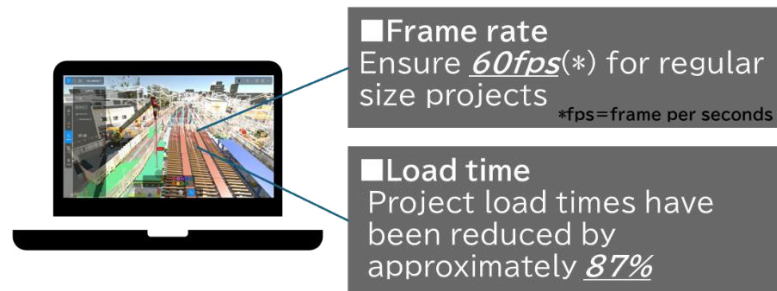


Fig. 5 Results of Validation with BIM Software

4. UI/UX DEVELOPMENT TO ADDRESS SOFTWARE CHALLENGES

Even if the hardware barriers are reduced, the implementation of on-site construction management using 3D models remains challenging if the time and cost required for application training and education are substantial. To address this, we collaborated with a UI/UX development team to repeatedly refine and improve the software's user interface and experience, focusing primarily on resolving software-related challenges.

Initially, we analyzed the features and operations included in existing BIM softwares and applications, then significantly narrowed down the functionality to those essential for the construction phase. This was necessary because many features in existing software were either irrelevant or overly sophisticated for use during this phase. For the selected features, we developed UI mockups from scratch using the UI design tool *Figma*. Subsequently, we conducted a "blind test" involving 14 participants—seven with construction experience and seven without (having IT backgrounds). In these tests, participants were instructed on the task they needed to accomplish (e.g., placing a object of heavy machinery at a specific location) but were given no guidance on how to operate the application. Their interactions were recorded using video cameras for analysis. By examining the results, we identified areas where users encountered difficulties and investigated the reasons for these challenges through follow-up interviews. Based on these findings, we iteratively revised and enhanced the UI to create a more intuitive and effective user experiences (Fig. 6).

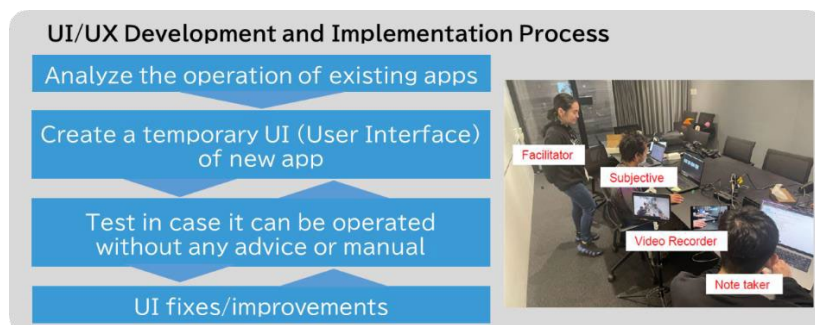


Fig. 6 UI/UX Development and Implementation Process

4.1. VALIDATION METHOD

We compared the usability of our UI with that of a conventional 3D viewer application. For the comparison, we created two scenarios: Work Case ① and Work Case ②, based on a series of tasks frequently performed in the construction field. After preparing integrated data that included BIM models, point cloud data, and terrain models, we configured Work Case ① with the task of placing four heavy machinery models (a hydraulic excavator, a crane, and a crane in two different orientations). Work Case ② involved the task of placing four boxes of different sizes. The operation time and the number of steps required to complete the tasks using the conventional application and ours were then measured and compared under identical conditions using standard engineering notebook PCs (Fig. 7)

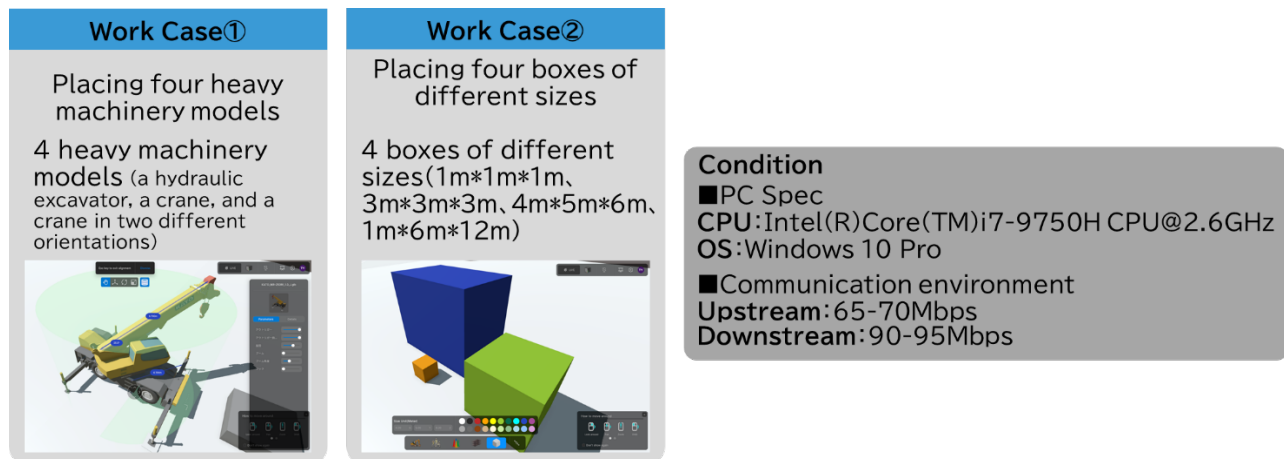


Fig. 7 UI Validation Method

4.2. VALIDATION RESULTS

The validation results are shown in Fig. 8. In Work Case ①, the conventional procedure, where the coordinates of the machine models created during data integration did not match the field coordinates, required rough repositioning of the machinery, followed by fine-tuning of the horizontal position and height. In contrast, our UI required only selecting the model from the heavy machinery list and clicking the location to complete the task. When placing the crane model in different orientations, the conventional procedure involved editing the model in 3D CAD software and importing it into the viewer application. However, with our UI, this entire operation could be completed within the application. As a result, our UI took only 132 seconds (18 steps) to complete the task, compared to 1,523 seconds (40 steps) using the conventional approach, reducing the time to less than one-tenth.

In Work Case ②, the conventional procedure took 557 seconds (48 steps), as it required model creation and coordinate adjustments using 3D CAD software, whereas our UI completed the task in 115 seconds (10 steps), reducing the time and steps to about one-fifth. These results indicate a significant improvement in operation time and steps compared to conventional BIM software, addressing the challenges primarily on the software side.

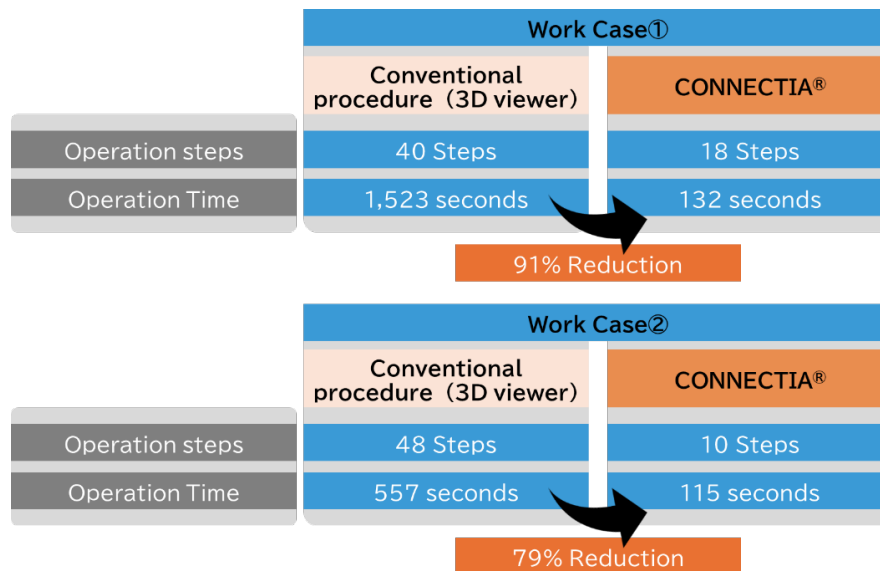


Fig. 8 UI Validation Results

5. VALIDATION OF CONNECTIA IN REDUCING CONSTRUCTION MANAGEMENT WORK

Through a series of development processes from FY 2022 to FY 2023, CONNECTIA has addressed certain hardware and software challenges, significantly lowering the barriers for users. However, since the use of 3D models in construction and on actual work sites was still in the experimental phase, it was necessary to further test CONNECTIA in real construction management settings to evaluate how effective it is in streamlining management tasks. The methods used in this assessment and the results are detailed below.

5.1. MAIN FUNCTIONS INSTALLED IN CONNECTIA

Prior to the validation, we will introduce some of the functions that have already been implemented and that are necessary during the construction stage based on interviews with users. These functions are categorized into three groups: machinery model placement, construction review, and real-time functionality. The main functions from each group are listed in Fig. 9



Fig. 9 Main Functions of CONNECTIA (by Function Group)

5.2. ON-SITE VALIDATION METHOD

To evaluate whether CONNECTIA, developed between FY2022 and FY2023, could be effectively utilized on construction sites, a field validation study was conducted during the first half of FY2024. The primary focus of this evaluation was not only to assess the functionality and completeness of CONNECTIA but also to identify the challenges faced by the sites and measure how effectively these challenges could be addressed. The evaluation also considered processes outside the scope of CONNECTIA, such as model creation, model uploads, and the preparation of materials for client presentations. For instances where CONNECTIA's current capabilities were insufficient, we also conducted interviews to explore what additional features or support systems would be necessary to complete these processes. The field trials followed a structured process: 1) Identifying on-site challenges, 2) Setting up a 3D utilization environment, 3) Trial implementation, and 4) Evaluation (via surveys)

Particular emphasis was placed on the challenge-identification phase, which required significant time. To minimize the burden on site offices, the setup of the 3D utilization environment was handled almost entirely by the headquarters team. This approach ensured consistent conditions for comparison and evaluation with traditional 2D workflows.

The study involved 14 operational civil engineering construction sites. Evaluations and interviews were conducted to gather insights and feedback from these sites. The evaluation was based on a questionnaire (shown in Table 1), with results organized into four rating categories: ◎ (very good), ○ (good), △ (fair), and × (poor). The main respondents of the survey were "Site Managers" (about 27% of all field staff) and "Chief Engineers" (about 12%), primarily representing the younger demographic, accounting for approximately 39% of the on-site supervisors.

5.3. RESULTS OF ON-SITE VALIDATION

As an overall assessment, 71% of the 14 surveyed sites responded that they expect a reduction of 10 or more hours per person per month in the near future by using CONNECTIA. Regarding specific reasons, 36% of the sites answered that the "work required for construction review and preparation of construction plans (TYPE A)," would be reduced by 30-50%, while 43% expect a 10-30% reduction, compared with the conventional approach using 2D drawings. Similarly, for the "prevention of omissions in the review process and reworking (TYPE B)," 43% of the sites expect a 30-50% reduction, while 50% expect a 10-30% reduction. As for "expediting the consensus-building process among stakeholders by promoting mutual understanding (TYPE C)," 43% expected a 30-50% reduction, and another 43% expected a 10-30% reduction. Regarding "reducing the workload of preparing explanatory materials during construction reviews (TYPE D)," 54% answered that there was a reduction of 10% or more. In addition, 38% answered "no change," however we believe that if they can create 3D explanatory materials, which handle an increasing amount of information, in the same amount of time as traditional 2D work, users will be able to improve the quality of explanatory materials without increasing the burden. Furthermore, in the future, we plan to continue development with a view to speeding up work compared to traditional 2D work by incorporating assistance, such as automatically arranging models of safety equipment and guards that need to be installed in the vicinity in conjunction with the placement of cranes. (Fig. 10).

The challenges faced at each site varied widely, and due to space limitations, detailed descriptions cannot be provided here. However, it became evident that compiling the types of challenges encountered and the solutions implemented as use cases would be valuable for scaling the application to other sites. Additionally, the study revealed that the significant workload associated with preparatory tasks such as model creation and uploading highlighted the critical importance of establishing support systems to drive broader adoption.

TYPE (A)	work required for construction review and preparation of construction plans	<p>Compared to traditional construction planning using 2D drawings, to what extent do you think the workload of construction planning (construction planning) work will increase or decrease when using CONNECTIA? Please rate on a 4-point scale: ◎, ○, △, ×.</p> <p>Based on the amount of time previously spent on construction planning, by what percentage will it be reduced?</p> <p>◎:30-50% reduction (imagine completing 10 hours of study and discussion in 5 hours) ○:30% or less reduction (imagine completing 10 hours of study and discussion in 7 hours) △:No change ×:Takes longer</p>
(B)	prevention of omissions in the review process and reworking	<p>Compared to traditional construction planning using 2D drawings, to what extent do you think CONNECTIA can prevent oversights and rework? Please rate on a 4-point scale using ◎, ○, △, or ×.</p> <p>◎:Greatly prevented (Imagine overlooking something being reduced by 30% or more) ○:Slightly prevented (Imagine overlooking something being reduced by 10% or more) △:No change ×:Increased rework</p>
(C)	expediting the consensus-building process among stakeholders by promoting mutual understanding	<p>When creating explanatory materials for customers and partner companies, how much do you think your workload will increase or decrease by using CONNECTIA compared to traditional 2D models? Please rate on a 4-point scale: ◎, ○, △, ×.</p> <p>Based on the amount of time you've spent on reviewing the work up until now, by what percentage will it be reduced?</p> <p>◎:30-50% reduction (imagine completing 10 hours of study and discussion in 5 hours) ○:30% or less reduction (imagine completing 10 hours of study and discussion in 7 hours) △:No change ×:Explanation takes longer</p>
(D)	reducing the workload of preparing explanatory materials during construction reviews	<p>When explaining construction plans (creating construction step diagrams, etc.) to customers and partner companies using conventional 2D drawings, how much do you think it would be possible to improve the other party's understanding of the situation and, as a result, to what extent would it be possible to speed up or simplify discussion times? Please rate on a 4-point scale using ◎, ○, △, ×.</p> <p>◎:30-50% reduction (imagine completing 10 hours of study and discussion in 5 hours) ○:30% or less reduction (imagine completing 10 hours of study and discussion in 7 hours) △:No change ×:Explanation takes longer</p>

Table 1 Questionnaire Sheet

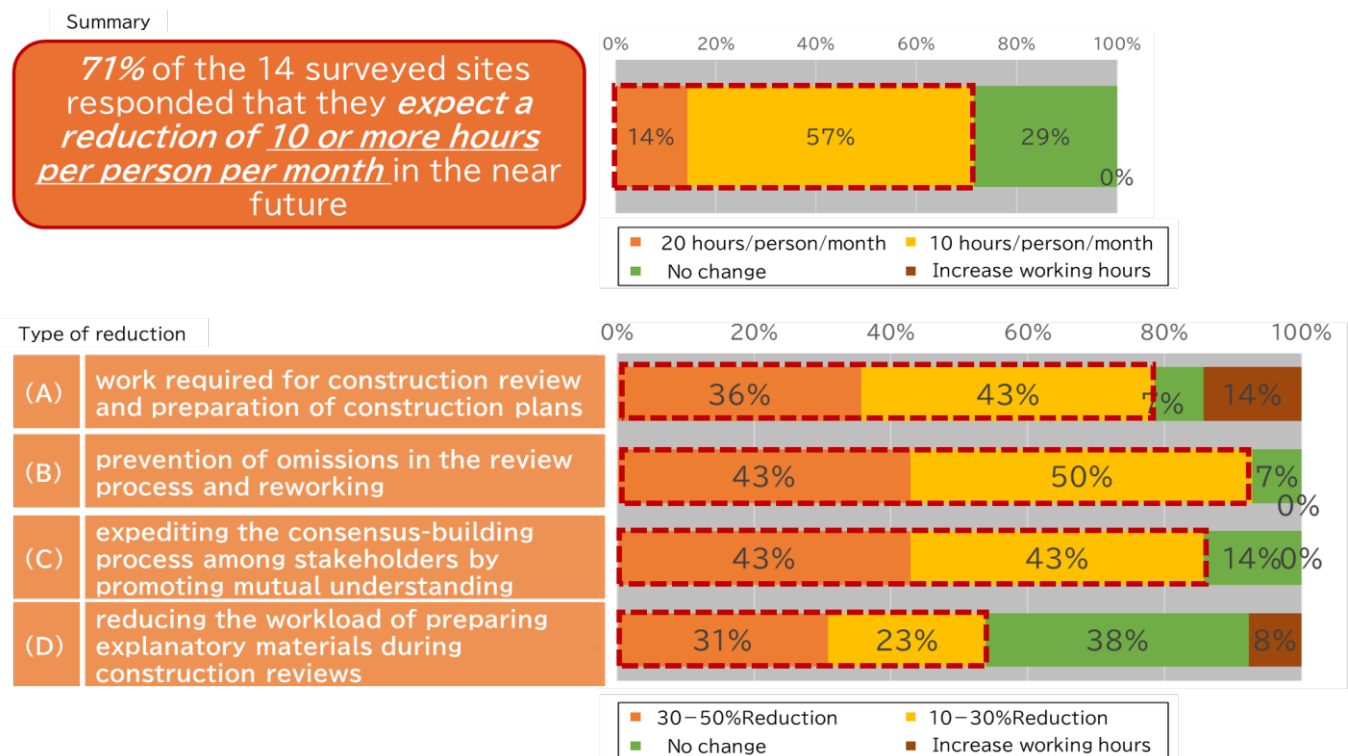


Fig. 10 Results of Effect Validation

6. DATA LINKAGE PLATFORM

In the previous chapter, we discussed how to use CONNECTIA to solve site issues, mainly in relation to level 2 of construction DX. In this chapter, we will explain the backend of CONNECTIA, with the realization of level 3 in mind. Constructing a digital twin requires to handle both "static information" with little time variation, such as 3D models and point cloud data used in BIM received at the start of design and construction, and "dynamic information" with large time variation, including the position data of moving objects like field workers, heavy machinery, and vehicles, as well as sensing information from the surrounding environment. Since we are also building an environment for data linkage platform that integrates these types of information to aggregate, convert, link, store, and render data in various formats, we provide below a brief description of the platform.

6.1. OUTLINE OF THE PLATFORM

The data linkage platform currently under development is a system that collects data from various sources, including internal systems, external services, and cloud platforms, and stores it in an integrated data repository. Additionally, it has a function to convert data with different formats and protocols into an interoperable format so that it can be uniformly processed by the platform. This ensures data consistency while minimizing the load on resources even when processing large volumes of data, enabling effective data utilization.

To minimize vendor dependency, the system uses FIWARE, a globally renowned open-source software (OSS) product commonly used as an "Urban OS", as its core component. Furthermore, it is structured by combining microservices that provide various functions, such as linking applications used in the construction sector, utilizing data via APIs, authentication and authorization, and storage. The data linkage platform leverages AWS, offering high availability and scalability (Fig. 11).

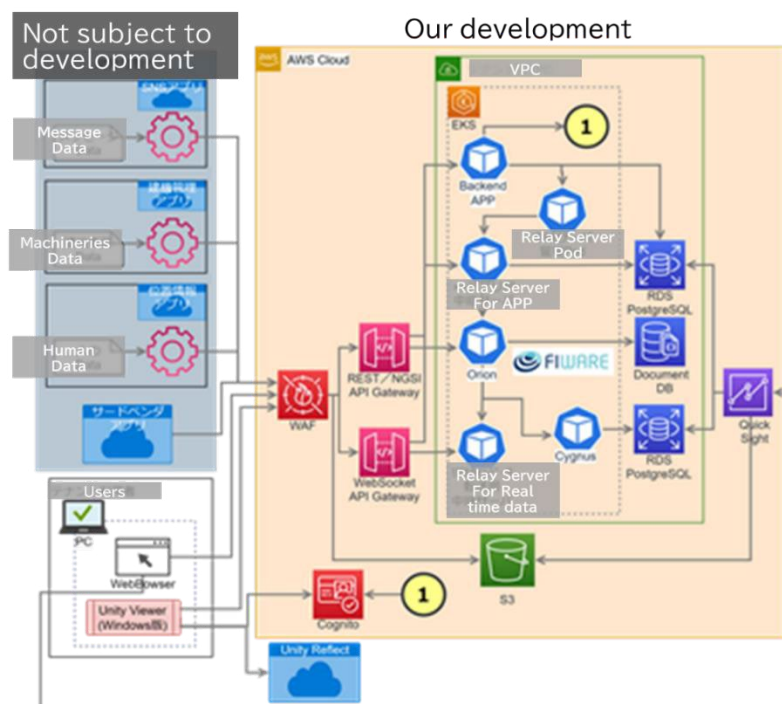


Fig. 11 Data Platform Configuration of the System

6.2. DATA STRUCTURE USED IN CORE PART

This system utilizes FIWARE and NGSI as the data structure for linking applications in the core part of the data linkage platform. NGSI is an open international standard for network APIs, originally developed in Japan. It was standardized by the

Open Mobile Alliance in 2010 and has been continuously updated since then. NGSI serves as an interface for linking data between different applications. The data model mainly consists of "entities" and "contexts," which are standardized data (attributes and metadata) representing the essential attributes of objects. This format ensures data interoperability and generality between different applications. Fig. 12 shows an example of a data model in NGSI format. This system uses the stable version NGSI-v24.

Additionally, this system provides APIs for master data managed in an Relational data base (RDB), which is not intended for application linkage, as well as for historical data used in Business Intelligence (BI) analysis and their combined view definitions. Almost all data operations—registration, retrieval, updating, and deletion—can be performed via the APIs. Moreover, both the NGSI API and the above-mentioned API are designed with flexible authentication and authorization settings. All necessary APIs for data linkage are planned to be made available to application vendors, so it is expected that application vendors will develop their own data linkage and utilization solutions in the future.

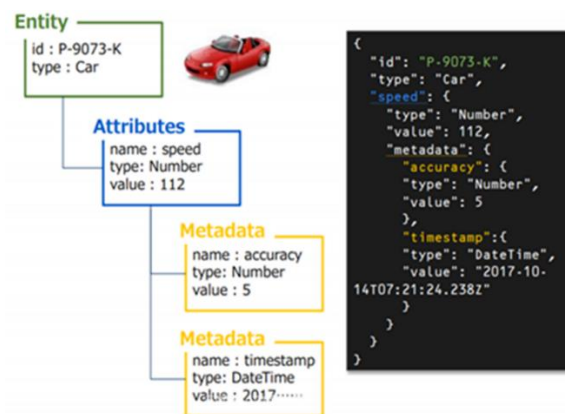


Fig. 12 Example of Data Model in NGSI Format

7. SUMMARY AND FUTURE DEVELOPMENTS

CONNECTIA adopts a hybrid system configuration, using the cloud for data management and a native application for rendering. Through the use of game engine-specific rendering processes, it effectively resolves hardware challenges. Additionally, by focusing on UI/UX development, it addresses software challenges, lowering the barriers to adoption. As a result, 71% of trial sites indicated that they expect a reduction of 10 hours/person/month or more in the near future, thanks to the implementation and improvement of functions specialized for various construction management tasks. We see CONNECTIA as one of the solutions to various issues such as labor shortages due to a declining and aging workforce, long working hours, and difficult working conditions. Moving forward, we will continue to explore on-site needs, identify and implement highly useful functions, strengthen the data linkage function with the goal of creating a CDE, systematize the operation of the app, and build an organizational structure to further promote its use.

REFERENCES

- Aki, M., Tomohide Y., Satoshi Y. (2023). Realization of a general-purpose digital twin during the construction stage - Development of a CPS construction management system (in Japanese), AI and Data Science Papers, 4(2), 89-96.
- Tomohide, Y., Satoshi Y., Akira, N., Susumu Y., and Ryota M. (2023). Development of a data platform for constructing a digital twin during the construction stage (in Japanese), AI and Data Science Papers, 4(3), 924-931.
- Satoshi Y., Akira N., and Susumu M. (2024). Development of a platform for constructing a digital twin (in Japanese), Monthly "Clean Technology" May issue.

Design Strategies for Enhancing Spatial Vitality through the Integration of Sports Venues and Commercial Spaces

Xuefei Yuan^{1,2,†,*}, Benchen Fu^{3,4,†}, Shan Man⁵, and Yuanyuan Wei⁶

1) Ph.D., Assoc. Prof., College of Civil Engineering and Architecture, Jiaxing University, Jiaxing, China. Email: 277312039@QQ.com

2) Ph.D., SN ENGR, Shanghai Cuiyunyuan Architectural Design Consulting Co., Ltd. Shanghai, China.

3) Ph.D., Prof., School of Architecture and Urban Planning, Shenzhen University, Shenzhen, China. Email: benchen_fu@126.com

4) Ph.D., Prof., Shenzhen Key Laboratory of Architecture for Health & Well-being (in preparation), Shenzhen, China.

5) Ph.D., Asst. Prof., College of Civil Engineering and Architecture, Jiaxing University, Jiaxing, China. Email: 408059530@QQ.com

6) Ph.D., Asst. Prof., College of Civil Engineering and Architecture, Jiaxing University, Jiaxing, China. Email: 00007652@zjxu.edu.cn

† These authors contributed equally to this work.

* Correspondence: 277312039@QQ.com,

Abstract: With the evolving functional attributes of sports venues in China and the promotion of national fitness strategies, sports venues are increasingly transitioning into comprehensive entertainment and leisure facilities. However, due to spatial constraints and other factors, many of these facilities suffer from suboptimal operational effectiveness and limited spatial vitality. This study employs literature review, case data collection, comparative analysis, and inductive analysis to propose a design strategy for the integration of sports venues and commercial spaces, emphasizing the "optimized horizontal space utilization, vertical compound utilization, and scene content creation." Through specific case discussion, the study emphasizes the need for these three strategies to work in synergy and highlights the importance of continuous optimization across the entire building lifecycle—from decision-making, strategic planning stages to design, construction, operation, maintenance, and feedback optimization. These findings provide a valuable reference for the organic integration of sports venues and commercial spaces in China and beyond, fostering new vitality in sports venues and promoting healthier urban environments.

Keywords: Sports venues, Commercial spaces, Spatial vitality, Design strategies

1. INTRODUCTION

Over several decades, sports architecture in China has evolved from serving primarily as landmark structures for major sporting events to becoming facilities that prioritize public access and enhance urban quality of life (Mo & Chen, 2020). This transformation is driven by government policies aimed at increasing public participation in sports activities and expanding the availability of sports venues (Zhang et al., 2011). With the growing number of people engaged in physical activities, the functional demands for sports architecture have diversified, shifting from a "sports+" model to a "+sports" approach. In this shift, the roles of sports venues and commercial spaces have been redefined, with many sports venues increasingly transforming into multi-functional leisure and entertainment complexes.

However, due to traditional sports venues' limited consideration of commercial integration and the lack of commercial operation experience in post-event scenarios, significant spatial resources remain underutilized, often resulting in suboptimal facility performance and, in some cases, even partial abandonment. This can negatively impact the surrounding urban environment (Turner & Rosentraub, 2016). Although previous research has highlighted the importance of post-event operations and the integration of commercial spaces (Ma & Zhuang, 2004; Luo & Mei, 2006), there has been little in-depth, systematic research specifically focusing on design strategies aimed at enhancing spatial vitality to boost both venue utility and regional urban vibrancy.

With the advancement of the "Healthy China" and National Fitness strategies, sports consumption has become an increasingly important part of the broader consumer market (National Development & Reform Commission et al., 2024). Recent statistics indicate that China's sports consumption market is currently valued at approximately RMB 1.5 trillion, with projections to grow to RMB 2.8 trillion by 2025 (Xinhuanet Client, 2023). In light of this, there is an urgent need to explore effective methods for integrating sports venues with commercial spaces, thereby upgrading both sports facilities and their surrounding urban consumption environments. One of the critical factors for successful sports-commercial integration is spatial vitality. Active and vibrant spaces not only foster greater sports participation but also attract more consumers, enhancing commercial atmosphere and economic returns, and thereby invigorating the entire area. This study aims to provide insights into enhancing spatial vitality through a comparative analysis of representative cases domestically and abroad, synthesizing models to outline design strategies that support the integration of sports and commercial spaces from three dimensions: optimized horizontal space utilization, vertical compound utilization, and scene content creation.

2. METHOD

This study employs a mixed-methods approach (Bryman, 2016), integrating literature review, case data collection, comparative analysis, and inductive synthesis to comprehensively examine the integration of sports and commercial spaces.

2.1 Data Collection

This study conducts a thorough review of literature to collect data on sports venues built or planned globally and domestically since 1990, resulting in a dataset of approximately 100 cases. These cases include information on location, urban context, construction date, building area, seating capacity, overall layout, floor plans, and where available, data on revenue composition and business configurations, forming the foundational research dataset. No limitations were placed on venue size, ensuring a comprehensive basis for subsequent model synthesis. From this dataset, around 30 prominent and influential sports venues with notable integration of commercial spaces were selected as key cases for in-depth analysis, focusing on spatial characteristics, factors influencing spatial vitality, and primary design strategies to establish a robust research sample database.

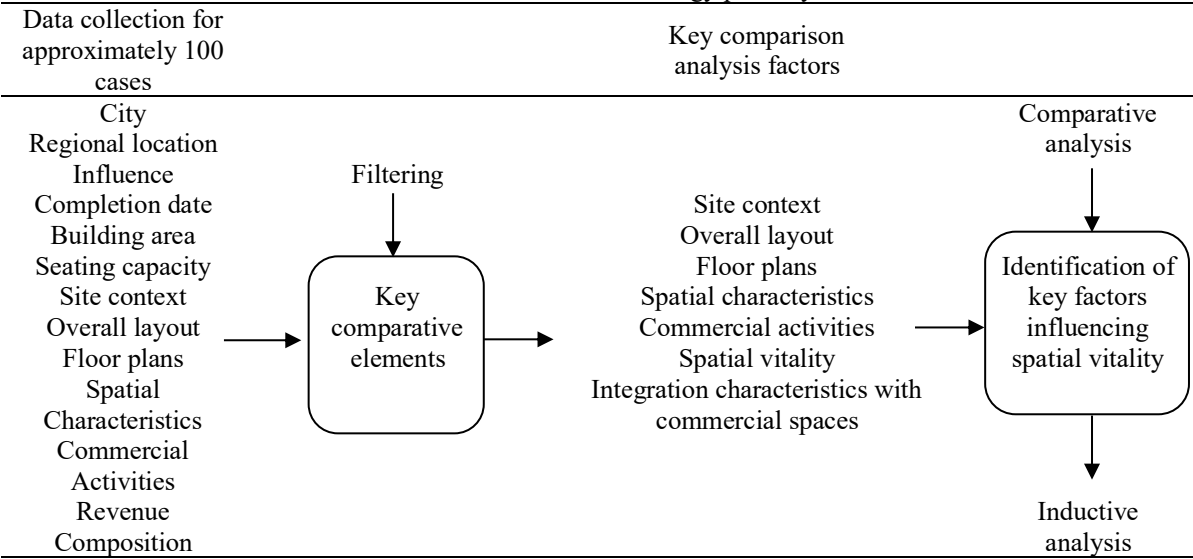
2.2 Comparative Analysis

Based on the research sample database, 30 completed sports venues with successful post-event operations were examined in detail. Key factors for comparison included site context, overall layout, floor plans, spatial characteristics, commercial functions, spatial vitality, and integration features with commercial spaces. This comparative analysis identifies the critical spatial factors that enhance the fusion of sports and commercial spaces, contributing to improved spatial vitality.

2.3 Inductive Analysis

Following the comprehensive comparative analysis of the sample database, design strategies for enhancing the integration of sports venues and commercial spaces with a focus on spatial vitality are synthesized. First, sports venues with ample available land should prioritize efficient use of horizontal space. Second, those with limited land should focus on maximizing vertical compound space utilization. Third, beyond spatial optimization in both dimensions, attention should be given to content-driven scene creation. These three elements – horizontal space, vertical compound use, and content creation – are interdependent, collectively shaping a vibrant public activity space.

Table 1. Research methodology pathway



3. RESULTS





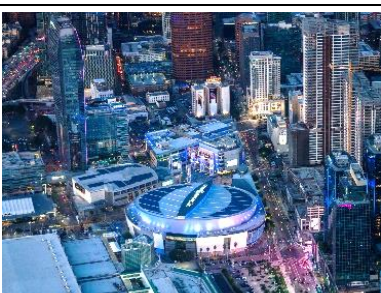
Sports venues are generally developed on two primary types of land use scales in varying spatial and geographic contexts: compact sites and expansive sites (Wei et al., 2023). Integration of sports venues with commercial spaces should be tailored to the specific land scale, maximizing the use of multi-dimensional site space and interior architectural space. This approach aims to create a dynamic sports complex characterized by "optimized horizontal space utilization, vertical compound utilization, and scene content creation," establishing vibrant spaces that serve both the city and its citizens. The main design strategies for integrating sports venues and commercial spaces with the goal of enhancing spatial vitality include the following three components:

3.1 Optimal Horizontal Space Utilization of Site Planning for Commercial Integration

Sports venues located on expansive urban sites often feature large transitional areas between the venue and surrounding city spaces, creating a spatial buffer from city streets. For such venues, integration with

commercial spaces should emphasize effective site planning to make full use of the horizontal dimensions. This can involve incorporating a variety of sports facilities, natural ecological environments, or even outdoor theme parks, commercial complexes, or shopping streets within the site. Examples of this approach include the "sports venue + multi-sport facility" model, the "sports venue + ecological environment" model, the "sports venue + theme park" model, and the "sports venue + commercial complex or shopping street" model (Yuan et al., 2023). These configurations extend the range of facilities accessible to the public, creating diverse, engaging areas for city-wide sports, leisure, and entertainment, thus establishing a rich urban sports and recreation district (Table 2).

Table 2. Optimal horizontal space utilization of site planning for commercial integration

Project	Real/Rendering image	Model analysis	Strategy description
Kai Tak Sports Park, Hong Kong		Sports Venue + Multi-Sport Facilities Model	<ol style="list-style-type: none"> 1. Analysis of the sports and consumer characteristics of surrounding demographics. 2. Optimized overall layout to maximize site utilization with rational zoning. 3. Strategic placement of sports facilities and additional functional areas.
Quzhou Sports Park, China		Sports Venue + Ecological Environment Model	<ol style="list-style-type: none"> 1. Analysis of natural environmental features. 2. Analysis of sports and consumer characteristics of nearby populations. 3. Maximized use of topography, landscape, and water features for sports facilities.
Tokyo Dome City, Japan		Sports Venue + Theme Park Model	<ol style="list-style-type: none"> 1. Analysis of sports and consumer characteristics of surrounding demographics. 2. Planning of entertainment theme and functional zoning. 3. Strategic configuration of theme park zones to optimize site use.
Sport Society, Dubai		Sports venue + Commercial Complex Model	<ol style="list-style-type: none"> 1. Analysis of sports and consumer characteristics of surrounding demographics. 2. Planning and design of the sports venue and commercial complex combination. 3. Strategic placement of functions to enhance synergy between the complex and sports venue, driving regional consumption upgrades.
L.A. Live, Los Angeles, USA		Sports Venue + Shopping District Model	<ol style="list-style-type: none"> 1. Analysis of sports and consumer characteristics of surrounding demographics. 2. Planning and design of the overall combination of shopping district and sports venue. 3. Strategic configuration of functions to enhance synergy between the district and sports venue, driving regional consumption upgrades.

(1) Diverse Sports Facilities Tailored to User Demographics

Sports venues on expansive urban sites often have large land areas, yet historically these spaces have lacked efficient, intensive land use. In China, large sports centers are common; however, many experience underutilization or even vacancy after events. To maximize these sites, a variety of sports facilities and functional spaces should be planned to meet the activity preferences and consumption characteristics of diverse surrounding user groups. Integrating additional outdoor sports areas, plazas, and leisure spaces can help boost operational revenue, creating a multi-use, dynamic urban sports district (Shi & Li, 2020).

For example, Hong Kong's Kai Tak Sports Park features multiple sports fields, urban leisure areas, park facilities, retail outlets, and dining options around its main stadium, providing high-quality facilities for professional athletes, amateur enthusiasts, and the public. This design establishes the venue as a world-class activity hub for the city. Similarly, Shanghai's Xujiahui Sports Park utilizes outdoor spaces between the main stadium and city streets, introducing lush aerobic parks and diverse sports areas connected by a running track, greatly fulfilling the fitness and recreational needs of city residents.

(2) Integrating Natural Environments to Create Ecological Parks

With a growing emphasis on preserving natural environments, the overall layout of sports venues should integrate with surrounding ecological elements, reducing the potential negative impact of large-scale sports architecture on urban spaces (Fan et al., 2022). Integrating sports venues with commercial spaces should also take full advantage of natural features, such as terrain, landscape corridors, water bodies, and wetlands. Thoughtful placement of outdoor sports facilities alongside vibrant commercial spaces allows people to engage in both recreational and sports activities within natural surroundings, fulfilling their desire to connect with nature (Zhuang et al., 2020). For instance, Qucheng Sports Park in Quzhou, China, spans a generous site where multiple sports venues blend with natural landscapes. Between the sports facilities and the city streets, artistic landscapes resembling volcanic formations and mirrored lakes have been created. This sports-focused ecological park offers various facilities, such as tracks, badminton and basketball courts, climbing walls, cycling paths, roller skating areas, and jogging tracks, alongside retail and dining spaces, forming a vibrant urban space that merges sports, fitness, commerce, and natural ecology.

(3) Introducing Theme Parks to Energize Urban Spaces

The integration of theme parks within sports venues provides a dynamic “sports + entertainment” theme that attracts a broader audience, particularly youth and families, and encourages active participation. This approach not only enhances event attendance but also creates a social hub for activities that cater to younger audiences. Tokyo Dome City exemplifies this model with a theme park, Korakuen Hall, LaQua spa facilities, restaurants, and hotels surrounding the main stadium, transforming the area into a vibrant urban center that combines sports, recreation, and leisure, attracting 35 million visitors annually. By creating a comprehensive space that combines sports, entertainment, and theme park attractions, sports venues transcend their traditional roles, becoming interactive and entertainment-rich urban landmarks. This design draws a wider audience, turning the venue into a multifunctional destination for viewing, participating, and socializing, thereby activating urban vitality. In addition to sports and entertainment, the area also serves as a community gathering space, enhancing regional cohesion and promoting the development of multiple industries within the area.


(4) Integrating Commercial Complexes to Foster a Community Environment

Introducing commercial complexes or shopping districts around sports venues is another effective approach to boosting regional vibrancy. In this model, the sports venue is either integrated within or closely connected to a commercial complex or shopping district, creating a symbiotic relationship that functions as a multi-use community hub. Such integration not only establishes the sports venue and commercial area as iconic urban spaces that strengthen the city's brand but also drives local economic growth. For example, Dubai's Sport Society integrates sports facilities with a commercial complex, creating a multifunctional area that combines sports, retail, and wellness, with sports as its defining feature. This blend cultivates a unique atmosphere and sense of belonging for the community, energizing the area and fostering a vibrant ambiance that encourages physical activity among all age groups. Similarly, L.A. Live in Los Angeles surrounds the Crypto.com Arena (formerly the Staples Center) with a dense network of shopping, dining, hotels, apartments, museums, and office spaces. This combination transforms the area into a significant urban gathering point, where visitors attending major events at the arena also patronize nearby businesses, boosting foot traffic in the surrounding shopping district and significantly enhancing its vitality, serving as a catalyst for regional economic development.

3.2 Utilization Vertical Compound Utilization for Commercial Integration

For sports venues on compact sites within dense urban areas, spatial constraints limit horizontal openness but provide closer integration with the city. In such cases, integrating commercial spaces should prioritize an efficient, compact arrangement of vertical spaces to create a multi-level, interactive commercial interface, enhancing vertical engagement and spatial entertainment value. Effective use of underground space can expand the venue's functional areas and attract foot traffic; multi-level utilization of above-ground spaces optimizes spatial efficiency, reducing potential wastage; strategically designed VIP areas improve service quality; and innovative use of rooftop spaces adds new elements to both the venue and urban environment (Table 3).

Table 3. Vertical compound utilization for commercial integration

Project	Real/Rendering image	Model analysis	Strategy description
Mercedes-Benz Arena, Shanghai		Underground + Above-Ground + VIP Area + Rooftop Commercial Integration	<ol style="list-style-type: none"> 1. Strategic planning of functions and revenue models. 2. Analysis of available space on each level. 3. Maximal use of underground space to attract foot traffic and rooftop for scenic value. 4. Multi-level integration to create a comprehensive center.
Barclays Center, New York, USA		Underground + Above-Ground + VIP Area Commercial Integration	<ol style="list-style-type: none"> 1. Analysis of spatial interaction between the venue and city streets. 2. Creation of an attractive street-corner entrance. 3. Evaluation of available space on each level. 4. Effective integration of commercial functions across multiple levels.
Crypto.com Arena, Los Angeles, USA		Above-Ground + VIP Area + Rooftop Commercial Integration	<ol style="list-style-type: none"> 1. Analysis of primary renovation focus—enhancing club seating and suite features. 2. Evaluation of available space on each level. 3. Creative utilization of all levels. 4. Unique suite and VIP areas within lower seating areas, seating zones, and rooftop.
O2 Arena, London, UK		Above-Ground + VIP Area + Rooftop Commercial Integration	<ol style="list-style-type: none"> 1. Analysis of key renovation areas to enhance VIP offerings within the existing sports-commercial complex. 2. Replacement of select suites with expanded member experience space, The Residence. 3. Addition of retractable cantilevered walkway.
Shenzhen Bay Sports Center, China		Above-Ground + Rooftop Commercial Integration	<ol style="list-style-type: none"> 1. Integration of commercial spaces with maximum utilization of scenic value. 2. Creation of the “Sea Gate” viewing area. 3. Inclusion of aerial commercial spaces for award ceremonies, high-end receptions, client appreciation events, and brand launches.

(1) Effective Utilization of Underground Space

In contemporary sports venue design, the effective use of underground space has become a key element for enhancing both overall venue utilization and foot traffic, especially in venues that connect directly with metro systems. Underground spaces not only provide additional functional areas but also create unique consumer experiences, drawing in more visitors. For example, at the Mercedes-Benz Arena in Shanghai, the B2 level features an open ice rink, while the B1 level hosts a distinctive underground food court and music theater, successfully attracting metro passengers and increasing the venue’s vibrancy and commercial value. Similarly, at the Barclays Center in New York, the basketball court and training facilities are located underground, allowing the building height to blend into the urban landscape while creating a captivating street-level entrance that enhances the urban vitality of the venue through a dramatic visitor experience.

(2) Multi-Level Utilization of Above-Ground Space

The multi-level utilization of above-ground space is a key strategy to enhance the vibrancy of sports venues. This approach not only optimizes physical space usage but also diversifies spatial functions, reducing the occurrence of "large but empty" spaces. Sports venues can assign functions based on specific height levels, such as parking on platforms at 4 meters, restaurants and training gyms at 6 meters, basketball and soccer courts as well as large recreational areas at 9 meters, and badminton courts and cinemas at 12 meters. Additional facilities like cinemas, children's play areas, and fitness centers can be incorporated based on spatial characteristics, enhancing the venue's ability to serve a diverse public.

Function distribution across vertical space should also align with the venue's thematic features to create an attractive, multi-layered environment. For instance, at the Mercedes-Benz Arena in Shanghai, different levels host music clubs, VIP suites, and restaurants, allowing 24/7 operation and establishing the venue as a key cultural and entertainment landmark in Shanghai (Toutiao, 2019). The Suzhou Industrial Park Sports Center includes parking, restaurants, cinemas, and training gyms on the sub-2nd floor, commercial and sports stores on the 2nd floor, dining areas on the 3rd floor, and retail outlets on the 4th floor. London's O2 Arena incorporates retail, dining, leisure, and sports spaces in the surrounding area, with additional interactive sports spaces like soccer training areas, indoor trampolines, bowling, and mini-golf, creating a versatile sports and commercial complex that attracts a wide range of visitors.

(3) Strategic Layout of VIP Areas

The strategic layout of VIP areas is crucial for enhancing the audience experience and service level within sports venues. Well-placed VIP suites provide a more comfortable and exclusive viewing experience that appeals to high-end consumers and generates significant revenue. In some venues, VIP area income has even exceeded naming rights revenue.

For instance, the Mercedes-Benz Arena in Shanghai has suites on the third and fourth floors, with capacities of 12, 21, and 30 seats, respectively. VIP suite holders enjoy access to various events and brand-exclusive experiences. At the Barclays Center in New York, luxury suites are positioned within 25 rows of the court, offering premium dining and optimal views. The venue also provides exclusive clubs, music bars, and experiential activities, adding unique value for guests while generating additional revenue. Similarly, the Crypto.com Arena in Los Angeles (formerly the Staples Center) includes 2,500 club seats, most of which are in the lower tier, and 172 suites between the upper and lower tiers, including 56 premium suites, 62 standard suites, and 54 value suites. Two tunnel suites allow close interaction with players, contributing to over \$300 million in long-term lease revenue. The O2 Arena in London recently upgraded its VIP area by replacing 12 suites with The Residence, a new club experience featuring a retractable walkway that allows VIP guests to enjoy elevated views of the arena.

(4) Innovative Utilization of Rooftop Space

Innovative rooftop spaces provide unique vantage points and introduce new functions to sports venues, amplifying urban vitality. At the Mercedes-Benz Arena in Shanghai, the rooftop features clubs, cinemas, bars, and viewing platforms, creating a distinctive experience and enriching the venue's recreational offerings, with panoramic views of the Huangpu River skyline. Shenzhen Bay Sports Center's "Sea Gate" rooftop offers an elevated vantage point over the bay and serves as a venue for award ceremonies, high-end gatherings, client appreciation events, and brand launches, enhancing both the spectator experience and commercial value. Crypto.com Arena in Los Angeles recently added three rooftop suites, allowing members to gather on terraces overlooking the downtown skyline during events (Xie & Kang, 2011). Atlanta's Mercedes-Benz Stadium includes the Maybach Terrace and Delta SKY360 Terrace on its rooftop, offering spectators a view of the cityscape while enjoying events, blending sports entertainment with iconic urban vistas.

3.3 Engaging, Sustainable Scene Content Creation

The vibrancy of sports venues integrated with commercial spaces largely stems from the continuous refreshment of underperforming spaces and the creation of diverse, event-driven scenes. Research indicates that revenue for most sports venues is positively correlated with both time spent on-site and visitor spending (Thibaut et al., 2017). Varied functions and rich scene settings increase visitors' dwell time, thereby enhancing profitability and addressing operational challenges faced by many sports venues in China. Beyond maximizing horizontal and vertical spatial utilization, sports venues should prioritize the strategic arrangement of diverse functional spaces to create engaging, dynamic content, making the venue a lively, attractive destination for the public (Table 4).




(1) Continuous Renewal of Commercial Functions

Updating commercial functions is critical to improving the operational efficiency and market competitiveness of sports venues. Regularly refreshing underperforming spaces serves to create a richer, more diverse functional environment and user experience, aligning with evolving public demands and social trends to mitigate issues of low utilization and poor revenue performance. This, in turn, extends the life cycle of the venue. Key factors guiding functional updates should include: (1) Changing Social Needs – updates should align with evolving societal interests, such as increasing demand for healthy lifestyles, cultural and artistic experiences, and senior-friendly recreational spaces (CHEN, et al., 2021); (2) User Demographics – understanding the age and

occupation of local residents can help tailor facilities, such as more leisure spaces for senior communities or educational and recreational facilities for young families; (3) Technological Advancements – new technologies enable venues to incorporate innovative elements, such as virtual reality zones and interactive exhibits, attracting younger visitors and tech enthusiasts; and (4) Local Cultural Characteristics – updates should also consider regional heritage and culture, with dedicated exhibition and experience zones in culturally rich areas.

For example, in 2022, Xujiahui Sports Park in Shanghai underwent an upgrade of both indoor and outdoor functions, adding versatile sports areas to meet the needs of a wide user base. These include the "Wantihui" multi-purpose hall with badminton, table tennis, tennis, soccer, and basketball courts. The swimming facility added a play pool and sun courtyard, while the outdoor area was enhanced with expanded greenery. Sixteen soccer and basketball courts were added to the "Sports Park" on the south side, while the "Aerobic Park" on the north provides a relaxing space for residents. A flexible, scenic jogging track connects the facilities, creating a unique running experience. At Xujiahui, visitors can now enjoy a wide variety of activities, establishing the venue as a vibrant sports and cultural hub in downtown Shanghai.

Table 4. Engaging, sustainable scene content creation

Project	Real/Rendering image	Model analysis	Strategy description
Xujiahui Sports Park, China		Continuous Update of Commercial Functions	1. Analysis of changing social demands. 2. Assessment of demographic and occupational attributes of local users. 3. Update and add dynamic functional areas. 4. Maximize use of indoor and outdoor spaces for updates and renovations to introduce lively functions.
HuaXi LIVE·Wukesong, China		Continuous Update of Commercial Functions + Multi-Stimulus Event Scene Creation	1. Analysis of changing social demands. 2. Assessment of demographic and occupational attributes of local users. 3. Create various interactive functions and "event scenes" across indoor and outdoor spaces. 4. Host diverse events to attract different user groups.
Xili Sports Complex, Shenzhen, China		Multi-Stimulus Event Scene Creation	1. Optimize use of multi-level spaces, arranging various open activity areas. 2. Connect venues and public spaces with an elevated running track. 3. Host a variety of "event scenes" to engage different user demographics.

(2) Creating Multi-Stimulus Event Scenes

Creating "event scenes" within sports venues is a key strategy for attracting public participation and enhancing venue vitality (Lu et al., 2023). Event scenes encompass more than major sporting events or large concerts; they include a variety of activities and gatherings that provide a multidimensional sense of enjoyment for visitors (Guenzi, 2007). These scenes increase the use frequency and operational capacity of the entire venue and its commercial spaces. Key considerations for event scene creation include:

Diverse and Innovative Activities – regular events such as amateur sports tournaments, community art exhibitions, fitness classes, and public lectures, as well as contemporary events like e-sports tournaments, billiard competitions, and night walking festivals.

Enhanced Interactive Experiences – offering zones for product demonstrations, sports tryouts, VR experiences, and other engaging activities to attract individuals with diverse interests and showcase the broader value of sports and related industries.

For example, HuaXi LIVE·Wukesong in Beijing exemplifies successful event scene creation with a wide array of updated functions. Outdoor areas include the Hi-Park basketball park, HI-ICE ice rink, HuaXi LIVE

Adventure Park, and HI-FUN amusement park, providing spaces for all age groups. Inside, the venue offers the Huiyuan Small Theater, M-Space, executive clubs, bars, VIP suites, and dining options, regularly hosting theater performances, fashion shows, fan meetups, and commercial exhibitions. Iconic events such as HI-Art Festival, Beer Carnival, and HI-Water Festival, along with the integration of Times Art Museum and commercial streets, create an immersive “event scene” encompassing entertainment, sports, lifestyle, education, and art, ensuring stable revenue and elevating regional vitality (Lanxionsports, 2022). Similarly, Shenzhen's Xili Sports Complex integrates a cultural and sports center, theater, multi-sport gymnasium, swimming pool, and a rooftop jogging track connecting buildings. Each structure offers open rooftop spaces, while a green stepped roof forms an outdoor theater or cinema, creating a multi-dimensional event space that merges with the urban environment. No clear boundary remains between the venue and the city, fostering community engagement.

4. DISCUSSION

4.1 Synergistic Integration of "Optimal Horizontal Space Utilization, Vertical Compound Utilization, and Scene Content Creation"

Based on the preceding results, enhancing spatial vitality relies on the synergistic interaction of “optimal horizontal space utilization, vertical compound utilization, and scene content creation.” Designs focused on a single dimension often fall short of fully activating the venue's potential. In contrast, the complementary integration of multi-dimensional spaces and content fosters a cohesive, attractive environment that offers users diverse activity options and experiences, leading to sustained spatial value. This synergy not only improves overall venue efficiency but also fuels industry growth and urban vitality in surrounding areas, providing a continuous source of dynamism (Bradbury, 2021) (Figure 1).

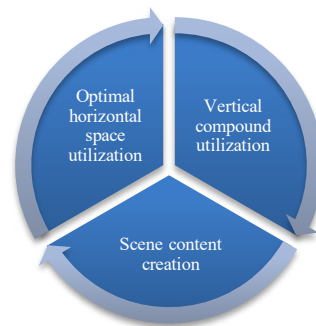


Figure 1. The need for a coordinated role of the three parties

As exemplified by the design of the Mercedes-Benz Arena in Shanghai, the synergy among the three components is essential for enhancing the spatial vitality of sports venues (Figure 2).

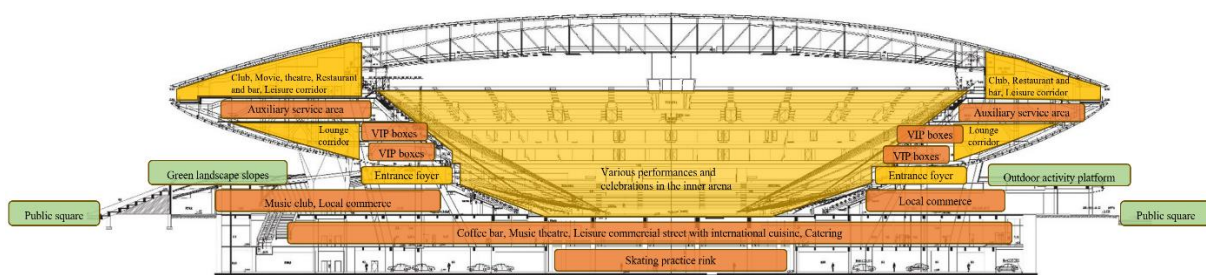


Figure 2. Multidimensional integration of the Mercedes-Benz Arena with commercial spaces, Shanghai.

Firstly, the arena emphasizes the efficient use of vertical space. The B2 level houses an ice rink, utilized for skating training and public activities. The B1 level connects to the metro, featuring an international gourmet food street, retail spaces, and a music theater. The first floor includes a music club, while the third and fourth floors accommodate 82 VIP suites and various restaurants. The sixth floor features clubs, cinemas, restaurants, bars, and a scenic leisure walkway. These spaces collectively create a vibrant urban complex that integrates performing arts, sports, commerce, and culture, providing diverse leisure experiences for a wide range of visitors.

Secondly, despite its limited site area, the arena maximizes horizontal space. The first floor and underground levels are extended to accommodate a variety of functions, while the second-floor outdoor platform is seamlessly connected to the landscaped green slopes, forming plazas, green spaces, and activity platforms. These

outdoor areas host various community activities, enhancing the venue's accessibility and appeal.

Finally, scene content creation is a key driver of the arena's spatial vitality. The 18,000-seat main arena can be reconfigured to host diverse events, such as concerts, music performances, and celebratory gatherings. The venue also features a music club, cinema, ice rink, international gourmet food street, Hans Christian Andersen Children's Park, NBA interactive exhibit, and nearly 20,000 square meters of commercial retail and cultural entertainment spaces. These facilities offer opportunities for artistic exchange, academic research, leisure, tourism, and more. Externally, the venue interacts dynamically with the city by hosting activities like summer camping and riverside running, establishing itself as a trendy and youthful urban hub that attracts and engages the public.

4.2 Comprehensive Optimization Across the Full Lifecycle is Essential for Sustaining Spatial Vitality

Sustaining spatial vitality of sports venues is not solely a consideration during the architectural design phase but requires a coordinated effort throughout the entire lifecycle of the building. Achieving the synergistic effect of "optimal horizontal space utilization, vertical compound utilization, and scene content creation" demands a holistic approach. From the decision-making and strategic planning stages to design, construction, operation, maintenance, and feedback optimization, the sustained spatial vitality of sports venues requires close collaboration and continuous refinement among multiple stakeholders, including governments, developers, design teams, construction firms, operations management companies, and end-users. Only through this comprehensive, lifecycle-wide coordination can truly attractive and dynamic urban sports spaces be created (Figure 3).

The Mercedes-Benz Arena in Shanghai exemplifies this holistic approach, with its vibrant spatial vitality stemming not only from its design phase but also from meticulous attention across the entire building lifecycle. During the planning and construction stages, the arena was conceptualized as the performance venue for Expo 2010 Shanghai. With a forward-looking vision, both the government and developers aimed to integrate functions such as performing arts, sports, and entertainment to ensure that the venue would continue to bring vitality to the city even after the Expo concluded. In the operational phase, the arena collaborated with internationally renowned management companies to optimize its operations across five key areas: naming rights, VIP suites, food and beverage, merchandising, and integrated development. Management models, functional configurations, and event programming have been continuously updated to meet the diverse needs of its user groups. Notably, revenue from naming rights, sponsorships, and VIP suite rentals now accounts for 80% of the venue's total income, providing a robust foundation for its sustainable vitality and long-term operation.

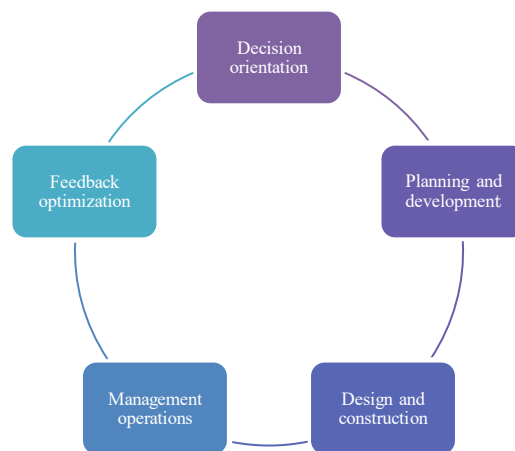


Figure 3. Framework for enhancing spatial vitality through multi-stage collaboration across the venue lifecycle

5. CONCLUSIONS

As the functional value of sports architecture in China shifts and national fitness initiatives advance, many venues are transitioning from a "sports+" model to a "+sports" model, emphasizing widespread public fitness, entertainment, and recreation. The integration of sports venues with commercial spaces has become a vital means of stimulating urban vitality. "Optimal horizontal space utilization, vertical compound utilization, and scene content creation" constitute the essential design strategies for enhancing spatial vitality in sports-commercial integration. This tripartite synergy is critical, and as sports venue functions evolve, new innovative strategies will likely emerge, further advancing the organic integration and development of sports and commercial spaces. On the other hand, the spatial vitality of sports venues is a dynamic and evolving process. It necessitates continuous optimization and adjustments at every stage of the venue's lifecycle—ranging from government decision-making, planning, and development to design, construction, operational management, and feedback optimization—with collaboration among all stakeholders.

ACKNOWLEDGMENTS

This research was funded by Jiaxing University Research Initiation Grant (Research on the Integration Mode of Sports Venues and Commercial Spaces) (grant number CD70523037).

REFERENCES

- Mo, L., and Chen, B. (2020). Principle and practice for everyday life dissolving “the grand” in the Qingpu Sports and Cultural Activity Center in Shanghai, *Time + Architecture*, 06, 132–139.
- Zhang L., Xie Q., Liu N., and Mancini, M. (2021). Sustainability in Beijing 2022: A Chinese Framework for Cross-Scale Design Intervention, *Architectural Journal*, (Z1), 7-13.
- Turner, R.S., and Rosentraub, M.S. (2016). Tourism, sports and the centrality of cities, *Journal of Urban Affairs*, 24(5), 487–492.
- Ma, J., and Zhuang, W. (2004). Analysis of the comprehensive utilization of space under fixed seats in large sports buildings in China. *Architectural Journal*, 05, 70–72.
- Luo, P., and Mei, J. (2006). Research on the dynamic adaptive design framework of large sports venues. *Architectural Journal*, 05, 61–63.
- National Development and Reform Commission (NDRC); Ministry of Agriculture and Rural Development (MAFRD); Ministry of Commerce (MOFCOM); Ministry of Culture and Tourism (MCT); State Administration for Market Supervision and Administration (SAMSA) (2024). *Circular of the National Development and Reform Commission and Other Departments Issuing Measures on Creating New Scenes of Consumption and Cultivating New Growth Points of Consumption*. Retrieved from Ministry of Culture and Tourism website: https://www.mct.gov.cn/preview/whhlyqyzcxfw/yshjxf/202407/t20240708_953949.html
- Xinhuanet Client. (2023). *Cultivate new business forms, create new scenarios, explore new modes Stimulate new vitality of sports consumption*. Retrieved from Xinhua net website: <https://app.xinhuanet.com/news/article.html?articleId=40ac606f2ee2bce3bd3ded7687ac9554>
- Bryman, A. (2016). *Social research methods* (5th ed.). Oxford University Press.
- Wei, Z., Liu, D., and Luo, P. (2020). A Study on strategy of space utilization of large-scale sport/ performance centers based on locational and environmental analysis, *Journal of World Architecture*, 09, 110–113.
- Yuan, X., Pan, H., and Fu, B. (2023). Models and analysis of stadium and commercial space integration: a case study from China, *Building Research and Information*, 51 (8), 914-936.
- Shi J., and Li, B. (2020). Sports Building Design Strategies Based on the Concept of Post-Competition Operation--Taking Jingjiang Sports Center as an Example, *Architecture & Culture*, (04), 266-269.
- Fan, Z., Liu, Z., and Yang, S. (2022). Interaction & Symbiosis: The Design Strategy of “Interaction” between Architecture and Landscape of the Exterior Space of Sports Park Based on the Spirit of Place, *New Architecture*, (02) , 76-81.
- Zhuang, W., Li, X., Ding, J., et al. (2020). Reflection on Current Development of Sport Architecture: From Increment to Inventory, *Contemporary Architecture*, (06), 6-15.
- Toutiao. (2019). *Venue Research: How Mercedes-Benz Arena has become a model for venue operations*. Retrieved from Toutiao website: <https://www.toutiao.com/article/6752431782701826571/?wid=1668075123079>
- Xie, S., and Kang, X. (2011). Spring cocoon—Interpretation of the design of Shenzhen Bay Sports Center. *Architectural Journal*, (09), 79–80.
- Thibaut, E., Eakins, J., Vos, S., et al. (2017). Time and money expenditure in sports participation: The role of income in consuming the most practiced sports activities in Flanders, *Sport Management Review*, 20, 455–467.
- Chen, Y., Chen, L., Li, J., Li, J., He, Y., and Shi, X. (2021). Main Performance, Strategies of American Sports Stadiums Promoting Urban Renewal and Their Enlightenment on China, *Journal of Shanghai University of Sport*, 45(2): 78-89.
- Lu, S., Du, X., and Guo, Q. (2023). The Development of Major Sports Events Venue Construction and Post-Match Utilization Mode in China, *Contemporary Architecture*, (09), 32-35.
- Desbordes, M., and Chadwick, S. (2007). *Marketing and Football*. Routledge.
- Lanxionsports. (2022). *Exclusive Dialogue with Huaxi Zhao Yan, Behind the Scenes of Wukesong, a Model Student of Venue Operations*. Retrieved from lanxionsports website: <http://www.lanxionsports.com/posts/view/id/22551.html>
- Bradbury, J.C. (2022). The impact of sports stadiums on localized commercial activity: Evidence from a Business Improvement District. *Journal of Regional Science*, 1: 194-217.

REAL-TIME ON-SITE INDOOR MODELING AND QUALITY VISUALIZATION USING MIXED REALITY DEVICE WITH TOF SENSOR

Norihiko Goto¹, Hiroaki Date², and Satoshi Kanai³

1) Master Course Student, Graduate School of Information Science and Technology, Hokkaido University, Sapporo, Japan. Email: n_goto@sdm.ssi.ist.hokudai.ac.jp

2) Dr. Eng., Assoc. Prof., Faculty of Information Science and Technology, Hokkaido University, Sapporo, Japan. Email: hdate@ssi.ist.hokudai.ac.jp

3) Dr. Eng., Prof., Faculty of Information Science and Technology, Hokkaido University, Sapporo, Japan. Email: kanai@ssi.ist.hokudai.ac.jp

Abstract: In this study, we propose a real-time on-site indoor modeling method and a modeling quality visualization method that use a mixed reality (MR) device with a time-of-flight (ToF) sensor to realize efficient and reliable indoor model generation. The modeling method is based on rule-based fast voxel labeling and consists of two steps: local and global modeling. In the local modeling, space occupancy labels and attribute labels (ceiling, wall, wall opening, wall candidate, floor, object, and space) are assigned to each cell of the local voxel from point clouds in each frame of laser scanning by the MR device. In global modeling, the results of the local modeling are integrated with the global voxel, considering the label assignment quality based on entropy and probability. The quality of the resulting model is presented to the user through MR visualization to verify the modeling results. For real-time visualization of the quality of the modeling results, a simple textured polygon model is created and used for MR visualization. The polygon model is efficiently generated from the local and global voxels, and the texture represents the quality of the modeling results. Our method was applied to a real indoor environment and its performance was evaluated. The modeling frame rate was 2 fps, and the labeling precision was 98%. A polygon model for MR visualization was created and updated within 20 msec, and real-time MR visualization was achieved. These results show that real-time indoor modeling and on-site quality checks using an MR device with a ToF sensor can be realized.

Keywords: Indoor modeling, Laser scanning, Mixed reality, Point cloud, Voxel, Quality visualization.

1. INTRODUCTION

Three-dimensional (3D) scanning technologies for large-scale environments and structures have been rapidly developed over the past decade and are used in a wide range of fields, including plant engineering, surveying, architecture, civil engineering, forests, and construction. Laser scanning and image-based 3D reconstruction techniques allow us to acquire point clouds or meshes of real environments and structures, and several types of platforms are available, such as drones/vehicles for large outdoor areas and terrestrial/backpack/handheld platforms for indoor environments and large-scale structures. Recently, time-of-flight (ToF) sensors have been installed in handheld mobile devices and head-mounted mixed reality (MR) devices. MR devices are especially suited for interactive and real-time applications because of their ability to acquire and show 3D geometric information in both cyber and physical spaces.

With the advancement of 3D scanning technology, many methods for modeling indoor environments have been developed and used in practical applications. The 3D model of the indoor environment based on 3D scanning technology has many applications in fields such as maintenance, drawing generation, navigation, simulation, planning, and MR. Many studies on 3D indoor modeling and recognition from point clouds acquired by laser scanning have been conducted (Akiyama et al., 2024; Cui et al., 2023; Hübner et al., 2021; Macher et al., 2017; Martens & Blankenbach, 2023; Monszpart et al., 2015; Pan et al., 2022; Qi et al., 2017; Quintana et al., 2018; Takahashi et al., 2019). The targets of modeling and recognition are indoor structures such as rooms, ceilings, walls, floors (Cui et al., 2023; Macher et al., 2017; Martens & Blankenbach, 2023), doors (Quintana et al., 2018), windows (Takahashi et al., 2020), and small facilities (Akiyama et al., 2024; Giovanni et al., 2018; Pan et al., 2022). In addition to the rule- or algorithm-based methods, machine-learning-based methods have become widely used (Qi et al., 2017). To realize accurate and efficient reconstruction, Manhattan assumptions and regularities are available (Monszpart et al., 2015; Takahashi et al., 2019). Recently, speed-specific algorithms that use voxels have been proposed (Hübner et al., 2021; Martens & Blankenbach, 2023).

The modeling of indoor environments based on 3D scanning is often performed offline after the scanning operation has been completed on-site and point clouds of the entire environment have been acquired. However, in this approach, it is difficult to guarantee the quality of the point clouds and the resulting 3D models. Point clouds from scanning often have problems related to density, scanning errors, and lack of points owing to occlusions; therefore, in the worst case, additional scanning may be required back on-site. To solve these problems, we propose a real-time on-site indoor reconstruction and quality check method using an MR device with a ToF sensor. Figure 1 shows an outline of the proposed method. This method consists of real-time voxel-based indoor modeling and MR visualization of the quality of the modeling results. The scanning of the environment is performed by the MR

device at approximately 5 fps while walking and seeing the environments (Figure 1(i)), and voxel models of the indoor reconstruction are generated using the rule-based method in real-time (Figure 1(ii)). The modeling is realized using the point cloud for each frame, and the results are successively merged with previous results. The quality of the modeling is shown to the user using on-site MR visualization (Figure 1(iv)). To realize efficient MR visualization, a method for generating a simple textured polygon model that can represent the quality of the indoor reconstruction is developed (Figure 1(iii)).

The details of the proposed method are described in Section 2, and the experimental results and discussions are presented in Section 3.

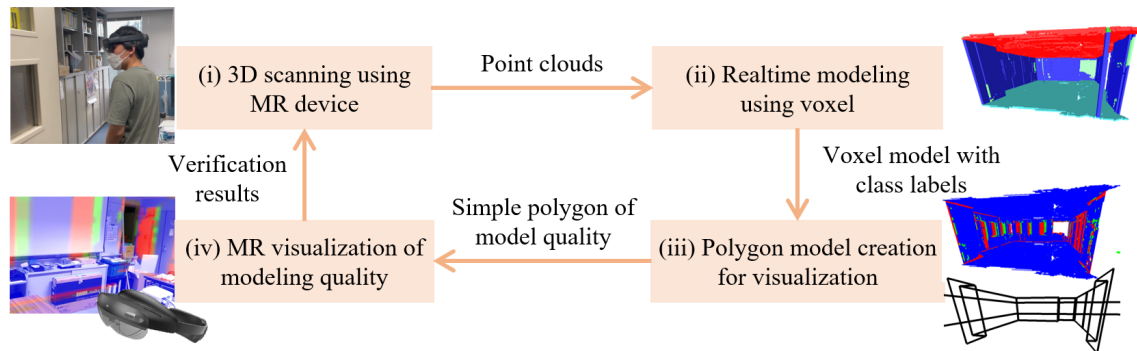


Figure 1. The proposed real-time on-site indoor modeling and quality visualization method using a mixed reality (MR) device

2. METHOD

2.1 Mixed Reality Device and Method Overview

The input for this method consists of a series of point clouds acquired using a head-mounted MR device. A Microsoft HoloLens2 (Microsoft, 2024) is used in this study. The device mounts a ToF LiDAR sensor, and real-time point cloud acquisition at approximately 5 fps is realized. The point cloud per frame contains approximately 60,000 points. Each point has a unit normal vector, and the measurement range is approximately 3.5 m. The point cloud in each frame is automatically registered using the self-localization function of the device in a global coordinate system.

In this study, a real-time modeling method for point clouds acquired from MR devices and an MR visualization method of the quality of the modeling results are developed. The flow of the proposed method is illustrated in Figure 2. The output of the modeling method is an indoor reconstruction voxel model, in which one of the attribute labels is assigned to each cell. Similar to the existing method (Hübner et al., 2021), the attribute labels include *ceiling*, *wall*, *wall_opening*, *wall_candidate*, *floor*, *object*, and *space*. The *wall_opening* label corresponds to the entrances of a room or windows. The *wall_candidate* label represents the candidates for a wall, such as the regions occluded by a shelf in front of a wall. The proposed indoor modeling method consists of local modeling, which generates a local voxel with labels assigned from a point cloud in a frame, and global modeling, which integrates the labels from the local modeling and updates some labels that cannot be estimated in only one frame (Figure 2, A-1,2). As different labels may be assigned to the same cell of a global voxel in the local modeling, a measure of the reliability of the label is calculated during the integration process to evaluate the quality of the resulting model. In contrast to common modeling approaches (Hübner et al., 2021; Martens & Blankenbach, 2023), this method creates a voxel model immediately while acquiring the point clouds to provide online feedback on the modeling results and their quality to the user with the MR device.

The MR device allows us to superimpose 3D digital information onto the real world using MR visualization. Because the user is always on-site during scanning, we try to show the modeling results and their quality to the user on-site and in real-time utilizing MR. Unfortunately, voxel models and point clouds are not appropriate for MR visualization because of the large amount of data. Therefore, a 3D model suitable for MR visualization of modeling quality is proposed. In this study, a simple polygon with texture is used and, an efficient generation method is proposed for our modeling framework (Figure 2, A-3).

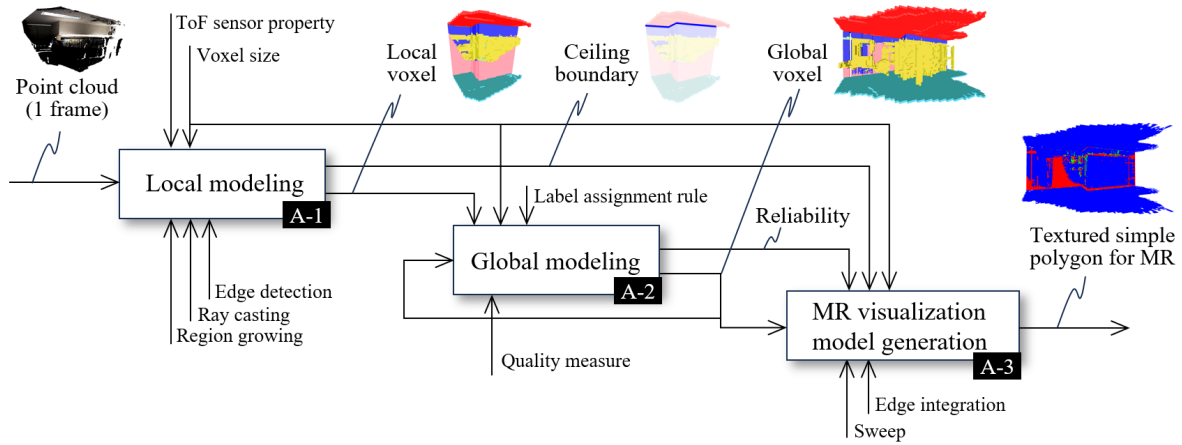


Figure 2. A flow of voxel-based modeling and MR model generation

2.2 Voxel-based Indoor Modeling

(1) Definition and Assumptions

Figure 3(a) shows a scene and the definitions of certain terms. Our modeling method is based on the following observations and properties of each entity: The ceiling is a large plane with a downward normal above the device (ToF sensor) position. The floor is a large plane with an upward normal below the device position. The walls are vertical planes below the ceiling boundary. The region on a wall plane through which the laser passes is recognized as a wall opening. The objects are represented by the scanned points between the ceiling and the floor, and free space is recognized as the region through which the laser passes during laser scanning. The occluded regions can be detected by ray casting from the device position to points in the point cloud, and the vertical occluded region below the ceiling boundary is recognized as an occluded wall. The occluded floor can be estimated by interpolation/extrapolation at the floor height within the ceiling range. This method assumes that the floor and ceiling heights of the environment to be reconstructed are constant. In the coordinate system of the point clouds and modeling, the vertical direction corresponds to the z-axis.

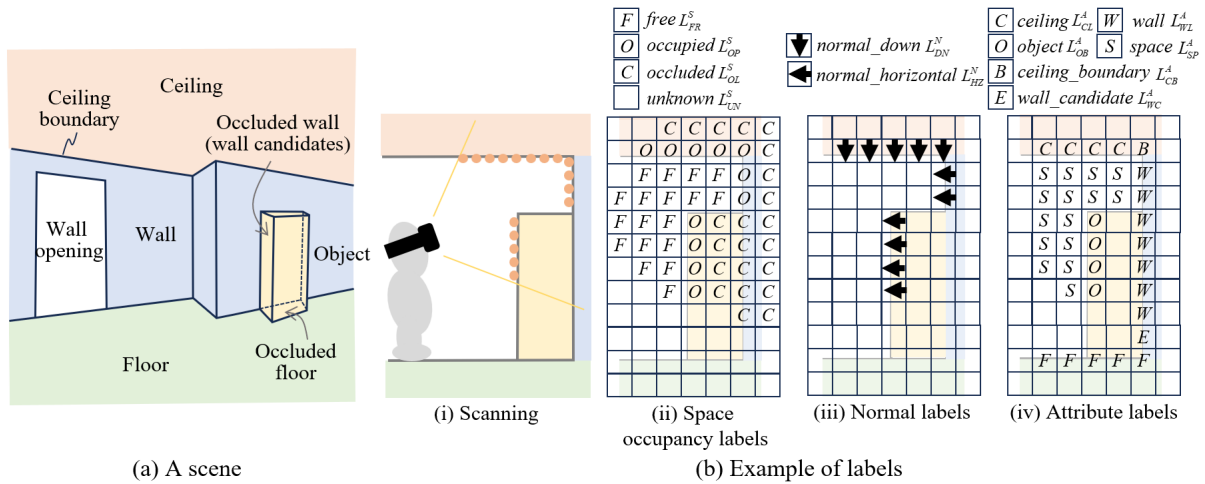


Figure 3. A scene and an example of labels

(2) Local Modeling

Local modeling consists of three steps: initial voxel generation, floor/ceiling detection, and wall/object detection. First, the initial voxel with space occupancy and normal labels is generated (Figure 3 (b) (ii)-(iii)). Next, the ceiling, ceiling boundaries, and the floor are detected based on the normal labels. Finally, walls, wall openings, wall candidates, objects, and spaces are detected from the ceiling information (Figure 3 (b) (iv)). The basic labeling strategy for the ceiling and walls is based on an existing method (Hübner et al., 2021), but the method is simplified to realize real-time labeling for a one-frame point cloud. In addition, by using ray casting, reliable estimation of occluded regions and labeling are achieved using our method.

During the initial voxel generation, a voxel with space occupancy and normal labels is generated. First, a voxel grid is defined by uniformly dividing the space of the scanner's measurement range (3.5 m for HoloLens2)

using a specified voxel size. The voxel size is set to 5 cm in our implementation, according to the accuracy of the point clouds from the experiments. The center of the voxel space is the device position. Then, using the ray casting from the device position in the voxel space, the space occupancy label, *occupied* L_{OP}^S , *unknown* L_{UN}^S , *free* L_{FR}^S , or *occluded* L_{OL}^S , is assigned to each cell as shown in Figure 3(b) (ii). In this process, L_{UN}^S is first assigned to all cells, and L_{OP}^S is assigned to the cells containing the points. L_{FR}^S and L_{OL}^S are then assigned using ray casting from the device position to the cells within the scanner's field of view and measurement range. Finally, in each cell c with L_{OP}^S , the unit average normal vector $\mathbf{n}_c = (x_{nc}, y_{nc}, z_{nc})$ of the points in it is calculated. For the fast computation of region detection, normal information is classified into three types and labeled in the voxel. The label *normal_up* L_{UP}^N , *normal_down* L_{DN}^N , or *normal_horizontal* L_{HZ}^N is assigned as the normal label l_c of cell c according to Equation (1), where τ_c is a given threshold (0.75 is used in the experiments).

$$l_c = \begin{cases} L_{UP}^N & (z_{nc} > \tau_n) \\ L_{DN}^N & (z_{nc} < -\tau_n) \\ L_{HZ}^N & (otherwise) \end{cases} \quad (1)$$

In floor/ceiling detection, the floor and ceiling regions are detected quickly from the initial voxel. In this process, the attribute labels of *ceiling* L_{CL}^A , *ceiling_boundary* L_{CB}^A , and *floor* L_{FL}^A are assigned, and the height of the floor h_f is calculated. The ceiling region is defined as a set of more than a certain number of connected L_{DN}^N cells above the device position. The connected L_{DN}^N cells are detected by using region growing (Poux et al., 2022). When a ceiling region is detected, the boundaries of the ceiling region are recognized as its boundary cells that have L_{HZ}^N cells below them. An example of this labeling is shown in Figure 3(b) (iv). On the other hand, the floor region is detected as a set of more than a certain number of connected L_{UP}^N cells below the device position using region growing. If a floor region is detected from the point cloud once, its average height value h_f is computed and no further floor detection processing is performed, and the label of *floor* L_{FL}^A is assigned to the cells at height h_f under the ceiling cells.

In wall/object detection, a downward traversal is performed from the L_{CL}^A (*ceiling*) and L_{CB}^A (*ceiling_boundary*) cells to the floor level h_f . According to the space occupancy labels of the visited cells in the traversal, *wall* L_{WL}^A , *wall_opening* L_{WO}^A , *wall_candidate* L_{WC}^A , *object* L_{OB}^A , or *space* L_{SP}^A are assigned according to the rules shown in Table 1. For example, the attribute label of a cell with an occupied label under the ceiling is *object* L_{OB}^A , and the label of a cell with an occluded label under the ceiling boundary is *wall* L_{WL}^A . For the cells at the floor level under the *ceiling* and *ceiling_boundary* cells, the *floor* label L_{FL}^A is assigned. Examples of the labeling results for a scan are shown in Figure 3(b) (iv). This label assignment process is performed when a floor has been detected at least once and a ceiling is detected in the current point cloud.

Table 1. Attribute label assignment rule

An attribute label of the start cell of traversal	A space occupancy label of the visited cell			
	<i>occupied</i> L_{OP}^S	<i>free</i> L_{FR}^S	<i>occluded</i> L_{OL}^S	<i>unknown</i> L_{UN}^S
<i>ceiling</i> L_{CL}^A	<i>object</i> L_{OB}^A	<i>space</i> L_{SP}^A	—	—
<i>ceiling_boundary</i> L_{CB}^A	<i>wall</i> L_{WL}^A	<i>wall_opening</i> L_{WO}^A	<i>wall</i> L_{WL}^A	<i>wall_candidate</i> L_{WC}^A

(3) Global Modeling

The space occupancy and attribute labels of the local voxels are transferred to the corresponding cells in the global voxel. First, an initial global voxel is generated by dividing a sufficiently large space with a specified voxel size (5 cm in our implementation) centered at the device position of the first frame of the point cloud acquisition. After each local voxel generation, the label of the global voxel is successively updated. In the experiments, different local voxel labels were often assigned to a cell of the global voxel because of the scanning noise and outliers in the point cloud, different scan directions, and voxel discretization errors. Therefore, in this study, a labeling method based on the reliability of labels from the local voxels is introduced.

For each cell in the global voxel, the reliability is calculated if the minimum number of the assigned labels from the local voxel is satisfied. In our study, the state probability in Equation (2) and entropy in Equation (3) are evaluated.

$$P_c^l = \frac{B_c^l}{B_c} \quad (2)$$

$$E_c = -\sum_{l \in L} P_c^l \log P_c^l \quad (3)$$

Where, B_c^l is the number of assignments of label l from the local voxel for cell c in the global voxel, B_c is the total number of label assignments for cell c , and L is the attribute-label set. A smaller entropy means that the label distribution is non-uniform, and the number of assignments of a label becomes higher compared with that of other labels, that is, the reliability of the label with the highest state probability is high. Therefore, in this method, if the entropy becomes lower than the given threshold τ_E , the attribute label l_c of the global cell is assigned using Equation (4).

$$l_c = \arg \max_{l \in L} P_c^l \quad (4)$$

2.3 Mixed Reality Visualization of Modeling Quality

A method to show the modeled regions and the quality of the modeling results and to support the additional scanning of the user in real-time using MR is proposed. Because voxel models and point clouds with large data volumes are not suitable for MR visualization from the viewpoint of data amount (Ohno et al., 2022), a simplified polygon model with texture is used. In this method, a polygon model for MR visualization is generated from the local and global voxels. The polygon model is defined by large rectangles for the ceiling and floor and rectangles for walls defined by each ceiling boundary edge, as shown in Figure 4 (right). The reconstruction results and their reliability information are represented by texture colors. The texture color is assigned according to the entropy defined by Equation (3), and the non-reconstructed areas are transparent.

Figure 4 presents an overview of the polygon model generation method for MR visualization. The ceiling and floor polygons can be defined using large horizontal rectangles when their heights are determined. Wall polygons are created by sweeping the straight ceiling boundary edges to the floor height. In this method, the ceiling boundary edges are detected in the local modeling step. The edges detected first are directly copied to and stored in the global voxel and are updated by the edges from successive local modeling.

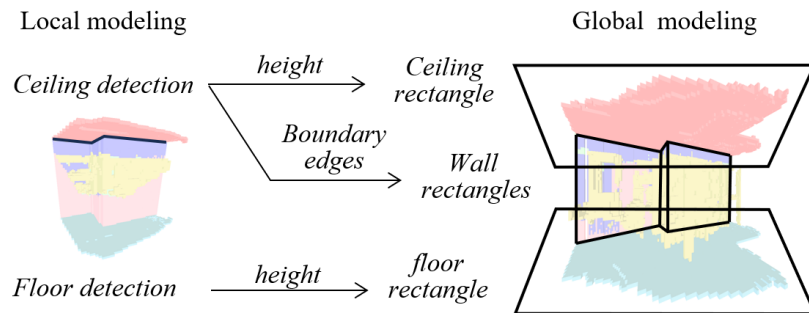


Figure 4. Polygon model generation for MR visualization

The straight-line segments of the ceiling boundary edges, which represent the boundary between the ceiling and walls, are detected by line fitting to the points in the ceiling boundary cells. For a fast computation, the barycenter points of the scanned points in each cell are used. Straight lines are detected using RANSAC (Fishler & Bolles, 1981; Schnabel et al., 2007), and the barycenter points in the cells of the endpoints of the detected lines are used as endpoints of the straight-line segments. The initial line segments from the local modeling are used as the initial line segments in the global voxel. If the new edges from the local modeling overlap with and have the same direction for the current line segments in the global voxel, the positions of the endpoints in the global voxel are modified. Otherwise, the edges are added as new edges to the global voxel. This process is performed during global modeling.

For each polygon, a texture image with a pixel size equal to the cell size of the voxel is generated. From each pixel p in the texture image, the cell c_p of the corresponding global voxel is found, and the color of p is determined based on the value of the entropy of c_p . As it is difficult to distinguish slight changes of color in MR visualization, three colors are assigned in our implementation: blue for high quality, green for medium quality, and red for low quality or for an insufficient number of labeling from the local voxel. The quality becomes higher when the scanning times are increased; therefore, red regions are identified as regions that require additional scans. Transparency is assigned to the pixels that correspond to the *wall_opening* cells or cells without the attribute labels.

3. RESULTS AND DISCUSSIONS

The proposed method was applied to a room with an area of approximately 70 m² as shown in Figure 5(a). Scanning was performed for approximately 2 min while walking around the room. During the scanning, 607 frames and approximately 30 million points were acquired, as shown in Figure 5(b). In the experiment, to verify the feasibility of real-time modeling and polygon generation for MR visualization, voxel modeling and polygon model generation were performed using a desktop PC with a Ryzen 5600X CPU after acquiring all the point clouds.

Figure 5(c) shows the voxel model of the resulting indoor reconstruction. The results indicated that walls, floors, ceilings, and objects were recognized and represented by the voxel model. Figure 6 shows the results of local and global modeling for each frame. The top figures in each figure show the results of the local modeling, and the middle figures show the results of global modeling. In local modeling, indoor environments are recognized from the point cloud of one frame, and the global modeling generates the integrated voxel model from the local modeling. The processing times for local and global modeling were approximately 0.5 s per frame. The acquisition rate of the point cloud was 5 fps, and real-time processing reflecting the acquired point cloud was possible for every two or three frames. In our framework, local modeling can also be performed using parallel processing to improve the modeling frame rate, which will be included in future work.

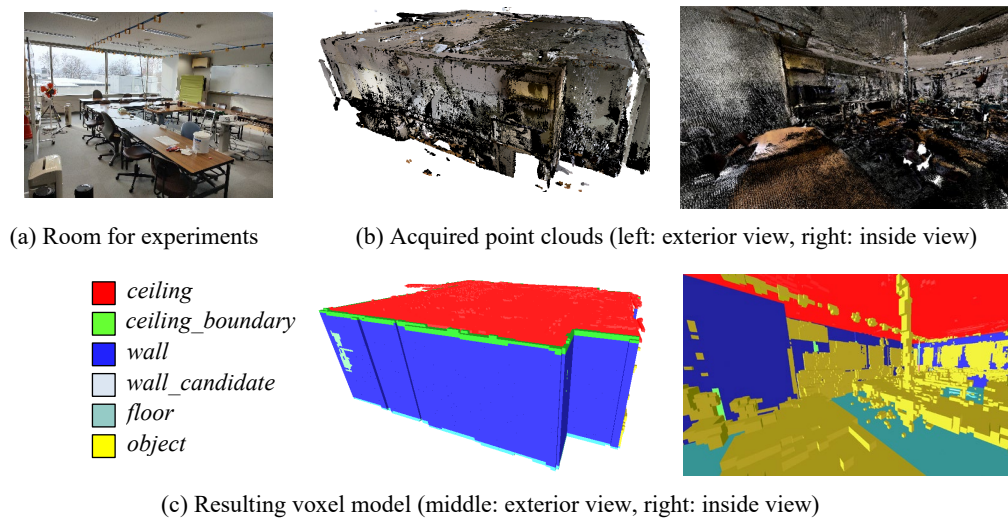


Figure 5. Point clouds and voxel modeling results

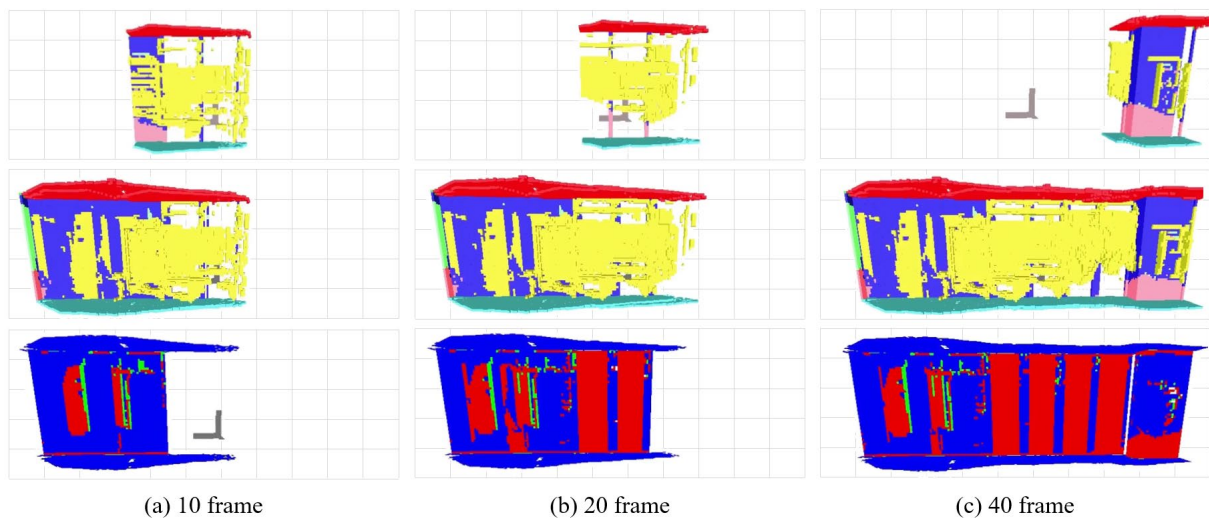


Figure 6. Local voxels, global voxels, and MR models for different frames (top: local modeling results, middle: global modeling results, bottom: textured polygon models for MR visualization of modeling quality)

To confirm the accuracy of the generated voxel model, a ground truth model with the walls, floor, and ceiling of the room was generated by accurate laser scanning using a terrestrial laser scanner and a manual modeling operation using 3D modeling software (Blender). The ground truth model and the results reconstructed by our method were registered, and the precision was evaluated. Adjacent cells were also considered correct in the

evaluation to mitigate the effects of discretization and registration errors. The precision of labeling the global voxel was approximately 98%, and it was confirmed that accurate 3D modeling could be performed using our method.

The bottom figures in Figure 6 show the textured polygon models used to visualize the reconstruction quality for each frame. Only the walls, floors, and ceilings are shown in each figure. The results show that the textured model adequately represents the quality of the global modeling of the walls, floors, and ceilings. The quality of the wall is low in some regions where the labeling is unstable or where labels are not assigned because of tall bookshelves that make ceiling boundary estimation difficult. Figure 7 shows a wireframe representation of the polygon model. A simple polygon model is generated. The textured polygon model was generated at 20 ms per frame, which was sufficient for real-time processing. Figure 8 shows the MR visualization results of the textured polygon model. The MR visualization could be performed at a high frame rate, and the simple textured polygon model was effective in showing the modeling quality on-site using MR.

The proposed method assumes that the indoor environments have a horizontal ceiling. In addition, sufficient measurement of the ceiling boundary leads to an accurate reconstruction model. Real environments may include sloped ceilings, cases where the ceiling boundary is occluded by tall objects, and objects or structures that are difficult to scan, such as glass doors. The development of a method for applying complex environments by using regularity recognition and beautification (Takahashi et al., 2019), and image-based recognition is included in the future works.

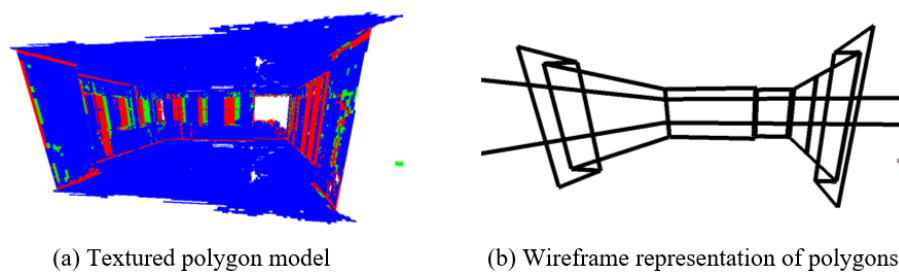


Figure 7. Textured polygon model and wireframe representation of polygons

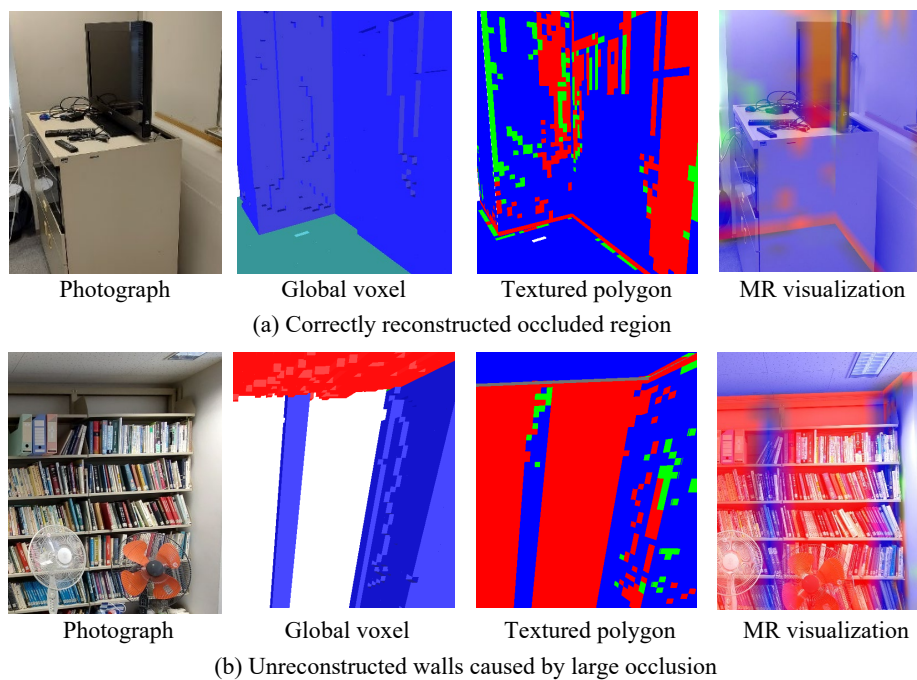


Figure 8. MR visualization of modeling quality

4. CONCLUSIONS

In this study, a real-time on-site indoor modeling method based on fast voxel labeling and a quality visualization method based on simple textured polygons using an MR device with a ToF sensor are proposed. The modeling method consists of a local modeling method that rapidly assigns labels such as walls, floors, and ceilings from the point cloud in each frame and a global modeling method that integrates the results of the local modeling considering the quality of labeling based on the entropy and state probability. For quality visualization, a method

for generating simple textured polygon models is proposed.

The experimental results showed that indoor reconstruction could be performed at 2 fps with a labeling precision of 98%. These results demonstrate the capability of real-time voxel modeling using point clouds acquired by an MR device. In addition, polygon models for MR visualization can be generated and updated within 20 ms, and MR visualization using polygons can be performed at a high frame rate. The MR display of the reconstruction quality was found to intuitively convey areas that were well reconstructed and those that were difficult to measure and not well reconstructed.

Future work will include the implementation of a server-client system to transfer the acquired point clouds and the 3D polygon model wirelessly between the MR device and a computer for modeling and the application to large-scale and complex indoor environments.

ACKNOWLEDGMENTS

This work was supported by JSPS KAKENHI Grant Number JP24K01044.

REFERENCES

- Akiyama, R., Date, H., Kanai, S., and Yasutake, K. (2024). Ceiling Equipment Extraction from TLS Point Clouds for Reflected Ceiling Plan Creation, *International Journal of Automation Technology*, 18 (5), 603-612.
- Cui, Y., Yang, B., Liu, P., and Kong, L. (2023). A Review of Indoor Automation Modeling Based on Light Detection and Ranging Point Clouds, *Sensors and Materials*, 35 (1), 247-268.
- Fischler, M. A. and Bolles, R. C. (1981). Random Sample Consensus: A paradigm for model fitting with applications to image analysis and automated cartography, *Communications of the ACM*, 24 (6), 381-395.
- Giovanni, P., Ruggero, P., Fabio, G., Roberto, S., and Enrico, G. (2018). Recovering 3D existing-conditions of indoor structures from spherical images, *Computers & Graphics*, 77, 16-29.
- Hübner, P., Weinmann, M., Wursthorn, S., and Hinz, S. (2021). Automatic voxel-based 3D indoor reconstruction and room partitioning from triangle meshes, *ISPRS Journal of Photogrammetry and Remote Sensing*, 181, 254-278.
- Macher, H., Landes, T., and Grussenmeyer, P. (2017). From Point Clouds to Building Information Models: 3D Semi-Automatic Reconstruction of Indoors of Existing Buildings, *Applied Sciences*, 7, 1030.
- Martens, J. and Blankenbach, J. (2023). VOX2BIM+ - A Fast and Robust Approach for Automated Indoor Point Cloud Segmentation and Building Model Generation, *Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 91, 273-294.
- Microsoft HoloLens2. (2024). website: <https://www.microsoft.com/ja-jp/hololens>
- Monszpart, A., Mellado, N., Brostow, G. J., and Mitra, N. J. (2015). RAPter: rebuilding man-made scenes with regular arrangements of planes, *ACM Transactions on Graphics*, 34 (4), Article No 103.
- Ohno, K., Date, H., and Kanai, S. (2022). MIXED REALITY VISUALIZATION OF POINT CLOUDS FOR SUPPORTING TERRESTRIAL LASER SCANNING, *The International Archives of the Photogrammetry Remote Sensing and Spatial Information Sciences*, XLIII-B2-2022, 251-258.
- Pan, Y., Braun, A., Brilakis, I., and Borrmann, A. (2022). Enriching geometric digital twins of buildings with small objects by fusing laser scanning and AI-based image recognition, *Automation in Construction*, 140, 104375.
- Poux, F., Mattes, C., Selman, Z., and Kobbelt, L. (2022). Automatic region-growing system for the segmentation of large point clouds, *Automation in Construction*, 138, 104250.
- Qi, C. R., Yi, L., Su, H., and Guibas, L. J. (2017). PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space, *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 5105-5114.
- Quintana, B., Prieto, S.A., Adán, A., and Bosché, F. (2018). Door detection in 3D coloured point clouds of indoor environments, *Automation in Construction*, 83, 146-166.
- Schnabel, R., Wahl, R., and Klein, R. (2007). Efficient RANSAC for Point-Cloud Shape Detection, *Computer Graphics Forum*, 26 (2), 214-226.
- Takahashi, H., Date, H., and Kanai, S. (2019). Automatic Indoor Environment Modeling from Laser-scanned Point Clouds Using Graph-Based Regular Arrangement Recognition, *Proceedings of the 4th International Conference on Civil and Building Engineering Informatics*, 368-375.
- Takahashi, H., Date, H., Kanai, S., and Yasutake, K. (2020). DETECTION OF INDOOR ATTACHED EQUIPMENT FROM TLS POINT CLOUDS USING PLANAR REGION BOUNDARY, *The International Archives of the Photogrammetry Remote Sensing and Spatial Information Sciences*, XLIII-B2-2020, 495-500.