

Automated Progress Monitoring and Reporting for Construction Projects

Dena Shamsollahi

A Thesis

In the Department

of

Building, Civil, and Environmental Engineering

Presented in Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy (Building Engineering) at Concordia University

Montreal, Quebec, Canada

September 2024

© Dena Shamsollahi, 2024

CONCORDIA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: **Dena Shamsollahi**

Entitled: **Automated Progress Monitoring and Reporting for Construction Projects**

And submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY (Building Engineering)

complies with the regulations of the University and meets the accepted standards with

respect to originality and quality.

Signed by the final examining committee:

_____ Chair

Dr. Lyes Kadem

_____ External Examiner

Dr. Jeff Rankin

_____ External to Program

Dr. Amin Hammad

_____ Examiner

Dr. Mazdak Nik-Bakht

_____ Examiner

Dr. Sang Hyeok Han

_____ Thesis Supervisor

Dr. Osama Moselhi

Dr. Khashayar Khorasani

Approved by

Dr. Chunjiang An

Graduate Program Director

September 17th 2024

Dr. Mourad Debbabi, Dean

Faculty of Engineering and Computer Science

ABSTRACT

Automated Progress Monitoring and Reporting for Construction Projects

Dena Shamsollahi, Ph.D.
Concordia University, 2024

In complex and dynamic construction sites, efficient progress monitoring and reporting play an important role in minimizing schedule delays and cost overruns. Such reporting requires detailed and accurate records from job sites to help project managers in comparing project's current state to its as-planned state. Manual traditional progress reporting is time-consuming, costly, labour-intensive, and error-prone. In recent years, advancements in technologies and methods have been introduced in an effort to overcome the challenges of manual methods and to automate the processes of progress monitoring and reporting. These introduced levels of automation still lack capabilities to provide complete and accurate information about the project's current status and available resources on job sites. To address these challenges, this thesis introduces a novel framework for automated progress reporting in construction. This framework provides detailed information for each tracked building element, enabling the identification of its current status and the generation of timely progress reports. The developments integrated into the framework focus on challenges associated with congested mechanical components in indoor environments. Monitoring these components is crucial because their complex and time-consuming installation procedures can lead to project delays.

The developed framework consists of three main modules: (i) Object Recognition (ii) Object Localization, (iii) Integrated Object Recognition and Localization. In the "Object Recognition" module, two deep learning algorithms, YOLACT++ and Mask R-CNN, were utilized in processing digital images captured at construction sites for the automated recognition of tracked building elements. YOLACT++ proved superior to Mask R-CNN and was accordingly utilized in the developed framework. In the "Object Localization" module, a Real-time Locating System (RTLS) is utilized to identify the location of each recognized element along with its ID. The Ultra-wideband (UWB) system was selected as an RTLS, and different laboratory and field experiments were conducted to validate the UWB system's localization performance. Finally, in the "Integrated Object Recognition and Localization" module, a user-friendly application was developed to

integrate the outputs from the YOLACT++ model and the UWB system and automatically generate status reports of tracked elements. These reports include visual and location information, along with the unique ID of each element.

The framework was tested and validated using 3,632 images. The results demonstrate good performance and effectiveness of the developed framework under challenging conditions; yielding recognition accuracy of close to 85% in precision and recall for HVAC duct and slightly less than that for pipes. Similar performance was achieved in localization, yielding errors ranging from 0.03 to 1.22 meters in two-dimensional (2D) coordinates and from 0.15 to 1.6 meters in three-dimensional (3D) coordinates in the field test. The developed framework can be easily extended to other building elements, and the excel format of its output can facilitate linkage with Building Information Modeling (BIM) systems.

ACKNOWLEDGMENTS

I would like to thank my supervisors, Prof. Moselhi and Prof. Khorasani, for their invaluable guidance, support and supervision throughout my PhD studies. Their continuous mentorship, insight, and encouragement have made me feel truly honored to be their student, teaching me life lessons I will never forget.

I would also like to express my appreciation to my examining committee members – Dr. Amin Hammad, Dr. Sang Hyeok Han, Dr. Mazdak Nik-bakht and Dr. Jeff Rankin for their insightful comments and feedback, which significantly contributed to the improvement of my work.

Special thanks go to my parents, whose unconditional love and guidance have been my constant support. Their encouragement has lightened my every day and helped me stay positive and strong through different stages of my life. Special thanks to my uncles, Ardeshir and Babak, for all their support and motivation. I am very grateful for my grandparents' endless kindness, and care from when I remember. It is with deep regret that they are not here to see this achievement. Lastly, I extend my deepest thanks to Ashkan Baadi, who stood by my side during this journey with love, encouragement and support.

*Dedicated to my loved ones: my dear parents, my uncles, my grandparents
for all love, courage and support.*

Table of Contents

LIST OF FIGURES	ix
LIST OF TABLES	xii
LIST OF ABBREVIATIONS.....	xiii
CHAPTER 1: INTRODUCTION	1
1.1 Problem Statement and Motivations.....	2
1.2 Objectives	3
1.3 Research Methodology	4
1.4 Thesis Organization	5
1.5 Summary and Conclusions	6
CHAPTER 2: LITERATURE REVIEW	8
2.1 Overview.....	8
2.2 Construction Progress Monitoring and Reporting	8
2.2.1 Computer Vision Techniques	10
2.2.2 Application of Location Tracking Technologies in Construction	24
2.2.3 Application of BIM in Monitoring Construction Sites	42
2.3 Summary and Conclusions	45
CHAPTER 3: RESEARCH METHODOLOGY.....	48
3.1 Introduction.....	48
3.2 Object Recognition Module.....	49
3.2.1 Image collection and labeling.....	52
3.2.2 Model Training	55
3.2.3 Model Evaluation.....	64
3.3 Object Localization Module	66
3.3.1 The UWB System	66
3.3.2 The UWB System’s Configurations	71
3.4 Integrated Object Recognition and Localization	73
3.4.1 The Integrated Model Process	75
3.5 Summary and Conclusions	77

CHAPTER 4: MODEL IMPLEMENTATION and VALIDATION.....	79
4.1 Automated Recognition of MEP Components in Indoor Job Sites.....	79
4.1.1 Overview.....	79
4.1.2 Object Recognition with Synthetic Image Datasets.....	80
4.1.3 Object Recognition Using Mixed Synthetic and Real Images with Two Deep Learning Models.....	83
4.1.4 MEP Recognition using YOLACT++	88
4.1.5 Summary and Conclusions	95
4.2 Object Localization using the UWB System	96
4.2.1 Overview.....	96
4.2.2 Laboratory Experiments under Line-of-Sight (LOS) Conditions.....	96
4.2.3 Field Experiments under Line-of-Sight (LOS) and Non-Line-of-Sight (NLOS) Conditions.....	108
4.2.4 Summary and Conclusions	122
4.3 Integrated Object Recognition and Localization	123
4.3.1 Overview.....	123
4.3.2 The Developed Application for Data Integration	127
4.3.3 Summary and Conclusions	130
4.4 Limitations	131
CHAPTER 5: SUMMARY AND CONCLUDING REMARKS	134
5.1 Conclusions.....	134
5.2 Research Contributions.....	136
5.3 Future works	137
APPENDICES	140
REFERENCES	148

LIST OF FIGURES

Figure 1-1. The research methodology.	6
Figure 2-1. Construction progress monitoring and reporting process (Shamsollahi et al. 2022) ...	9
Figure 2-2. a) Original Image b) Object detection result c) Semantic segmentation result d) Instance segmentation result (Shamsollahi et al. 2022).	14
Figure 2-3. Comparison between the two techniques: a) Feature-based algorithms, b) Deep learning models (Wang et al. 2018).	16
Figure 2-4. Synthetic composite image sample (Hwang et al. 2023).	21
Figure 2-5. Virtual synthetic image sample.	21
Figure 2-6. Indoor localization technologies and their performance metrics.	27
Figure 2-7. RFID hardware components including the reader and different types of tags and RFID tag printer (Montaser and Moselhi 2014).	31
Figure 2-8. Sample of (a) UWB receivers and (b) tag from the DecaWave Company (DWM1000) (Yao et al. 2021).	32
Figure 2-9. Line-of-sight and non-line-of-sight conditions, and multipath effects (Sang et al. 2020).	34
Figure 2-10. Samples of different UWB commercial products: a) Ubisense, b) BeSpoon and c) DecaWave (Ruiz and Granja 2017).	36
Figure 2-11. Sample of integration of BIM, RTLS and cloud-based system (Fang et al. 2016)..	44
Figure 3-1. Overview of the research methodology.	48
Figure 3-2. Overview of the object recognition module.	51
Figure 3-3. Generated synthetic images under different conditions: (a) Lighting condition; (b) Complexity level; (c) Scale level; (d) Occlusion degree; (e) Viewpoint level (Shamsollahi et al. 2024)	54
Figure 3-4. Mask R-CNN network architecture (Shamsollahi et al. 2021).	56
Figure 3-5. Samples of applied data augmentation techniques.	58
Figure 3-6. 3*3 Standard and deformable convolution architecture (Shamsollahi et al. 2024). ..	60
Figure 3-7. YOLACT++ network architecture (Shamsollahi et al. 2024).	61
Figure 3-8. Original images and the Protonet outputs (Shamsollahi et al. 2024).	62
Figure 3-9. Sample of applied data augmentation techniques for the YOLACT++ model (Shamsollahi et al. 2024).	63

Figure 3-10. Box IoU and mask IoU (Shamsollahi et al. 2024).	65
Figure 3-11. Process of implementing the UWB system for object localization.....	69
Figure 3-12. Sample of the UWB output file.....	70
Figure 3-13. Trek1000 tags and receivers' components.	72
Figure 3-14. User interface of the UWB application.	73
Figure 3-15. Overview of the integrated method.....	75
Figure 3-16. Steps of the integrated framework.	76
Figure 4-1. 3D BIM models and their MEP networks.....	80
Figure 4-2. Training and validation losses during 90 epochs in the Experiment #2 (Shamsollahi et al. 2021)	82
Figure 4-3. Results of the Mask R-CNN prediction	83
Figure 4-4. The presence of a different number of HVAC ducts in each image.	84
Figure 4-5. Loss curves of Mask R-CNN: (a) Without data augmentation (b) With data augmentation.....	85
Figure 4-6. Examples of inaccurate recognition of objects by Mask R-CNN.....	86
Figure 4-7. Output images from Mask R-CNN and YOLACT++ models.	88
Figure 4-8. The types and numbers of images within the dataset.....	89
Figure 4-9. Comparison of bounding box and mask mAP values across different datasets (Shamsollahi et al. 2024)	90
Figure 4-10. Output results of YOLACT++ trained on two different datasets (Shamsollahi et al. 2024)	92
Figure 4-11. Examples of failures in the MEP recognition using YOLACT++ (Shamsollahi et al. 2024)	94
Figure 4-12. (a) Lab area (b) Layout plan and location of the receivers.	98
Figure 4-13. Placement of the tags in the test area: a) Experiment #1 and b) Experiment #2....	101
Figure 4-14. UWB data readings for tags on 2D and 3D planes in (a) Experiment #1 and (b) Experiment #2.....	102
Figure 4-15. 2D and 3D errors of the tags in Experiment #1 and Experiment #2.	103
Figure 4-16. Placement of the tags in the test area: a) Experiment #3 and b) Experiment #4....	104
Figure 4-17. UWB data readings for three tags on 2D and 3D planes in (a) Experiment #3 and (b) Experiment #4.....	107

Figure 4-18. Mechanical room at Concordia University.	108
Figure 4-19. Mechanical room plan and the locations of the receivers in the room.....	110
Figure 4-20. Distribution of the tags in the mechanical room.	112
Figure 4-21. Examples of conditions between tags and receivers.	114
Figure 4-22. The tags' actual locations, UWB average locations, and data points on a 3D Plane in (a) Test #1 and (b) Test #2.....	117
Fig 4-23. Histogram of 2D and 3D errors of the UWB tags: (a) Tag-7 and (b) Tag-6.....	119
Figure 4-24. Histogram of 2D and 3D error values for Experiments #1- Experiments #4.....	121
Figure 4-25. An overview of the integrated framework.	126
Figure 4-26. Sample images taken from the mechanical room and their corresponding predicted images by YOLACT++.	127
Figure 4-27. Visual studio environment for designing the application user interface.....	128
Figure 4-28. The application user interface.	129

LIST OF TABLES

Table 2-1. Applied object recognition algorithms in the construction industry.	18
Table 2-2. Overview of object recognition models for construction monitoring.	22
Table 2-3. Different applications of location tracking technologies in construction.....	28
Table 2-4. Comparison between different UWB manufacturers' system configurations (Jiménez & Seco, 2016; Ruiz & Granja, 2017; J. Wang et al., 2015).....	34
Table 2-5. Summary of the research studies that utilized the UWB system in construction projects.....	39
Table 2-6. Capabilities and limitations of RTLS technologies.....	41
Table 3-1. Data augmentation techniques and selected parameters for Mask R-CNN training. ..	57
Table 3-2. Applied data augmentation techniques, their parameters, and values.....	62
Table 4-1. Results of HVAC duct recognition with synthetic test dataset.	81
Table 4-2. The results of Mask R-CNN and YOLACT++ algorithms for HVAC duct recognition.	87
Table 4-3. Distribution of synthetic and real images in each dataset and the number of instances.	89
Table 4-4. The results of YOLACT++ with two object classes	93
Table 4-5. Collected datapoints in each experiment.....	99
Table 4-6. True location of tags, mean and standard deviation of UWB data.....	99
Table 4-7. UWB localization errors on 2D and 3D planes.....	106
Table 4-8. Tagging details of each tracked component.....	113
Table 4-9. Actual locations of tags within the mechanical room.....	114
Table 4-10. Distances between the tags and receivers.....	114
Table 4-11. Captured datapoints in Test #1 and Test #2.....	115
Table 4-12. Mean, standard deviation and error of the UWB system for Test #1 and Test #2. .	118
Table 4-13. The tags' information in each experiment.....	122
Table 4-14. YOLACT++ detection performance for HVAC ducts and pipes.....	127
Table 4-15. Record samples from the generated reports.....	129

LIST OF ABBREVIATIONS

2D	Two Dimensional
3D	Three Dimensional
ACID	Alberta Construction Image Data Set
AI	Artificial Intelligence
AOA	Angle of Arrival
AP50	Average Precision at 50% Overlap.
BCE	Binary Cross Entropy
BIM	Building Information Modeling
BLE	Bluetooth Low Energy
BRIEF	Binary Robust Independent Elementary Features
C#	C-sharp
CLAHE	Contrast Limited Adaptive Histogram Equalization
CV	Computer Vision
DCN	Deformable Convolutional Neural Networks
DCNv2	Deformable Convolutional Neural Networks Version 2
DSOD	Deeply Supervised Object Detector
Fast R-CNN	Fast Region-based Convolutional Network
Faster R-CNN	Faster Region-based Convolutional Neural Network
FCN	Fully Convolutional Network
FN	False Negative
FP	False Positive
FPN	Feature Pyramid Network
FPS	Frames Per Second
GPS	Global Positioning Systems
HOG	Histogram of Gradients
HSV	Hue, Saturation, and Value
ID	Unique Identifier
IoU	Intersection over Union
IR	Infrared

ISeg	Instance Segmentation Capability
KNN	K-Nearest Neighbors
LBP	Local Binary Pattern
LOS	Line-of-Sight
mAP	mean Average Precision
mAP^{bbox}	Bounding Box mean Average Precision
mAP^{mask}	Mask mean Average Precision
Mask R-CNN	Mask Region-based Convolutional Neural Network
MEP	Mechanical, Electrical, and Plumbing
MOCS	Moving Objects in Construction Site
MS COCO	Microsoft COCO
NIBS	National Institute of Building Sciences
NLOS	Non-Line-of-Sight
PCB	Printed Circuit Board
M	Predicted Masks
M_{gt}	Ground Truth Annotations
RANSAC	RANdom SAmples Consensus
R-CNN	Region-based CNN
ResNet	Residual Neural Network
RFID	Radio Frequency Identification
RGB	Red, Green and Blue
RoIs	Regions of Interest
RPN	Region Proposal Network
RSSI	Received Signal Strength Indicator
RTLS	Real-Time Locating System
SFM	Structure From Motion
SIFT	Scale Invariant Feature Transform
SODA	Site Object Detection Dataset
SOLOv2	Segmenting Objects by Locations, version 2
SSD	Single-Shot Multi-box Detector
SURF	Speeded Up Robust Features

SVM	Support Vector Machine
TDOA	Time Difference of Arrival
T_{tag}	Timestamp of the UWB System
T_{image}	“Date taken” Metadata from Each Image
TOA	Time of Arrival
ToF	Time of Flight
TP	True Positive
TWR	Two-Way Ranging
UAVs	Unmanned Autonomous Vehicles
USB	Universal Serial Bus
UWB	Ultra-Wideband
VIA	VGG Image Annotator
VOC	Visual Object Classes
YOLACT++	You Only Look At CoefficientTs++
YOLO	You Only Look Once
YOLOv3	You Only Look Once, version 3
YOLOv4	You Only Look Once, version 4

CHAPTER 1: INTRODUCTION

Efficient progress monitoring systems in construction projects assist management teams to successfully meet their objectives, minimizing cost overruns and delays (T. Omar & Nehdi, 2016; W. Wei et al., 2023). These systems provide accurate and timely information about the project's current status, allowing for comparison with the planned state to identify progress deviations (Reja et al., 2022). Once deviations are identified at an early stage, problems can be addressed before they develop into more complex issues that necessitate costly remedies (Navon & Shpatnitsky, 2005). However, numerous projects have experienced cost overruns and delays during the construction phase due to inefficient monitoring systems. These systems are often incapable of collecting and processing data from construction sites in an accurate and timely manner (H. Omar et al., 2018). These issues arise from traditional approaches that are highly dependent on the expertise of site personnel, including manual site monitoring, assessments, and the preparation of reports (Kopsida et al., 2015; H. Omar et al., 2018). The required information is then extracted from these reports and submitted along with other documents, including paper-based drawings, sheets and charts (Kropp et al., 2012). As a result, manual progress reports can be unreliable and misinterpreted due to possible errors and delays in report submissions (Kopsida et al., 2015; H. Omar et al., 2018).

Progress monitoring in indoor construction sites is critical due to the involvement of various activities and their significant impact on the project schedule (Ekanayake et al. 2021; Hamledari et al. 2017). However, indoor progress monitoring is more challenging than in outdoor environments due to the presence of numerous detailed building elements (Ekanayake et al., 2021; Kopsida et al., 2015). Indoor construction activities often contain complex and interrelated components within congested spaces, which makes visual assessments significantly complicated. Therefore, developed monitoring systems suitable for outdoor environments may not be sufficient for indoor job sites (Ahmed et al., 2012; Koo & Fischer, 2000; Kopsida et al., 2015; Kopsida & Brilakis, 2020).

In recent years, new technologies and methods have been applied in construction projects to improve the performance of progress monitoring and reporting systems through innovative solutions. These technologies can enhance efficiency of collecting, recording, analyzing, and

displaying data captured from construction sites. Camera-based systems, Artificial Intelligence (AI), Computer Vision (CV), Real-time Locating System (RTLS), Building Information Modeling (BIM), and Cloud-based applications are some examples of these technologies. Additionally, these technologies can be integrated to mitigate their individual shortcomings and ensure that required information about tracked resources such as location and visual information is obtainable (Ibem & Laryea, 2014; Noruwa et al., 2020; Pour Rahimian et al., 2020; Rafiee et al., 2013). Such integrated systems can provide timely, accurate, and reliable information, assisting project managers in understanding project progress.

1.1 Problem Statement and Motivations

Monitoring progress in construction sites should be conducted repetitively to timely evaluate the ongoing development of projects (Alaloul et al., 2021; W. Wei et al., 2023). This repetitive process applies to data collection, analysis and reporting to determine deviations between the as planned and actual status of projects (Patel et al., 2021). However, traditional approaches such as manual field visits and reporting have led to issues like inconsistent and infrequent updates and reports which can contribute to project delays (Pal et al., 2023).

Technological advancements in data collection, recording and analysis has enhanced the efficiency of automated progress monitoring and reporting systems in construction (Pal et al., 2024). Laser scanners and camera-based systems provide visual data from construction sites. Laser scanners deliver precise 3D point clouds of the as-built status (Moselhi et al., 2020), however, they cannot provide the required information in a timely manner. They have challenges with reflective materials, and their hardware is expensive (Hamledari et al., 2017; Maalek & Sadeghpour, 2013; Z. Wang et al., 2021). In contrast, camera-based monitoring systems, are cost-effective, practical, and reliable alternative (Teizer, 2015). These systems capture images from construction sites, providing visual information essential for understanding the as-built state (M. W. Park & Brilakis, 2016; J. Yang et al., 2016). Advances in computer vision (CV) techniques and deep learning algorithms allow automated analysis of images for extracting needed information about the project's current status (Pal & Hsieh, 2021; Seo et al., 2015; J. Yang et al., 2016). Particularly, object detection and segmentation techniques have shown highly improved performance with deep learning algorithms (Pal & Hsieh, 2021). However, the following problems are associated with digital imaging and deep learning models: (1) Optimal performance from deep learning algorithms

requires a large, high-quality image dataset for model training (J. Kim et al., 2020), which is challenging in the construction industry due to lack of open datasets and confidentiality concerns, (2) Achieving adequate accuracy in the recognition of complex objects with irregular geometric or slender shapes within indoor environments, remains a challenge, (3) The complex computational requirements of some image processing techniques can be time-consuming and require high levels of hardware and software resources, and (4) Object recognition models are unable to accurately localize recognized objects within a 3D plane based on images (Zhou et al., 2021).

RTLS is another advanced technology which is mainly used for object localization and tracking in construction sites. These systems are differentiated by their localization accuracy, cost, scalability, level of consistency, robustness, and data protection (Alarifi et al., 2016). However, RTLS technologies have several limitations: (1) They lack visual information about tracked resources, which is a critical factor for the monitoring and validation of project progress, (2) Monitoring metallic elements which can cause signal blockage and reflections as well as high multi path environments are challenging for some RTLS types, especially Radio Frequency Identification (RFID) and Bluetooth Low Energy (BLE) systems (Moselhi et al., 2020; C. Zhang et al., 2020), and (3) They may fail to deliver accurate 3D location information at the centimeter level. This level of accuracy depends on the objects being localized and the specific applications.

Lastly, there is an absence of integrated vision-based technologies and RTLS that provide comprehensive information about the status of tracked components. This becomes even more critical when a BIM model is not available to provide information about the elements. An integrated system significantly increases capabilities as compared to using each system individually. This approach not only enhances the strengths of the entire monitoring system but also effectively minimizes their individual constraints. Additionally, the lack of organized and unified reporting of collected information prevents project managers from making timely and appropriate decisions.

1.2 Objectives

The main objective of the research is to develop a method for automated progress monitoring and reporting to timely recognize and localize the tracked building elements in indoor construction sites. The method aims to provide comprehensive information about these elements for identifying

their as-built status so that decision-making can be supported. This objective is divided into the following sub-objectives:

- 1- To strategically employ different data collection and analysis methods to gather comprehensive data from tracked objects in construction sites, thereby improving monitoring systems and supporting data-driven decisions.
- 2- Implement and evaluate novel object recognition models using digital imaging and deep learning algorithms to obtain visual data and recognize tracked building elements at construction sites in a timely manner.
- 3- Improve the performance of object recognition models in identifying complex objects in challenging construction environments through enhanced dataset quality. This includes generating synthetic images that represent actual indoor scenarios and finding the optimum mix of synthetic and real images for efficient model training.
- 4- Implement a reliable RTLS to accurately localize tracked elements within complex and dynamic job sites in real-time. Assess different factors that may affect its performance in construction sites and its integration capability with other available techniques.
- 5- Develop a model to integrate the outcomes of object recognition models and localization systems. This integration will report the required information about each component including its specific ID, location, visual information in an organized and understandable format. This assists in understanding their quantity, installation status and consequently understanding the actual progress of the project.

1.3 Research Methodology

Figure 1-1 shows the research methodology to meet the objectives stated in the previous section. This study started with a problem statement and the objectives. Then a comprehensive literature review was accomplished in the following domains:

- Recent advances in construction progress monitoring and reporting.
- Application of vision-based techniques for construction progress monitoring.
- Application of location tracking technologies in construction projects.
- Application of integrated monitoring methods and BIM in construction projects.

In the next step, shortcomings and limitations in the literature are identified to develop the research methodology which consists of three modules including “Object Recognition”, “Object Localization” and “Integrated Object Recognition and Localization”. Then the developed method is implemented and evaluated in different case studies. Finally, the results, conclusion, and future works are described in the last chapter.

1.4 Thesis Organization

This research study includes five chapters. Following Chapter 1, the chapters are as follows:

Chapter 2 provides an overview of the previous works focusing on automated progress monitoring and reporting, vision-based techniques, RTLS, integrated approaches and BIM within the construction industry. This chapter identifies advances in automated progress monitoring through the utilization of different technologies and methods. Finally, a summary of existing gaps in the literature is provided at the end of this chapter. This chapter assists in identifying the techniques and tools that have the capabilities to be implemented in this research study and highlights domains for possible improvements.

Chapter 3 describes the overview of the research methodology to support automated progress monitoring and reporting systems in construction projects. Three main modules were defined including “Object Recognition”, “Object Localization” and “Integrated Object Recognition and Localization”.

Chapter 4 implements and evaluates the methods described in Chapter 3 using different laboratory and field experiments. Section 4.1.4 of this chapter is a modified version of “Automated Detection and Segmentation of Mechanical, Electrical, and Plumbing (MEP) Components in Indoor Environments by Using the YOLACT++ Architecture” published by *Journal of Construction Engineering and Management* (Shamsollahi et al. 2024). Moreover, Sections 4.2 and 4.3 are extended versions of “A Data Integration Method Using a Deep Learning Algorithm and Real-Time Locating System (RTLS) for Automated Construction Progress Monitoring and Reporting” published by *Automation in Construction* (Shamsollahi et al. 2024).

Chapter 5 provides a summary and conclusion of this study and highlights its contributions along with recommendations for future improvements.

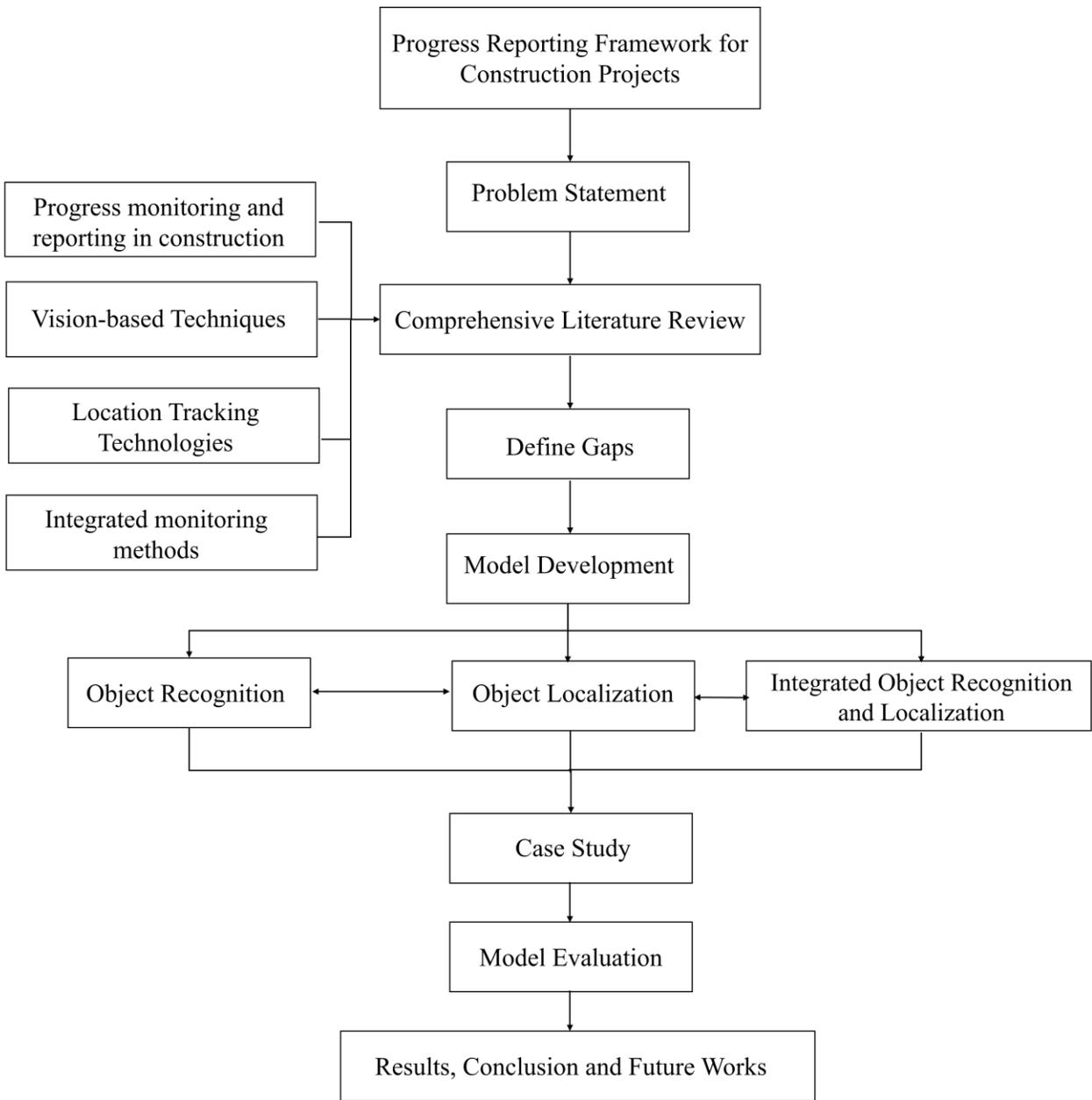


Figure 1-1. The research methodology.

1.5 Summary and Conclusions

The purpose of this chapter is to establish the focus of this research, which is automated progress monitoring and reporting in construction environments, to improve efficiency and address the problems associated with manual monitoring methods. Next, the problem statement is identified,

including challenges related to monitoring tracked elements in indoor environments and limitations of current technologies utilized for automated progress monitoring.

Following the problem statement, the research objectives were outlined. These objectives include developing a method that can accurately recognize and localize tracked building elements in indoor job sites in a timely manner. This method aims to provide detailed and comprehensive information about these elements for identifying their current status and supporting the decision-making process. The methodology to achieve these objectives is then provided. Finally, the organization of the thesis was presented, encompassing a comprehensive literature review in Chapter 2, the research methodology in Chapter 3, the model implementation and validation in Chapter 4, and the summary and concluding remarks in Chapter 5.

CHAPTER 2: LITERATURE REVIEW

2.1 Overview

In this chapter, related research on automated progress monitoring and reporting in construction projects is reviewed. The application of different technologies to enhance construction progress monitoring is discussed. These technologies include digital cameras and computer vision (CV) techniques for collecting and analysing visual information, Real-time Locating System (RTLS) for object localization, and integrated methods using both visual and location data. Furthermore, the role of Building Information Modeling (BIM) for further analysis of collected data from construction sites is detailed. The chapter ends by identifying the existing gaps in these methods.

2.2 Construction Progress Monitoring and Reporting

Construction progress monitoring and reporting are key managerial tasks for timely project completion and staying within budget. Through these tasks, discrepancies or unsatisfactory performance between the as-built and as-planned states of the project can be identified, allowing corrective actions to be taken at the right time (Golparvar-Fard et al., 2011; Moselhi et al., 2020). Inefficient progress monitoring has brought more than 53% delays and 66% over budget in construction projects (Alaloul et al., 2021; K. Han et al., 2018).

Monitoring the construction sites in a consistent manner assists the project managers in avoiding unforeseen expenses derived from schedule delays, poorly performed tasks, revisions, conflicts and improper resource management (Kopsida et al., 2015; Yates & Epstein, 2006). Nevertheless, in complex construction sites, identifying the actual status of the project provides challenges for decision-makers due to reliance on traditional systems which are manual, inaccurate, and slow. In such systems, humans are responsible for data collection from job sites and information extraction from both as-planned and as-built states. This information is derived from different documents such as drawings, surveys, schedules and site reports which all of them are submitted by different responsible parties in the project. Moreover, reports from job sites are often based on human assessments, which may not represent real site conditions. To address the above stated issues, it is crucial to have an automated system to collect and analyze data accurately, visualize and report the findings in an interpretable format for different responsible parties (Golparvar-Fard et al., 2015).

Recently, new technologies have been introduced and applied in the construction industry to automate the processes of progress monitoring and reporting. These processes are: (a) data collection from as-built/as-is scenes, (b) data analysis, (c) progress estimation by comparing as-built and as-planned information, (d) visualization of the results (Kopsida et al., 2015) as shown in Figure 2-1.

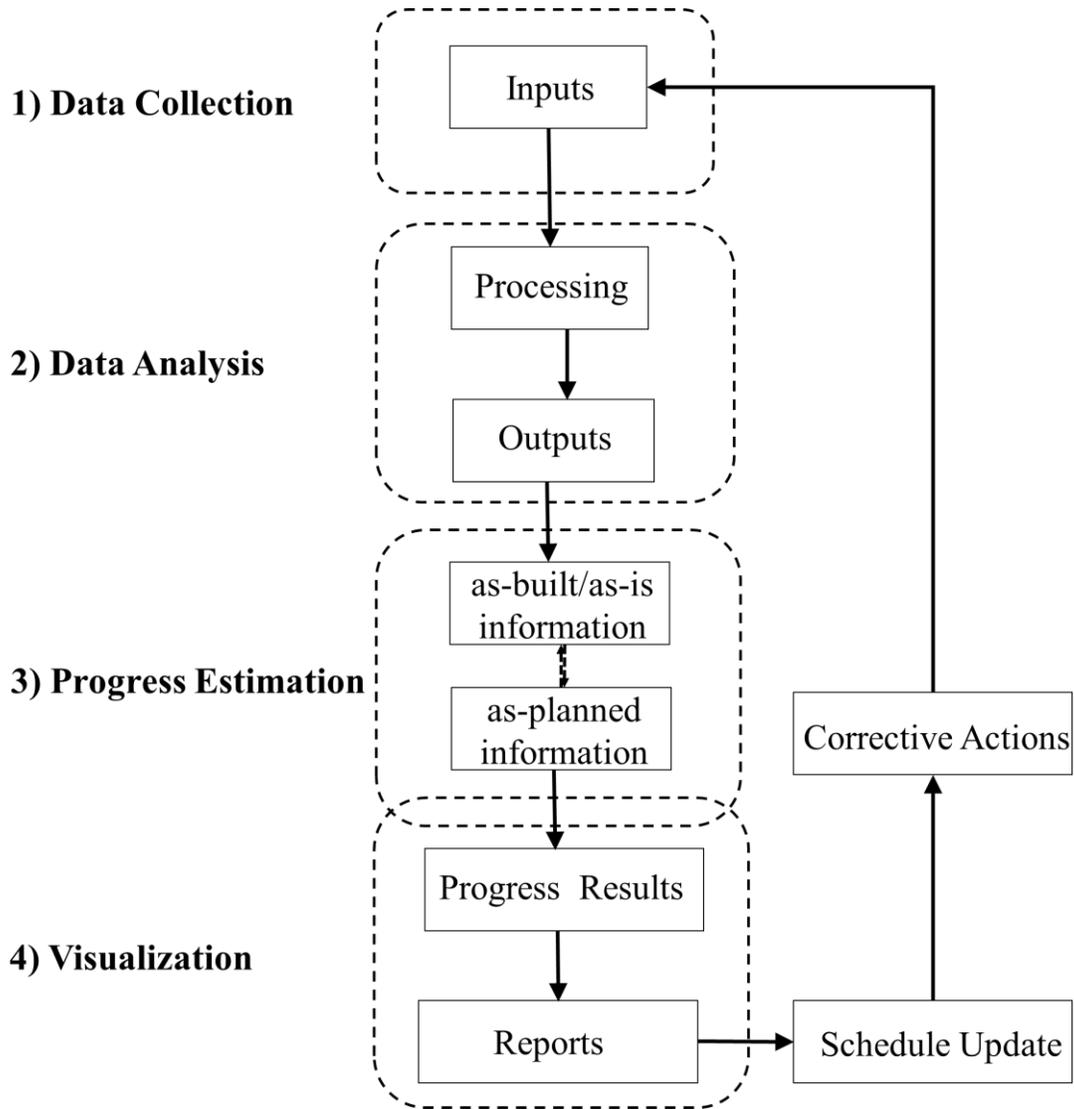


Figure 2-1. Construction progress monitoring and reporting process (Shamsollahi et al., 2022).

For evaluating automated progress monitoring systems, Kopsida et al. (2015) defined eight criteria, including:

1. Utility: The system's applicability and generalizability to both outdoor and indoor environments and encompass different objects for monitoring purposes.
2. Time efficiency: The duration of time spent using the system.
3. Accuracy: The level of accuracy and reliability the system provides for tasks involved in progress monitoring, from data collection to progress estimation.
4. Automation Level: The degree of human involvement in completing the required tasks.
5. Preparation requirements: The time required to set up the system.
6. Requirements for training the user: The level of expertise needed to operate the system.
7. Cost: All expenses related to the system's equipment, deployment, and maintenance.
8. Mobility: The capability to move the system without restrictions and effort.

2.2.1 Computer Vision Techniques

In recent years, low-priced and high-resolution digital cameras with high-capacity memory have enabled construction companies to effectively capture data from construction operations. Digital cameras can produce a large number of images and videos on a daily basis from as-built scenes containing useful and detailed information (Golparvar-Fard et al., 2015; Hou et al., 2020). However, due to challenges in image analysis tools such as computation time, accuracy and cost, images are analysed manually only for documentation and data recording purposes. As a result, only a small portion of this information is utilized, while the rest becomes unusable (Hou et al., 2020; Nieto et al., 2012; Paneru & Jeelani, 2021).

With developments in hardware platforms and algorithms, CV technology improved significantly (Feng et al., 2019). CV is a branch of artificial intelligence that uses computers to obtain high-level understanding from visual data, like human visual systems. During the last decade, CV has attracted many researchers due to its wide range applications in enhancing automation in construction. It can be applied for different project management purposes such as safety monitoring, quality control, productivity analysis and progress monitoring (Xu et al., 2020).

In progress monitoring systems using CV techniques, the collected visual data from job sites are analyzed to understand the project's current state. Computer vision techniques can be categorized into 3D scene reconstruction, object tracking, object detection and image segmentation

(Paneru & Jeelani, 2021). In recent years, the number of research studies related to automated progress monitoring and computer vision techniques in the construction industry has increased. Kopsida et al. (2015) provided an in-depth review of automated progress monitoring steps and their related technologies and methods. Patel et al. (2021) explored recent developments, existing challenges, and future works for automated progress monitoring in the construction industry. Reja et al. (2022) and Sami Ur Rehman et al. (2022), described the current computer vision algorithms that are applicable in construction projects and conducted a detailed comparison of them. Additionally, they discussed how CV techniques can be integrated with other technologies, including BIM, augmented reality, and virtual reality.

Gharib and Moselhi (2023), conducted a review of current practices for automated progress monitoring and reporting using CV techniques. As part of this research, devices for capturing visual data were described, including depth cameras, drones, and laser scanners. Ekanayake et al. (2021), described the challenges associated with indoor progress monitoring and reviewed various computer vision techniques that have been used in recent research studies for indoor job sites.

An in-depth literature review reveals that three major computer vision techniques are used for construction progress monitoring and reporting using digital images, namely (i) 3D scene reconstruction, (ii) object detection, (iii) image segmentation. In this section, recent algorithms and technologies in these areas are discussed and the challenges are highlighted.

2.2.1.1 3D Scene Reconstruction

In this technique, 3D representations (mesh models, point clouds and geometric models) are generated from one or multiple images taken from construction sites (Lu & Lee, 2017; Xue et al., 2021). These 3D representations contain critical information pertinent to the current state of the project, which can then be compared to the as-planned state to track and report the project progress. For this purpose, collected data from cameras (monocular, stereo, video, panoramic, and RGB-Depth) is required to generate the point cloud models (Ma & Liu, 2018; Xue et al., 2021). Many review papers have been published in the past few years that provide more information about the recent advancements in 3D reconstruction techniques, their capabilities, and limitations (Ham et al., 2019; Kang et al., 2020; Mirzaei et al., 2022; Wang et al., 2020; Wang & Kim, 2019; Xu et al., 2021; Xue et al., 2021).

Computer vision techniques and algorithms for generating 3D scene reconstruction are different due to the characteristics of the input images. The input images are categorized into single and multiple images (Ham et al., 2019). Single images can be taken using regular cameras or RGB-Depth cameras such as Azure Kinect. The Azure Kinect camera can easily create as-built 3D scenes using streams of depth and color images. However, for creating 3D reconstruction scenes using regular cameras, there is a need to calculate the depth of pixels in the images using computer vision techniques (Kang et al., 2020). Eder et al. (2019) developed and trained convolutional neural networks with a dataset containing RGB-D images to predict depth estimation of a single 360° image of an indoor scene that provides all information for creating the 3D as-is model.

The multiple images are divided into (i) multi-perspective 2D images and (ii) video sequences. In general, 3D scene reconstruction using multiple images have fewer challenges and is studied more frequently in the literature since they are more accurate with a higher level of detail compared to approaches using single images. In multi-perspective 2D images, several images with different perspectives of the objects are taken and create 3D scene representations based on parameters and poses of cameras (Ham et al., 2019; Kang et al., 2020). Fathi et al. (2011) proposed a framework to create 3D point clouds using two calibrated cameras. The feature points captured from two video frames were detected using the Speeded Up Robust Features (SURF) algorithm. Automatic point matching between two video frames using Euclidean distance was applied and the outliers were removed using the RANdom SAMple Consensus (RANSAC) algorithm. Then triangulation was used to estimate spatial coordinates of the points in the frames and create point clouds of the construction objects on site.

Video sequences can be utilized as an input for computer vision techniques to reconstruct 3D scenes. One of these techniques is the structure from motion (SFM) method that uses the shared information between consecutive frames by repeatedly extracting and matching features between two images, filtering outliers, and estimating poses of images and point clouds through image registration and triangulation (Ham et al., 2019; Jiang et al., 2020; Kang et al., 2020). In the study done by Golparvar-Fard et al. (2011) unordered daily photographs were used to reconstruct the as-built environment by using the SFM technique. Creating 3D point cloud models, enabled project management team to visualize the project's current state through different viewpoints.

Despite many advances in 3D scene reconstruction techniques, still some limitations exist that need to be discussed. Some related issues are (1) Lack of automation level in all the required steps for creating as-built models such as data collection and removing outliers. This increases, the operation time, and errors in the models (Lu & Lee, 2017; Ma & Liu, 2018); (2) Lighting conditions, occlusions, and cluttered backgrounds are unavoidable in construction sites which make 3D representations incomplete and noisy (Han et al., 2021; Xue et al., 2021); (3) Limited operating spaces in indoor environments (Xue et al., 2021); (4) Incapability of existed techniques in reconstructing of building elements with complicated geometric shapes (cylindrical, spherical, L-shaped, etc.) which mostly are in indoor environments (Kang et al., 2020).

2.2.1.2 Object Detection Algorithms

Due to numerous construction activities which use a wide variety of resources including materials, equipment, and workers, it is important to identify which resources are in the scene and which ones are involved in performing the task of interest (Seo et al., 2015). Object detection is used to identify tracked building components automatically on site from the captured images and videos. This technique facilitates analysis of tracked activities and material allocation to support progress monitoring and reporting (Lin & Golparvar-Fard, 2020). Object detection is a computer vision task that performs both classification and localization. Meaning that it classifies the objects in the captured image into pre-defined categories and predicts the location of each object in the image as shown in Figure 2-2.b (Athira & Khan, 2020; Seo et al., 2015; Wu et al., 2020).

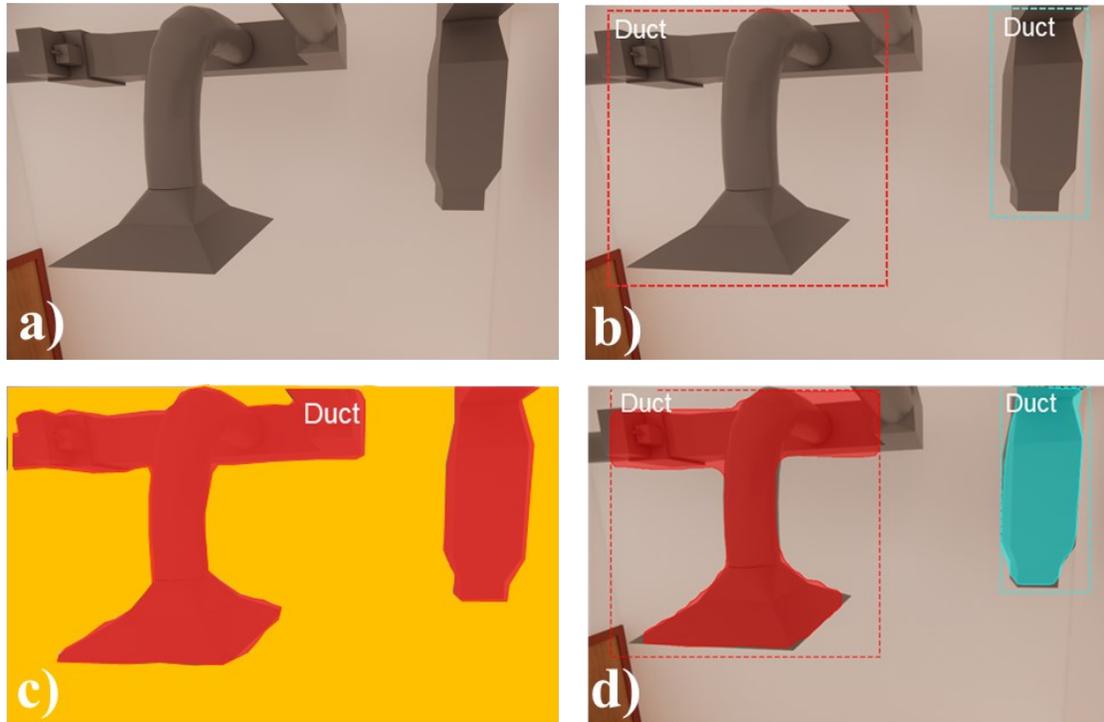


Figure 2-2. a) Original Image b) Object detection result c) Semantic segmentation result d) Instance segmentation result (Shamsollahi et al., 2022).

In the early stages of object detection, many researchers used traditional (feature-based) algorithms which are essentially performed in a step-by-step process, requiring a specific model for each task. In these algorithms, the image features are extracted using feature descriptors such as Scale Invariant Feature Transform (SIFT), Local Binary Pattern (LBP), Binary Robust Independent Elementary Features (BRIEF) and Histogram of Gradients (HOG). Next, these feature descriptors are combined with machine learning classifiers such as Support Vector Machine (SVM), K-Nearest Neighbors (KNN), naive Bayes classifiers and neural networks for classification tasks (Murphy et al., 2006; Seo et al., 2015; Wu et al., 2020). The performance of these models highly relies on the optimization of implemented algorithms and the selection of image features (Lee, 2015; Wang et al., 2018; Wang et al., 2021).

Several research studies have implemented feature-based object detection algorithms to detect construction resources from visual data for various construction applications (Wang et al., 2021). For instance, to improve productivity in construction projects, Zou and Kim (2007) used feature-based algorithms for automatic excavator detection and analysis of its idle time in job sites.

In this study, the excavators were extracted from images using the Hue, Saturation, and Value (HSV) color space and simple thresholding methods. Next, to determine if the excavator was moving, the centroid coordinates of the object were calculated and compared with those of the previous image frame.

Hamledari et al. (2017) utilized different feature extraction and classification techniques such as color space selection, thresholding, edge extraction, and support vector machine to detect partitioned elements such as drywall, insulation, studs, and electrical outlets to report their actual state. While the proposed model can detect different components, its generalization is challenging because it requires different preprocessing and model-based techniques that require specialized expertise in the CV domain. Hui et al. (2015) proposed a framework to detect and localize bricks in video frames using image processing techniques such as color thresholding and edge detection. They then estimated the number of bricks on the building façade automatically to improve performance of progress monitoring. Traditional methods have limitations in model generalization for detection since they are based on hand-crafted features and require significant expertise for feature selection and extraction (Kim et al., 2018; Nath & Behzadan, 2019).

In recent years, deep learning algorithms consisting of neural networks with many hidden layers, have provided solutions with better performance and reduced human involvement. This is achieved by introducing end-to-end learning process, which means that for completing feature learning, classification and regression tasks only a dataset of annotated images or video frames is required (Kim et al., 2018; O'Mahony et al., 2019; Wang et al., 2018). In these models, the features such as edges, corners, and contours are transferred from the input layer to higher layers. This transfer results in a more abstracted feature representation that allows the system to learn complex inherent structures (Nath & Behzadan, 2019; Wang et al., 2018). The differences between traditional and deep learning algorithms are depicted in Figure 2-3.

The deep learning algorithms are classified into two main groups: two-stage and one-stage algorithms. In two-stage algorithms, object classifications and detections are based on a set of generated region proposals then, each proposal is classified, and the bounding box coordinates are refined (Carranza-García et al., 2021; Ekanayake et al., 2022; Pal & Hsieh, 2021; Ren et al., 2015). These algorithms include region-based CNN (R-CNN) and its extensions, Fast Region-based Convolutional Networks (Fast R-CNN) (Girshick, 2015), Faster Region-based Convolutional

Neural Network (Faster R-CNN) (Ren et al., 2015), and Mask Region-based Convolutional Neural Network (Mask R-CNN) (He et al., 2017). In contrast, one-stage detection algorithms consist of a single fully convolutional neural network to perform classification and predict bounding boxes without relying on proposal-based tasks (Carranza-García et al., 2021; Ekanayake et al., 2022). Single-shot multi-box detector (SSD) (Liu et al., 2016) You Only Look Once (YOLO) (Redmon et al., 2016) and its subsequent versions belong to one-stage algorithms. Generally, two stage algorithms are more accurate compared to the one-stage algorithms, but the latter are faster and can be applied for real-time purposes (Pal & Hsieh, 2021). However, an important challenge for both of these algorithms is their capability in handling objects with small scale in the image. This limitation has been observed in algorithms such as YOLO and Mask R-CNN (Pal & Hsieh, 2021; Q. Yang et al., 2020) which demonstrate a need for exploring other algorithms.

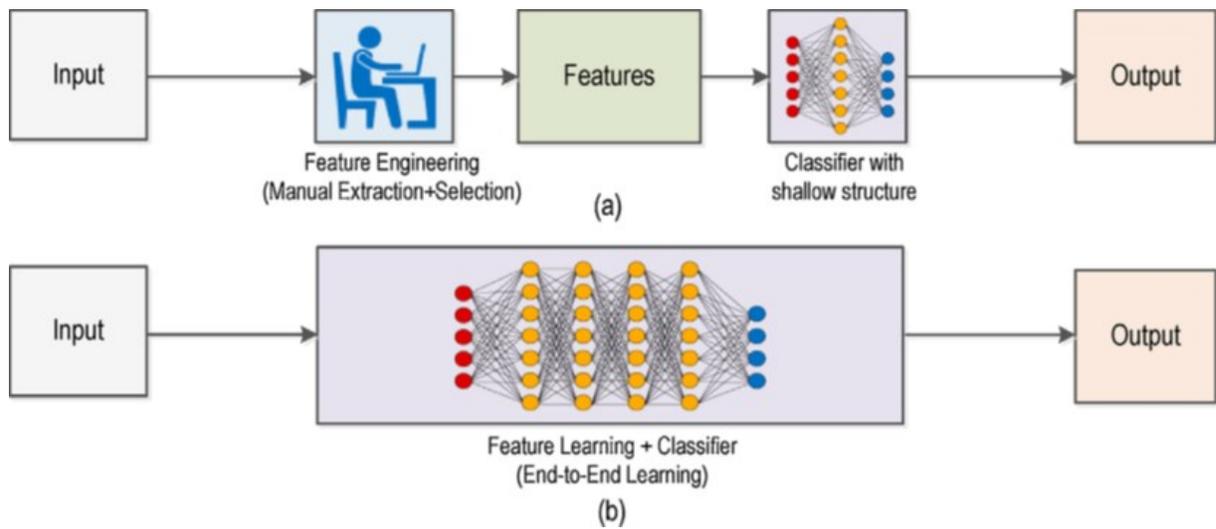


Figure 2-3. Comparison between the two techniques: a) Feature-based algorithms, b) Deep learning models (Wang et al., 2018).

Advances in deep learning algorithms and developments in hardware such as processing power, memory capacity, and high-resolution cameras have led to the rapid spread of CV techniques for various applications (O'Mahony et al., 2019). Many studies have applied deep learning-based object detection algorithms for construction progress monitoring and reporting. For example, Hou et al. (2020) trained Deeply Supervised Object Detector (DSOD) deep learning algorithm and detected building components including columns and beams automatically. Martinez et al. (2021) proposed a framework to track progress of construction tasks automatically

in offsite jobsites. In this research, Faster R-CNN is applied to detect and classify the construction resources that are utilized in each task. Pour Rahimian et al. (2020) developed a framework using CV techniques for building elements identification, integrating BIM and virtual reality to provide as-built information. Ekanayake et al. (2022) utilized You Only Look Once version 4 (YOLOv4), a real-time object detection model, in a cloud-based platform to automatically determine the wall partitions' status in indoor construction environments.

2.2.1.3 Image Segmentation Algorithms

In image segmentation, which is also named as pixel-level classification, a digital image is separated into different meaningful regions to find how objects are displayed in the image (Feng et al., 2019). Image segmentation can be divided into semantic segmentation and instance segmentation (Paneru & Jeelani, 2021).

Semantic segmentation refers to assigning a class label to each pixel in the image as shown in Figure 2-2.c (Hao et al., 2020). In instance segmentation, detection and segmentation are joined in one model, making detected objects distinguishable by pixel-wise masks. Compared to semantic segmentation, objects here from the same class, can be distinguished as separate instances as shown in Figure 2-2.d (Feng et al., 2019; Liu et al., 2019; Wu et al., 2020). Through segmenting objects of interest by predicted masks, shape and size of objects in the image can be identified. In addition, the object boundaries can be extracted, providing spatial information for further geometry analysis, localization, and tracking which can assist progress monitoring systems (Kang et al., 2022; Wang et al., 2021; Ying & Lee, 2019). Hence, as compared to object detection, instance segmentation is a harder task (Bolya et al., 2020).

Wang et al. (2021) developed an integrated framework using different CV tasks including instance segmentation and object tracking to monitor the progress of precast walls. Mask R-CNN was utilized for detecting and segmenting the walls and DeepSORT for tracking walls through consecutive frames. Wei et al. (2023) utilized Improved Segmenting Objects by Locations, version 2 (SOLOv2) which is a novel instance segmentation model to automatically monitor the progress of soil-foundation construction. Shamsollahi et al. (2021) implemented Mask-RCNN for automatic detection and segmentation of HVAC ducts from synthetic images. To overcome the overfitting problem and enhance the generalization of the model a set of data augmentation techniques was

applied during training the model. Table 2-1 provides more research studies that utilized deep learning-based object detection and segmentation algorithms for different construction domains such as safety management, productivity estimation, progress monitoring, and maintenance.

In recent years, many research studies focused on object recognition from point clouds as well (Mirzaei et al., 2023; Wei et al., 2022; Xie et al., 2023). For example, for automated monitoring of industrial facilities and implementing preventative maintenance, Agapaki and Brilakis (2020) developed and trained a deep learning algorithm called CLOI-NET in a cloud-based platform to automatically segment different types of industrial facilities from point clouds. Ma et al. (2020) investigated the performance of two novel deep learning models called PointNet and DGCNN for the automated recognition of different interior building components such as structural, openings, and furniture from point clouds. Both synthetic and real point cloud datasets were used for training of the models and DGCNN performed better in both cases. In addition, a mixture of synthetic and real dataset was used to train the DGCNN model, which yielded better results than using only real point cloud dataset.

Table 2-1. Applied object recognition algorithms in the construction industry.

Author	Objective	Training set	Test set	Key Algorithms	Object Type	Results			
						P	R	F1	AP ₅₀
Kang et al. (2022)	Safety	938 (Real images)	235 (Real images)	YOLACT	Excavator	-	-	-	0.86
					Truck	-	-	-	0.88
					Worker	-	-	-	0.89
					Safety Vest	-	-	-	0.88
					Hardhat	-	-	-	0.91
Techasarntikul and Mashita (2022)	Operation and Maintenance	529 (Real images)	95 (Real images)	Mask R-CNN	T-Rod	0.96	0.95	0.96	-
					Cable	0.86	0.92	0.88	-
					Metal bar	0.88	0.90	0.89	-
					Pipe	0.89	0.70	0.79	-
					Duct	0.59	0.83	0.69	-
					HVAC-E	0.82	0.82	0.82	-
Chen et al. (2020)	Productivity and Safety	10,000 (Real images)	1,500 (Real images)	IFaster R-CNN	Excavator	0.99	0.81	-	0.95
					Worker	0.98	0.79	-	0.91
Golkhoo (2020)	Progress Monitoring	560 (Real images)	62 (Real images)	Mask R-CNN	Duct	0.70	0.74	-	0.61

		804 (Synthetic/Real images)	62 (Real images)	Mask R-CNN	Duct	0.73	0.77		0.66
Wei Wei et al. (2022)	Progress Monitoring	500 (Real images)	78 (Real images)	Improved Mask RCNN	Plaster Person Putty	- - -	- - -	- - -	0.93 1.00 0.97
				YOLOACT	Plaster Person Putty	- - -	- - -	- - -	0.94 0.96 0.94
				PointRend	Plaster Person Putty	- - -	- - -	- - -	0.96 0.96 0.94
Hou et al. (2020)	Site Monitoring	4,378 (Synthetic images)	100 (Synthetic images)	DSOD	Column Beam	96.3 95.8	95.5 95.1	- -	- -

Note: P = Precision, R = Recall, F1 = F1 Score, and AP50 = Average Precision at 50% overlap.

2.2.1.4 Image Datasets for Training Deep Learning Models

Although deep learning algorithms are capable of robust performance, they require large image datasets for training. The limited availability of real images from construction projects makes the development of such datasets challenging (Xiao et al., 2021). Large-scale datasets like Microsoft COCO (MS COCO) (Lin et al., 2014), PASCAL Visual Object Classes (VOC) (Everingham et al., 2007), and ImageNet (Deng et al., 2010), lack construction-related objects and are not directly suitable for model training in the construction domain. However, they are helpful for transfer learning to pass basic image features to the models trained on specialized, smaller datasets containing the target objects (Nath & Behzadan, 2020; Oquab et al., 2014). There have been a number of research studies in the past few years that have created open image datasets specific to construction objects and made them available to the public.

Duan et al., (2022) created an image dataset named Site Object Detection Dataset (SODA) from different phases of construction sites. They developed a comprehensive annotated dataset consisting of 15 object classes related to material, labor, equipment, and layout with around 20,000 images. Xuehui et al. (2021) collected 41,668 images referred to Moving Objects in Construction Site (MOCS) with thirteen object classes presented in construction sites including workers and

machines. Xiao and Kang (2021), created a dataset consisting of 10,000 annotated images containing 10 object classes from construction machines referred to as the Alberta Construction Image Data Set (ACID). The dataset consists of images taken from construction sites using unmanned autonomous vehicles (UAVs), installed cameras, manual imaging, and downloading images from online sources. Czerniawski and Leite (2018), created a dataset consisting of 2D image frames and their corresponding depth images for a variety of indoor object classes, including HVAC ducts and plumbing.

Although construction image datasets have been released in recent years, they do not contain all types of construction components with enough images and viewpoints (Xiao & Kang, 2021). Moreover, the high-quality datasets created by construction companies are not publicly available due to confidentiality concerns. Without sufficient images for training the algorithms, a number of problems can arise, such as overfitting, poor performance, and limited generalizability (Kim et al., 2023). In such cases, synthetic images can be utilized in order to create construction-related datasets for training the deep learning models (Barrera-Animas & Davila Delgado, 2023).

There are two types of synthetic images: synthetic composites and virtual synthetic images. Synthetic composite images refer to real images that are digitally modified in a way that incorporates objects that are not initially included in the image. These objects can be either synthetically generated or obtained through splicing from various real images and subsequently added to the original image (Man & Chahl, 2022). A sample of these images is depicted in Figure 2-4. However, superimposing objects onto backgrounds can result in unrealistic images due to inconsistencies in geometry (e.g., mismatched size and orientation), lighting conditions, and semantic information that can negatively affect the model's performance (Niu et al., 2021). Conversely, virtual synthetic images are entirely generated artificially with no real elements (Man & Chahl, 2022) as shown in Figure 2-5.

Typically, 3D BIM models are utilized to generate synthetic images by capturing viewpoints within three-dimensional space containing the target objects (Soltani et al., 2016). Although object recognition algorithms using real, synthetic or mixed datasets have been investigated in many research papers, several challenges remain.

Regarding synthetic image generation, complex indoor objects are not investigated, and the images often fail to represent different real site conditions. Enhancing the diversity of synthetic images is crucial to improve model generalization across various construction site scenarios. In

addition, a more in-depth investigation of different synthetic and real ratios within datasets for training is required. Table 2-2 summarizes the gaps in the literature.



Figure 2-4. Synthetic composite image sample (Hwang et al., 2023).



Figure 2-5. Virtual synthetic image sample.

Table 2-2. Overview of object recognition models for construction monitoring.

Author	Training set	Indoor/ Outdoor	Target Objects	Key Algorithm	ISeg	Computing Platform	Limitations
Zheng et al. (2020)	Virtual and real images	Outdoor	Prefinished modules	Mask R-CNN	√	Physical	<ul style="list-style-type: none"> 1- Synthetic images lack diversity in real scenarios, such as lighting, weather, and occlusion. 2- Recognition is limited to simple objects (boxes). 3- Did not optimize the mix of synthetic and real images.
Wei Wei et al. (2022)	Real images	Indoor	Worker, wall elements	Improved Mask RCNN	√	Physical	<ul style="list-style-type: none"> 1- Cannot apply the preprocessing model to all objects. 2- Test data is limited (78 images). 3- Recognition is limited to simple objects (walls). 4- Carried out for ideal conditions; did not consider indoor scenarios like lighting, occlusion, and clutter.
Hwang et al. (2023)	Composite and Real images	Outdoor	Equipment	Faster R-CNN	-	N.A	<ul style="list-style-type: none"> 1- Unrealistic synthetic images (mismatch equipment with background in size, orientation, color, and semantic information). 2- Synthetic images lacked diversity in 3D-modeled objects and background images with lighting and weather variations.
Ekanayake et al. (2022)	Real images	Indoor	Wall partition elements	YOLOv4	-	Cloud	<ul style="list-style-type: none"> 1- Recognition is limited to simple objects (walls). 2- Improving the dataset quality or approaches to handle challenges in indoor sites are not investigated.
Kim et al. (2023)	Composite and Real images	Outdoor	Workers	YOLOv3	-	N.A	<ul style="list-style-type: none"> 1- Unrealistic synthetic images due to discrepancies in pose, orientation, and scale of the target object (worker) with the background. 2- Did not optimize the mix of synthetic and real images.

							<ul style="list-style-type: none"> 3- Synthetic images lacked diversity in occlusion and weather variation. 4- Concentrating on a class (workers) prevalent in datasets and did not consider less-represented, crucial construction elements.
Li and Chen (2022)	Real images	Outdoor	Pipes	YOLOv3	-	Physical	<ul style="list-style-type: none"> 1- Only considered objects (pipes) in storage sites not installed which is easier to identify as opposed to objects in challenging environments. 2- Images were taken from only limited shooting angles, failing to represent the variety of viewpoints found on construction sites.
(Hou et al., 2020)	Virtual images	Indoor	Structural elements	DSOD	-	Physical	<ul style="list-style-type: none"> 1- Only synthetic data were considered. 2- Lack of diverse visual attributes for target objects (shape, size, details). 3- Synthetic images fail to reflect real indoor conditions such as lighting and clutter conditions.

Note: ISeg = instance segmentation capability.

2.2.2 Application of Location Tracking Technologies in Construction

Tracking resources efficiently in complex construction environments requires identifying and localizing them accurately (Shahi et al., 2012). As construction projects become larger and the number of high-cost assets increases, the localization of resources becomes more serious and important (Cho et al., 2010).

Resource localization is required to be reliable in a consistent way to assist stakeholders in decision-making. For this purpose, RTLS technologies are introduced for tracking resources in job sites. These systems are used to overcome the challenges of traditional manual data collection methods in construction sites which are inaccurate, unreliable, and expensive (Shahi et al., 2012). RTLS is described as the integration of hardware components and software systems to automatically identify the position of an object in a device-installed monitored area in real-time. The data collected during this process can be used in real-time applications or analyzed later (Li et al., 2016).

Materials and equipment costs can account for 50-60% of the total cost of industrial construction projects (Kini, 1999). Hence, it is crucial to organize and control materials in construction projects to save unnecessary expenses (Georgy & Basily, 2008). Materials used in the construction industry are categorized into off-the-shelf, bulk, and engineered components. The materials in each category differ based on expense, procurement lead time, and exchangeability. In general, engineered components with specific properties are more costly and require more lead time and advanced scheduling. Mechanical, Electrical and Plumbing (MEP) components, such as pipe spools, are considered as engineered materials with a significant and expensive procedure (Song et al., 2006).

The importance of RTLS becomes higher in large projects, in which tracking thousands of material components such as MEP components (e.g., pipe spools and valves) and structural components on large area sites is required. In addition, components like pipe spools, due to their specific small shape and weight, are at a higher risk of being lost or not easily found compared to larger materials (Grau et al., 2009). Using RTLS technologies, different construction resources such as materials can be tagged for monitoring purposes, even in obstructed environments or from considerable distances (Teizer et al., 2008a).

RTLS enhances construction progress monitoring by increasing awareness of resource location and status, controlling activities for better planning, and implementing required actions in a timely manner (Teizer et al., 2008a). There are different types of RTLS technologies such as Global Positioning Systems (GPS), Infrared (IR), Radio Frequency Identification (RFID), UWB, and Bluetooth (Huang, Hammad, & Zhu, 2021). An experimental study conducted by Grau et al. (2009), compared traditional and automated materials tracking, and the benefits of an automated system were elaborated. For the automated system, they utilized 400 RTLS tags to track materials over a three-month period, demonstrating remarkable time and cost savings. The system could save 88% of the time spent per component on site, reducing it from 36.8 minutes to 4.56 minutes. Additionally, it showed an 8.98% enhancement in mitigating information loss about missing materials compared to traditional tracking systems.

The GPS system, a satellite based RTLS technology is used in construction projects to provide real-time object location information. The GPS consists of satellites placed above the Earth by the U.S. Department of Defense in the 1970s. The satellites emit radio signals to receivers attached to objects to identify their location. This system works based on the concept of trilateration, which needs four satellites to accurately determine an object's position through the geometric intersection of four spheres. The application of GPS in the construction industry is for safety management, activity tracking, site acquisition, and surveying (Kumar & Moore, 2002; Li et al., 2005; Moselhi et al., 2020; Omar & Nehdi, 2016). Due to the requirement of direct line-of-sight (LOS), GPS is most often applied in activities that take place outdoors such as earthwork operations. For this reason, other technologies are being considered for localizing objects in indoor construction environments such as RFID, Bluetooth, UWB, and IR (Moselhi et al., 2020; Teizer et al., 2020; Xu et al., 2018; Zhang et al., 2012, 2020). These systems are differentiated by their localization accuracy, cost, scalability, level of consistency, robustness, and data protection. Therefore, from available RTLS systems, one can be selected that meets the specific requirements of each project, such as being economical or providing high accuracy (Alarifi et al., 2016).

Recently, indoor localization has received considerable attention for identifying the location of workers, equipment, and materials continuously and in real time. However, localization in indoor job sites is more challenging and requires higher accuracy (Alarifi et al., 2016). This is due to the presence of different objects including equipment, structures, walls, and humans that densely

exist and cause signal reflections and scattering, degrading the performance of RTLS. Hence, the presence of different objects within the indoor environment can prevent signals from transmitting in a direct path resulting in non-line-of-sight (NLOS) conditions (Alarifi et al., 2016; Mautz, 2012). The other challenge in indoor localization is signal interference by other sources which impacts the RTLS and degrades the signals' stability. These sources include mobile devices, wireless networks, microwave systems and fluorescent lighting (Alarifi et al., 2016).

Different research studies introduced performance metrics to evaluate the RTLS in indoor environments. Wu et al. (2007) introduced six attributes of localization systems in indoor environments that can be used to evaluate their performance, which are: (1) accuracy and precision, (2) system coverage, (3) latency of location updates, (4) the building's infrastructure impact, (5) system calibration, and (6) random errors' impacts derived from multipath effects. Gu et al. (2009) introduced other attributes to evaluate these systems including (1) system cost, (2) data security, (3) complexity, and (4) robustness. They highlighted the importance of system cost, which is not limited to the cost of hardware components but also the costs of installation and maintenance. Other parameters involved in determining the cost of RTLS include time and space costs. Time costs relate to the time needed for the installation and calibration of the system and the time required for corrections when problems in the system occur. Space costs refer to the size of the components and their physical space requirements. Alarifi et al. (2016) emphasized the importance of systems' scalability and availability, in addition to the previously mentioned metrics. Figure 2-6 demonstrates the main performance metrics required to assess the performance of RTLS for different applications.

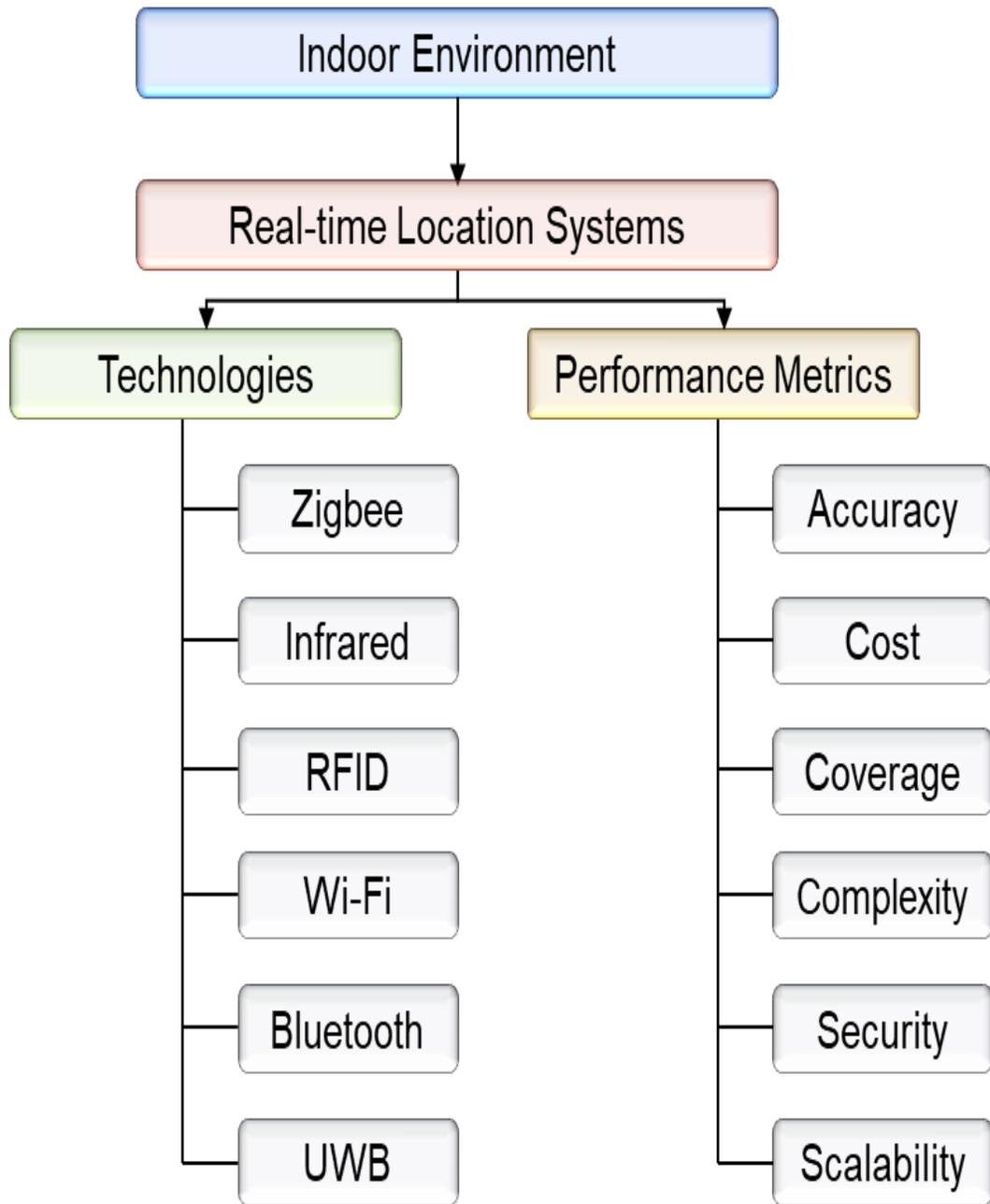


Figure 2-6. Indoor localization technologies and their performance metrics.

Table 2-3. Different applications of location tracking technologies in construction

Reference	Objectives	Tracked object	Method
Montaser et al. (2012)	Productivity Estimation	Equipment	GPS
Li et al. (2013)	Safety Management	Workers & Equipment	GPS / RFID
Song et al. (2015)	Material Management	Materials	GPS/Barcode
Teizer et al. (2015)	Safety Management	Workers & Equipment	GPS
Alshibani and Moselhi (2016)	Productivity Estimation	Equipment	GPS
Park et al. (2017a)	Safety Management	Workers	Bluetooth
Huang et al. (2021b)	Safety Management	Workers	Bluetooth
Zhao et al. (2019)	Production Control	Workers	Bluetooth
Fang et al. (2016)	Safety Management	Workers	RFID
Montaser and Moselhi (2014)	Material Management	Workers and Materials	RFID
Montaser and Moselhi (2012a)	Productivity Estimation	Equipment	RFID
Montaser and Moselhi (2012b)	Progress Monitoring	Materials	RFID
Chin et al. (2005)	Progress Management	Materials	RFID

One of the RTLS technologies appropriate for indoor object localization is Bluetooth Low Energy (BLE), which was introduced in 2010 and uses 2.4 GHz radio frequency (Huang, Hammad, & Zhu, 2021; Park et al., 2017). Due to its advantages such as being economical, having low energy use, internet connectivity, scalability, and the ability for signal penetration into walls, the BLE system is suitable for different construction applications. This system contains tags that can communicate with the reader. Mobile devices such as smartphones, and tablets that contain BLE can work as the reader for the tags (Topak et al., 2018).

There are two main methods for distance measurement using BLE including the Received Signal Strength Indicator (RSSI) and the fingerprint technique (Sergi et al., 2022). Different research studies utilized BLE technology to track the location of workers and equipment in indoor and outdoor job sites. For instance, Mohanty et al. (2020) implemented BLE technology in the field environment to track the location of workers for productivity monitoring. In another study done by Park et al. (2017b), BLE tags were attached to workers and equipment to identify hazardous situations and increase safety in construction sites. Khazen et al. (2023) conducted laboratory experiments to assess the developed proximity detection framework using BLE for improving safety and productivity monitoring in indoor environments. Furthermore, there are other research studies that used BLE for different construction applications which can be found in Table 2-3. However, since the accuracy of the BLE system is between 2-5 meters (Schjørring et al., 2022), and has a coverage limitation of up to 30 meters (Rao et al., 2022), it may not be accurate enough for all the applications in the construction projects and other technologies are required to be investigated.

RFID is another RTLS technology that is widely used for object tracking and localization in construction sites. In the RFID system, data transmission is achieved via radio frequency signals transmitted between RFID tags and readers (Montaser & Moselhi, 2014). While it is not as accurate and easy to implement as other technologies, it has received considerable attention (Li et al., 2016). In large-scale projects, RFID is recognized as the most common technology for localization objectives (Alarifi et al., 2016). RFID uses radio frequencies, ranging from low to super-high frequencies for automated data collection. RFID tags can be utilized to track and localize materials, workers, and equipment in construction projects (Li et al., 2016).

Each RFID tags consists of a microchip and an internal antenna with a predefined ID. The tags are scanned by a mobile reader handled by an inspector or by a fixed reader to collect and store the data (Lu et al., 2011; Moselhi et al., 2020; Omar & Nehdi, 2016; Song et al., 2006). Subsequently, the reader passes the captured data to the core computer for further analysis and specific application use. There are various types of RFID that have different properties such as power supply, signal frequency, reading range, data rate, service life, price, memory space, size, and shape (Song et al., 2006). According to the type of material and the conditions of the work

environment, appropriate RFID type and configuration can be selected (Valero et al., 2015). A sample of RFID tags and a mobile reader is depicted in Figure 2-7.

The RFID system is categorized into active and passive tags (Teizer et al., 2008a). Active tags utilize batteries and are capable of high reading range, data rates and memory storage. However, passive tags that do not require batteries are more economical and can be smaller in size (Song et al., 2006; Teizer et al., 2008). RFID systems can be used to track objects in construction sites without requiring line-of-sight (Song et al., 2006). Many studies have investigated the benefits and limitations of RFID, its use cases, future trends, and potential integration with other technologies (Sardroud, 2014; Sun et al., 2013; Valero et al., 2015; Wing, 2006)

RFID is used in many construction applications such as material management, safety management, and progress monitoring. Ghanem and Abdelrazig (2006), utilized RFID technology to track the progress of work in construction sites. Montaser and Moselhi (2012) used economical RFID tags, attached to the hauling units, and fixed RFID readers at specified gates to capture near real-time data for tracking earthmoving operations. Song et al. (2006) utilized active RFID tags to identify and track pipe spools using mobile and fixed readers. This study demonstrated that RFID is practical in complex environments and that the tags could work in congested areas filled with metallic objects. In addition, they determined the benefits of using RFID for tracking pipe spools during shipment, reception, and storage. These benefits include reduced identification time and errors, minimized search time, and fewer reproductions of misplaced items.

Razavi and Moselhi (2012) investigated the application of RFID in indoor job sites. In this study, economical passive RFID tags and a reader were used for localization purposes. Motamedi et al. (2013) studied the utilization of active RFID for localizing static and movable assets during the operational phase in indoor environments. Other studies that used RFID for construction applications can be found in Table 2-3. The RFID system has limitations, such as not being fully automated and not being accurate enough for 3D localization. Therefore, they are primarily used for 2D localization (Awolusi et al., 2018; Moselhi et al., 2020; Omar & Nehdi, 2016; Rao et al., 2022; Shahi et al., 2012; Zhang et al., 2020).

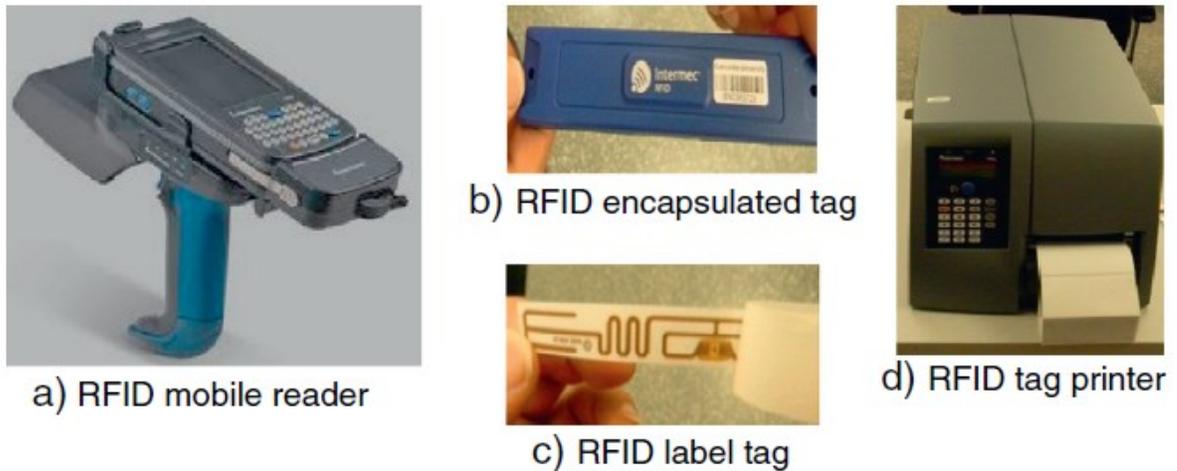


Figure 2-7. RFID hardware components including the reader and different types of tags and RFID tag printer (Montaser & Moselhi, 2014).

2.2.2.1 The UWB System

Between 1960 and 1990, the UWB system has been confined to military applications and the US Department of Defense (Kshetrimayum, 2009). The UWB was subsequently made available for commercial sale to the public (Alarifi et al., 2016). The UWB system is a promising technology for providing real-time, secure, and accurate distance and location estimations. The hardware components of the UWB system include tags and receivers. The receivers are static and fixed in predefined locations, while the tags can be installed in moving or static objects (Liu et al., 2022). Depending on the type of receivers used, the UWB system can have a range between 200 meters to 1000 meters, allowing users to utilize fewer receivers while covering a large space (Liu et al., 2022; Teizer et al., 2008). The UWB system has the capability to provide accurate 3D localization that can be beneficial for 3D material tracking on construction sites to improve decision-making (Teizer et al., 2008).

Moreover, the UWB system offers a longer range and better stability in different environmental conditions than RFID. The UWB tags and receivers communicate consistently over a bandwidth that exceeds 500 MHz (Awolusi et al., 2018; Bardareh & Moselhi, 2022; Moselhi et al., 2020; Omar & Nehdi, 2016; Rao et al., 2022; Shahi et al., 2012; Zhang et al., 2020). The broad bandwidth enables a high rate of data transfer, and the low frequency of UWB pulses allows signals

to penetrate objects such as walls (Alarifi et al., 2016). The other advantages of the UWB system are its anti-interference capability and high multipath resolution (Zhang et al., 2020).

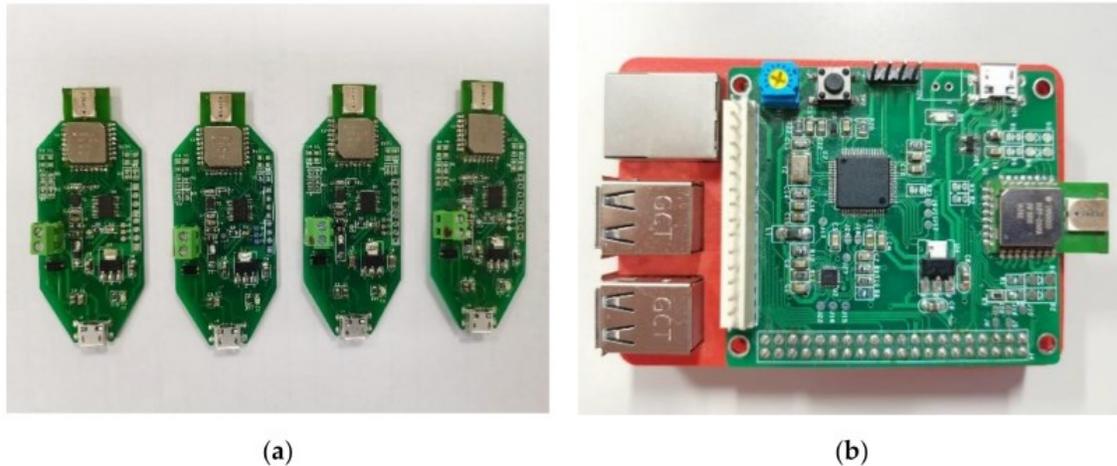


Figure 2-8. Sample of (a) UWB receivers and (b) tag from the DecaWave Company (DWM1000) (Yao et al., 2021).

UWB tags can be activated by operators and then attached to tracked objects for localization and tracking so that the system can collect data from them (Shahi et al., 2013a). In the UWB system, the tags send radio signals to the receivers and the location of tags can be determined by the utilization of various techniques such as Time of Arrival (TOA)/Time of Flight (ToF), Time Difference of Arrival (TDOA), Angle of Arrival (AOA) (Liu et al., 2022; Shahi et al., 2013; Zhang et al., 2020). Each of these techniques can be used separately or in combination for localization of the tags (Zhang et al., 2020).

Range measurements between tags and receivers can be based on ToF estimation. Several parameters that exist on job sites can reduce the accuracy of ToF estimation, leading to errors in range measurements and tag localizations. These parameters include non-line-of-sight conditions, multipath effects, synchronization difficulties and signal interference (Liu et al., 2022). In indoor environments, the impact of these parameters is higher on range measurement estimations, leading to outliers that may exceed 1 meter (Ruiz & Granja, 2017).

The concepts of line-of-sight, non-line-of-sight and multipath effects become important in improving the performance of these systems since they have direct impact on range measurement

accuracies (Sang et al., 2020). A line-of-sight condition is when a signal propagates through a direct path between the tag and receiver without any present obstacles (Dardari et al., 2009). A non-line-of-sight condition occurs when a clear and direct path between the tag and the receiver does not happen due to obstacles. In this condition, the signals pass through longer distance between the tag and receiver compared to the direct line-of-sight path (Kristensen et al., 2019). Through the multipath effect, radio frequency signals propagate from tags to receivers through various paths, with time delays caused by obstacles in the environment (Mautz, 2012; Sabri et al., 2012). Moreover, these multipath propagations can also be caused by obstacles that are not necessarily between the tag and the receiver (Sang et al., 2020). Figure 2-9 demonstrates the sample of the line-of-sight, non-line-of-sight conditions and multipath effects.

In indoor job sites, UWB localization accuracy can be reduced from centimeter-level in line-of-sight conditions to meter-level in non-line-of-sight conditions (Liu et al., 2022). Various obstacles can cause signal reflections, diffractions, and scattering within indoor environments. These include structures, walls, workers, and components (Sabri et al., 2012; Sang et al., 2020). In addition, human presence, and communication devices such as cordless phones can cause signal interference that can negatively affect the UWB system (Cho et al., 2010). It is also possible that signal interference occurs between different UWB systems (Jiménez & Seco, 2016).

The most recognized manufacturers that provide the UWB systems for commercial purposes are Ubisense, BeSpoon, Decawave, Time Domain, Sewio and Pozyx. Different research studies have utilized Ubisense (Cho et al., 2010; Shahi et al., 2013b; Siddiqui et al., 2019; Umer & Siddiqui, 2020; Xia et al., 2010) or Decawave (Bardareh & Moselhi, 2022; Jin et al., 2019a; Yao et al., 2021) as part of their experiments, specifically for localization and tracking purposes. Depending on their application, UWB tags may have compact, cubic, or microrectangular shapes, and their weight may be less than 12 grams (Teizer et al., 2008a). According to the manufacturer's information (BeSpoon; Decawave; Ubisense) and the details provided in (Jiménez & Seco, 2016; Ruiz & Granja, 2017; Wang et al., 2015), the configurations of some of these products are provided in Table 2-4.

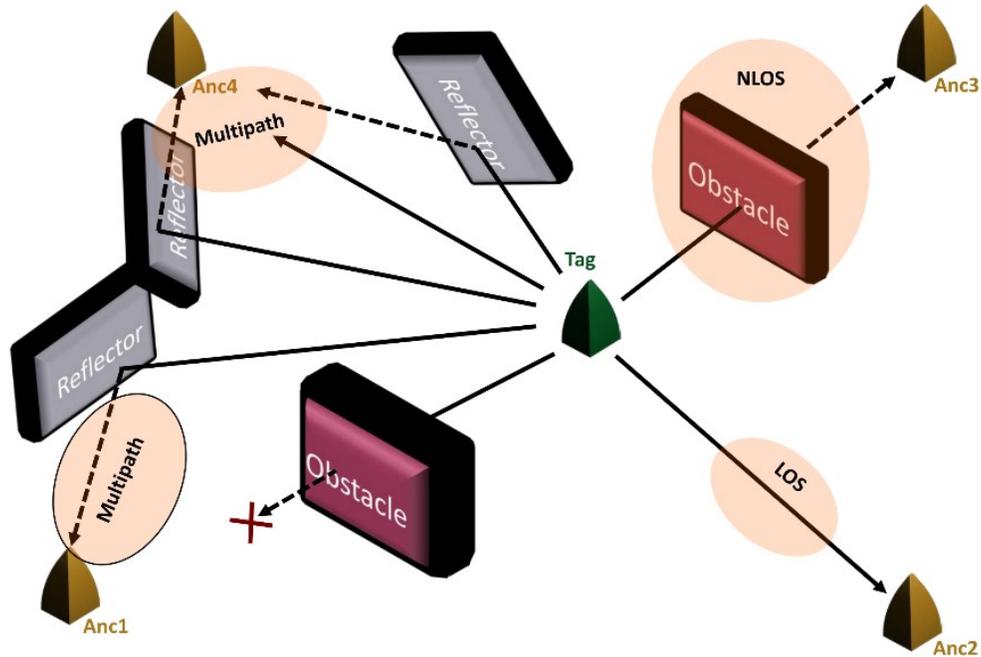


Figure 2-9. Line-of-sight and non-line-of-sight conditions, and multipath effects (Sang et al., 2020).

Table 2-4. Comparison between different UWB manufacturers' system configurations (Jiménez & Seco, 2016; Ruiz & Granja, 2017; Wang et al., 2015).

System	Origin	Founded Date	Channel (GHz)	Range (m)	Technique	Range Update (Hz)	Price
Ubisense 7000	United Kingdom	2002	6–8	>160	AOA - TDOA	N. A	26,900 Euros -6 receivers -10 tags
BeSpoon	France	2010	3.99	≤880	TOA	2.5	1,699 euros -6 tags -hardware system
DecaWave DW1000	Ireland	2007	3.99 & 6.489	≤300	TWR-TOF	3.5	925 U.S. Dollars

In some research studies, these products have been compared to evaluate their performance under the same environment and conditions to provide useful information for the users. Jiménez and Seco (2016) compared the localization and range measurement performance of BeSpoon and Decawave products under both line-of-sight and non-line-of-sight conditions. To conduct an experiment under line-of-sight condition, an outdoor environment without any obstacles was chosen, and for non-line-of-sight conditions, another experiment was conducted in an indoor laboratory with presence of walls, furniture, and humans as obstacles. Their findings indicated that the ranging errors in both line-of-sight and non-line-of-sight conditions for BeSpoon are larger than those for Decawave. Moreover, the performance of Decawave in positioning is better than BeSpoon. They found that operating both systems simultaneously would have a negative effect on each system due to signal interference, causing increased noise in range measurements. Moreover, in the Decawave system, communication between some nodes was disrupted.

They also expanded their work in Ruiz and Granja (2017), comparing the performance of three UWB system products including DecaWave, Ubisense and BeSpoon, under the same conditions in an industry-like environment. They demonstrated that, in a real job site where non-line-of-sight conditions exist, Decawave had superior performance to BeSpoon and Ubisense in terms of accuracy and outliers. Moreover, Ubisense had the poorest performance among the three. Schjørring et al. (2022) summarized the findings of recent research papers which applied UWB commercial products across different applications. Pertinent details regarding each conducted experiment and its outcomes were provided, encompassing the experimental area, the number of receivers used, the heights of the tags, and 2D/3D localization accuracy.

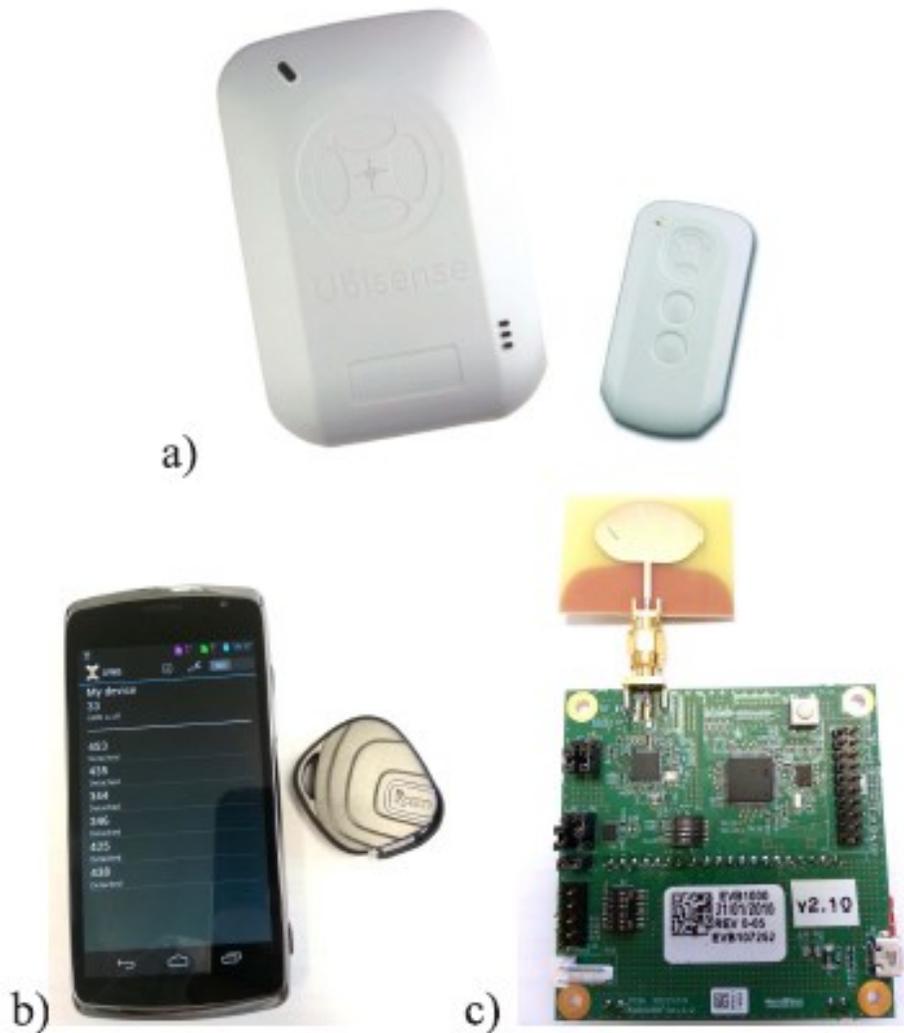


Figure 2-10. Samples of different UWB commercial products: a) UbiSense, b) BeSpoon and c) DecaWave (Ruiz & Granja, 2017).

Shahi et al. (2012) implemented different experiments to evaluate the performance of UWB technology based on occlusions, location dependency and time duration in indoor construction sites for material tracking and progress reporting. Siddiqui et al. (2019) implemented the UWB system in outdoor construction sites, highlighted different factors affecting the system and concluded with practical guidelines for its implementation on sites. In this study factors affecting the UWB system are categorized based on different aspects including the type of UWB system (e.g., wired or wireless), tag type and its setting (e.g., shape, update rate, number of tags), UWB system setting (e.g., number of receivers, calibration and measurement quality) and environmental

conditions (e.g., presence of electromagnetic devices and radio frequency noises, materials of objects existed within the environment and line-of-sight conditions). Following the experiments, they emphasized the importance of analyzing the UWB system during the planning phase to ensure high-quality installation at construction sites. The importance of collaboration with team entities and site managers to achieve optimal performance from the system was highlighted. Further, they indicated that increasing the number of receivers could increase the possibility of line-of-sight conditions between tags and receivers, improving localization accuracy. Finally, they recommended integrating additional data collection sources such as camera-based and computer vision techniques to enhance object localization performance.

Maalek and Sadeghpour (2013) conducted a series of experiments to determine the impact of several factors on the accuracy of the UWB system in an indoor environment with static tracked elements. These factors are (1) clear line-of-sight condition, (2) presence of metallic elements in the testing area, (3) signal blockage by elements, (4) metallic tracked elements, (5) the utilization of timing cables and AOA measurements, (6) the number of installed tags within the testing area, (7) the number of active UWB receivers. Another study by Umer and Siddiqui (2020) was conducted to assess the 2D and 3D localization performance of the UWB system in an outdoor environment. Several experiments were carried out, focusing on: (1) the number of tags in the testing area, (2) the locations of receivers (3) static and dynamic tags (4) the utilization of TDOA-AOA and AOA measurements only. They concluded that activating all tags in the experimental area degraded UWB localization performance. Moreover, using only AOA measurements for localization decreased the system's accuracy. They observed that sensor placements impact the system's accuracy. To reach optimal performance they suggested the following items: (1) installing the receivers outside the experimental area boundary, (2) adjusting the placement of receivers according to changes in the condition of the construction site during its lifecycle, (3) placing the receivers at the highest possible locations and (4) attaching the tags on the top of tracked resources to maintain direct line-of-sight with receivers.

The exploration of optimized UWB receiver installation is one of the key research directions (Liu et al., 2022). Yao et al. (2021) investigated the effect of receiver positions on tag localization accuracy and provided insights into the underlying reasons for this effect. Since the estimation of a tag's location is based on the measured distances between the tag and its receivers, the positions

of installed receivers significantly affect localization accuracy. Different experiments were conducted to assess the impact of receiver placements on 2D and 3D localization errors under both line-of-sight and non-line-of-sight conditions.

For real-time tracking of moving resources in harsh construction sites, Cheng et al. (2011) carried out several experiments on three outdoor sites. In these experiments, UWB tags with different frequencies between 1 Hz and 60 Hz were attached to the labor, materials, and equipment. It was recorded that the average error of the tags was between 0.36 meters and 1.82 meters depending on the type of construction site and the frequency of the tag. Cho et al. (2010) evaluated the performance of an untethered UWB system for tracking assets in indoor sites, in both static and dynamic modes of tags. They conducted different experiments to evaluate the accuracy of static tags at different heights in various indoor environments, such as wood and steel-framed buildings. In their experiments, the localization accuracy of tags improved at elevated positions due to better line-of-sight with the receivers, compared to when the tags were placed at ground level. In addition, an experiment was conducted in the furnished office lab to examine whether human traffic and devices could interfere with UWB signals.

Shahi et al. (2013b) implemented the UWB system for progress tracking of activities in indoor job sites. In an area filled with MEP components, they tracked pipeline-related activities in 3D by attaching UWB tags to pipelines. Many research studies have investigated the application of the UWB system in construction sites as depicted in Table 2-5. However, many of these studies have applied the UWB system in laboratory tests and outdoor field areas, which may not have the actual challenges in complex indoor environments where many occlusions and objects are present in limited spaces. Consequently, the application of the UWB system in real indoor environments requires further consideration to assess its performance in such scenarios and areas. Other RTLS technologies can be integrated with the UWB system for object localization to save cost and time in construction projects (Xia et al., 2010), however, it may affect the localization accuracy (Bardareh & Moselhi, 2022).

The integration of RTLS with vision-based technologies enhances monitoring systems' efficiency in construction sites by reducing data loss and inaccuracies that are derived from using a single technology for data acquisition (Ekanayake et al., 2021; Soltani et al., 2018). Soltani et al. (2018) introduced a data fusion framework that utilizes GPS, calibrated cameras, and computer

vision techniques to compute the 3D pose of an excavator in a construction field for improving productivity and safety. Cai and Cai (2020) presented a hybrid method that uses calibrated cameras and Faster R-CNN along with BLE technology to accurately detect and track workers in 3D for safety monitoring. Rafiee et al. (2013) proposed a data fusion model using a BIM model, a single fixed surveillance camera and a UWB system to detect and localize persons for enhancing security in indoor sites. In this model, the KNN algorithm was used to timely identify intruders and their location among verified individuals. Shahi et al. (2015) implemented a 3D object recognition model and a UWB system to facilitate the identification of construction activities' progress including piping by providing comprehensive information from vision-based and positioning technologies.

Table 2-5. Summary of the research studies that utilized the UWB system in construction projects.

Study	Method	Indoor/ Outdoor	Tagged object	Tag Status		Test area type		Accuracy (m)		UWB Product	UWB Technique	Research Limitations
				D*	S*	Lab	Field	2D	3D			
(Zhang et al., 2020)	UWB	Indoor and Outdoor	Test points, Vehicles, and Workers	√	√	√	√	-	0.17- 0.45	N.A	N.A	1. Indoor experiments were conducted in lab settings, not in congested and occluded environments. 2. Tags were not attached to challenging materials like metal in indoor environments. 3. 2D Accuracy was not reported. 4. Significant data loss occurred.
(Umer & Siddiqui, 2020)	UWB	Outdoor	Test points	√	√	√	-	0.18- 1.99	0.32- 2.65	Ubisense	TDOA and AOA	1. Only lab experiments under ideal conditions were performed. 2. Localization in indoor environment was not considered. 3. Tags were not attached to actual objects existed in job sites. 4. The wired UWB system is not practical in congested indoor environments.
(Shamsollahi et al., 2023)	UWB	Indoor	Test points	-	√	√	-	-	0.05- 0.13	Decawave (Qorvo)	TOF	1. Experiments were conducted in lab settings under ideal LOS conditions. 2. 2D accuracy was not reported. 3. Tags were not attached to challenging objects.

(Siddiqui et al., 2019)	UWB	Outdoor	Equipment	√	√	-	√	0.13-1.37	-	Ubisense	AOA	<p>4. A limited number of experiments were conducted under restricted conditions.</p> <p>1. The UWB system for 3D localization was not considered.</p> <p>2. It was applied in an outdoor environment with fewer obstructions than indoor.</p> <p>3. Challenging conditions where obstructions between tags and receivers occur were not addressed.</p> <p>4. Calibration and sensor placement issues occurred.</p> <p>5. Tags were attached to only one component, not across different areas of the site.</p>
(Bardareh & Moselhi, 2022)	UWB-RFID	Indoor	Test points	√	√	√	-	0.52	1.15	Decawave (Qorvo)	TOF	<p>1. Experiments conducted only in lab settings under ideal conditions.</p> <p>2. Tags were not attached to challenging objects.</p> <p>3. Limited experiments under restricted conditions were performed.</p> <p>4. System implementation is complex.</p> <p>5. Tag placements at various heights were not explored.</p> <p>6. RFID tags are susceptible to multipath effects.</p>
(Xia et al., 2010)	UWB-GPS	Indoor	Test points and Human	√	√	√	√	-	≤1	Ubisense	TDOA and AOA	<p>1. Tags were not attached to actual objects.</p> <p>2. The wired UWB system is not practical in congested indoor environments.</p> <p>3. Indoor field environments were not obstructed.</p> <p>4. Experienced calibration difficulty.</p> <p>5. Significant data loss occurred.</p>
(Jin et al., 2019)	UWB	Indoor	Test points	√	√	√	-	0.10	0.15-0.20	Decawave (Qorvo)	TOF	<p>1. Only indoor lab experiments were conducted, and the system was not applied under real conditions.</p> <p>2. Tags at different heights or positions were not considered.</p> <p>3. All static experiments were conducted under LOS conditions or limited NLOS conditions with few obstacles.</p> <p>4. Tags were not attached to actual objects.</p>

(Shahi et al., 2013)	UWB	Indoor	Pipes and Ducts	-	√	-	√	0.07-0.10	0.10-0.15	Ubisense	TDOA and AOA	<ol style="list-style-type: none"> 1. The wired UWB system is not practical in congested indoor environments. 2. Error analysis for tags at various locations was not conducted. 3. Errors associated with tags at different construction stages were reported. 4. Effect of NLOS conditions and multipath on tags were not provided.
(Zhang et al., 2012)	UWB	Outdoor	Crane	√	√	√	-	-	0.25-0.30	Ubisense	TDOA and AOA	<ol style="list-style-type: none"> 1. The wired UWB system is not practical in congested indoor environments. 2. The indoor environment was not investigated. 3. The outdoor site was not a real construction field and was not occluded.

Note: D = Dynamic, S = Static.

To select the appropriate RTLS type for each project based on specific application, construction site characteristics and type of tracked resources, it is crucial to understand each system's capabilities and limitations which are detailed in Table 2-6.

Table 2-6. Capabilities and limitations of RTLS technologies.

Technology	Capabilities	Limitations	References
RFID	<ol style="list-style-type: none"> (1) No line-of-sight requirement (2) Economical (3) Easy tag installation (4) Supports a high number of tags 	<ol style="list-style-type: none"> (1) Additional tag maintenance cost (2) Low accuracy in 3D localization (3) Difficulties in calibration (4) Limited range (5) Weak anti-interference capability (6) Not fully automated 	(Moselhi et al., 2020; Omar & Nehdi, 2016; Song et al., 2006; Wang et al., 2021; Yao et al., 2021; Zhang et al., 2020)
UWB	<ol style="list-style-type: none"> (1) Longer reading range than other technologies. (2) Applicable in both indoor and outdoor construction 	<ol style="list-style-type: none"> (1) High cost (2) Calibration difficulties (3) Tagging difficulties (4) Limited range in non-line-of-sight conditions. 	(Chong et al., 2023; Moselhi et al., 2020; Omar & Nehdi, 2016; Zhang et al., 2020)

	(3) High positioning accuracy in 2D and 3D planes (4) Resistance to Multipath effects (5) Anti-interference capability	(5) Necessity for battery replacement (6) Occurrence of missing data	
Bluetooth	(1) Economical. (2) Low-energy consumption (3) Easy to link with other technologies and devices	(1) Signals susceptibility to obstacles (2) Limited coverage range (3) Weak anti-interference capability (4) Low accuracy (5) Signal strength fluctuations	(Moselhi et al., 2020; Yao et al., 2021; Zhang et al., 2020)
GPS	(1) Globally accessible (2) Precise positioning (3) Simple installation process (4) Cost-effective (5) Minimal computational effort for data analysis	(1) Requires a clear line-of-sight (2) Limited to outdoor locations (3) Not economical for large-scale projects (4) Multipath errors caused by obstacles (5) Signal loss occurrences (6) High maintenance and installation cost	(Moselhi et al., 2020; Omar & Nehdi, 2016; Pradhananga & Teizer, 2013; Teizer et al., 2008b)

2.2.3 Application of BIM in Monitoring Construction Sites

Over the last few years, the use of BIM in different phases of construction has considerably increased due to its capability to save cost and time in projects (J. Chen et al., 2014; Fang et al., 2016). BIM enables stakeholders to develop 3D models and facilitates the documentation, analysis and management of construction-related information. Furthermore, it improves communication and collaboration between team members (Alizadeh Salehi & Yitmen, 2018; Deng et al., 2020; Oh et al., 2015; Wang & Love, 2012). Hence, using BIM to monitor construction sites has attracted many researchers in this domain. However, for construction projects, more complex systems that integrate BIM with other technologies are necessary to provide comprehensive information about the project's status (Boje et al., 2020).

4D simulation is one of the most commonly used BIM methods for progress monitoring, which allows project managers to visualize and compare as-planned and as-built information

through semantically enriched 3D models that are linked with project schedules (Alizadehsalehi & Yitmen, 2016; Braun et al., 2015; Campagna-Wilson & Botton, 2020). Many research studies integrated computer vision algorithms and 4D BIM for automated construction progress monitoring (Han & Golparvar-Fard, 2015; Kropp et al., 2013a, 2018; Tuttas et al., 2017). For example, Kropp et al. (2013b) utilized 4D BIM to find information related to the objects associated with specified activities as well as motion information to do a simple 2D classification. To evaluate the model, computer vision tasks including HOG features and SVM classifiers were applied to recognize heating devices in an indoor construction site from image frames.

Integrating BIM with RTLS can improve the efficiency of monitoring resources in construction sites and facilitate the decision-making process by providing visual information about tracked resources in a timely manner (Fang et al., 2016). Chin et al. (2005) integrated RFID technology with a 4D BIM model to determine the progress of building components such as structural elements, curtain walls, and cast-in-place concrete. Huang et al. (2021), utilized computer vision techniques for activity recognition and used RTLS to collect location data of workers and equipment to obtain information needed about performing activities. Subsequently, they developed detailed 4D simulations that could improve productivity estimation, safety, and progress monitoring in construction projects.

Cloud-based BIM technology is another BIM development that provides opportunities for users to have access to project progress information in real-time. It is also a cost-effective collaboration tool that enables project entities to share and exchange necessary information through devices such as tablets and smartphones in different locations. This allows decision-makers to track the progress, organize schedules and apply early corrective actions (Afsari et al., 2016; Matthews et al., 2015; Wong et al., 2014).

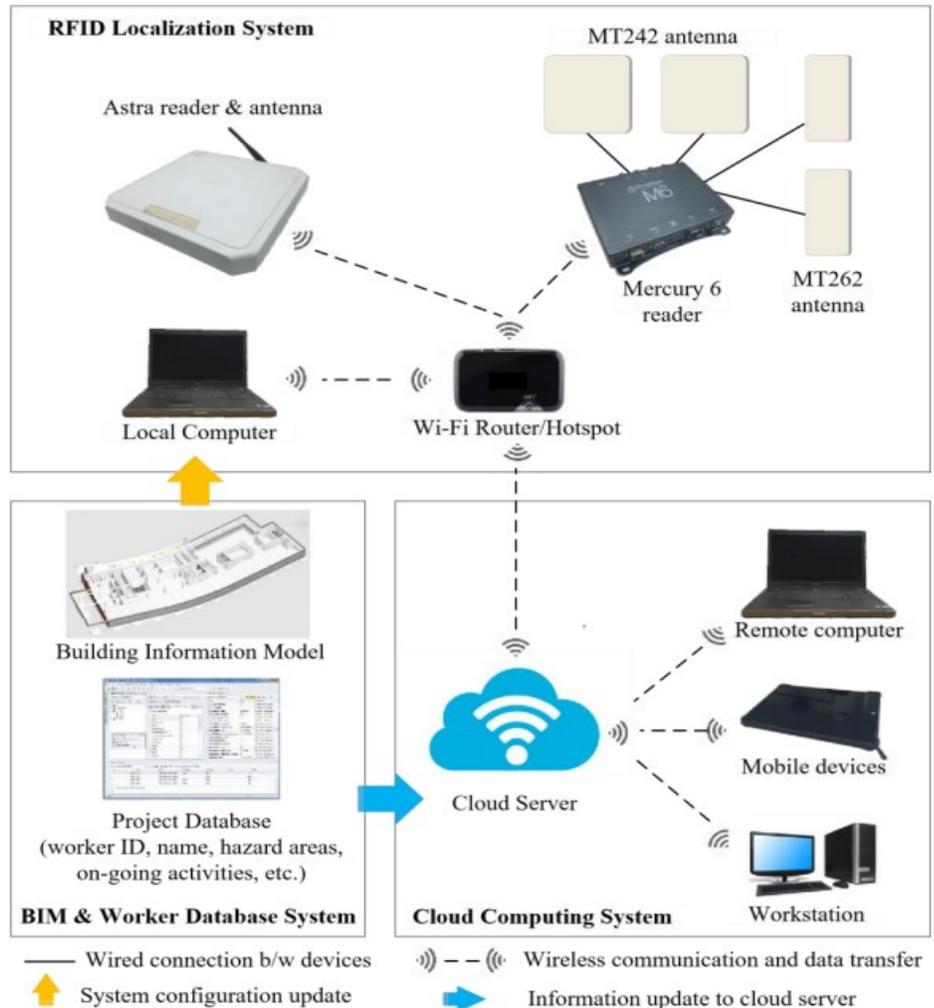


Figure 2-11. Sample of integration of BIM, RTLS and cloud-based system (Fang et al., 2016).

Deng et al. (2020) developed a method using computer vision and BIM to automatically measure and visualize the progress status of tiles. Computer vision techniques including LBPs and SVM classifier were used to identify tiles and the improved edge detection algorithm was applied to extract boundaries of the installed tiles from images. By using camera calibration and BIM model information the real tile area was calculated and the results were transferred to the BIM cloud platform for progress visualization.

Moreover, other integrated systems using BIM and computer vision techniques have the potential to support progress monitoring systems. Ying and Lee (2019) developed an automated framework that creates as-is BIM elements using images taken from construction sites. Mask R-CNN, a deep learning-based object recognition algorithm were applied to detect and segment

walls, doors, and lifts from images. After the segmentation task, the mask boundaries of detected objects were extracted to generate surface geometries and construct IFC building objects.

As mentioned in this section many studies used BIM models with different technologies to facilitate monitoring in construction sites. However, there is still a lack of automated link between sensor data and BIM models. The application of BIM in the construction industry can be increased significantly if it is linked with other devices to provide access to the latest information on the project automatically (Tomasi et al., 2015). Chen et al. (2014) highlighted the importance of “Dynamic BIM” which stores real-time accurate data derived from sensors to represent the project’s current state for applying necessary actions at the right time. They developed a framework to link data from temperature sensors to the BIM model for enhancing the facility management system. Natephra and Motamedi (2019) developed a framework using BIM, sensors, and virtual reality to visualize and analyze indoor thermal conditions such as humidity, temperature, and light intensity in real-time. Teizer et al. (2017) utilized BLE sensors and a cloud-based platform to enrich the BIM model with real-time data for monitoring the project’s progress.

2.3 Summary and Conclusions

In this chapter, recent studies on automated monitoring and reporting at construction sites using new technologies were investigated. The applications of computer vision techniques, RTLS, BIM and integrated systems were reviewed. In summary, although earlier works have demonstrated the potential of new technologies and methods for automated progress monitoring, significant knowledge gaps still exist in the following three domains:

- Digital imaging and object recognition algorithms:

Previous studies have mainly focused on outdoor site elements, resulting in rich datasets for automated progress monitoring systems. However, in outdoor environments factors such as visibility, occlusion, and lighting conditions are generally better compared to those in indoor environments. Poor lighting conditions and high levels of occlusion and clutter degree in indoor environments have negative effects on object recognition performance. Hence, the developed models for outdoor construction sites may not be adaptable to complex indoor environments.

Object recognition in indoor sites is often limited to basic shapes and neglects complex elements and specific challenges of these environments. Additionally, the unavailability of enough real images from indoor elements for training deep learning models intensifies this problem. Consequently, this necessitates an in-depth investigation into complicated and detailed components existing in indoor environments. Although efforts have been made to create synthetic images to enhance the quality of construction datasets, there is a notable lack in generating images from complex indoor components such as MEP elements. In addition, real scenarios typical in indoor environments were not adequately represented in synthetic images, making them less realistic, which limits the effectiveness of training deep learning models. Most studies used high-performance physical computing systems for model training and testing which are expensive and not accessible for everyone.

Finally, although deep learning algorithms are capable of accurate object recognition, they fail to provide precise geolocations of these objects within construction sites. It is essential to identify the exact location of tracked components in order to accurately assess their installation progress.

- RTLS technologies:

RTLS provides precise location information and a unique ID for each element but lacks the visual data needed to validate the proper installation of tracked objects in jobsites and reduce uncertainty in construction operations. Also, tags can be damaged or detached from tracked elements, and data loss is possible. This demands careful RTLS selection based on specific application requirements, system design and the incorporation of complementary data acquisition techniques to improve system robustness. Many research studies using RTLS have been conducted in laboratory or outdoor areas, which may not represent the challenges of complex indoor environments. In indoor sites, numerous objects in limited spaces can cause signal interferences for RTLS. Therefore, the performance of RTLS in real indoor environments requires further exploration. Additionally, the ability of RTLS for accurate 3D localization of challenging objects, such as metallic items in highly occluded indoor environments was not fully explored.

- Integrated methods using vision based and RTLS techniques:

Despite the advantages of vision-based systems and RTLS for monitoring tracked components, a single technology cannot provide the complete information needed to determine the status of elements on a job site. A review of existing studies demonstrates the great potential of integrating technologies to provide comprehensive data from resources in construction environments. However, few studies have applied such integrated models using vision-based models and RTLS to enhance the reliability and accuracy of monitoring systems. These studies were limited to ideal conditions where obstacles causing occlusion for vision-based techniques and creating multipath environments for the RTLS system are significantly less presented as compared to indoor field environments. Moreover, these integrated methods have not used the full advantage of each single technology. For example, fixed cameras were utilized in sites which may fail to monitor all required areas of a construction site where tracked components are located. Another challenge is the inefficiency of data analysis models in extracting necessary information from different sources, due to their complex computational requirements and the need for high-level computing power and resources. Due to these shortcomings, decision-makers may not be able to obtain the necessary information to take timely actions.

Based on the identified gaps, this study developed a method for automated progress monitoring and reporting that can timely collect and process data from tracked components to accurately provide their status information. The method employs digital imaging, novel deep learning-based object recognition algorithms and reliable RTLS for the automated recognition and localization of tracked components within challenging environments. In addition, a model was developed that can timely integrate data from the object recognition model and RTLS. It has the potential to deliver an organized and comprehensive report on the status of components. The research methodology and its modules are detailed in the following chapter.

CHAPTER 3: RESEARCH METHODOLOGY

3.1 Introduction

This chapter provides a comprehensive explanation of the research methodology to accomplish the objectives and address problems outlined in Chapters 1 and 2. The research methodology aims to support automated progress monitoring and reporting systems in indoor job sites during the construction phase. It specifically focuses on monitoring complex elements in challenging indoor environments, particularly MEP components. This area is less explored in current research, which has mainly concentrated on outdoor or simpler indoor settings. The overview of the methodology is depicted in Figure 3-1.

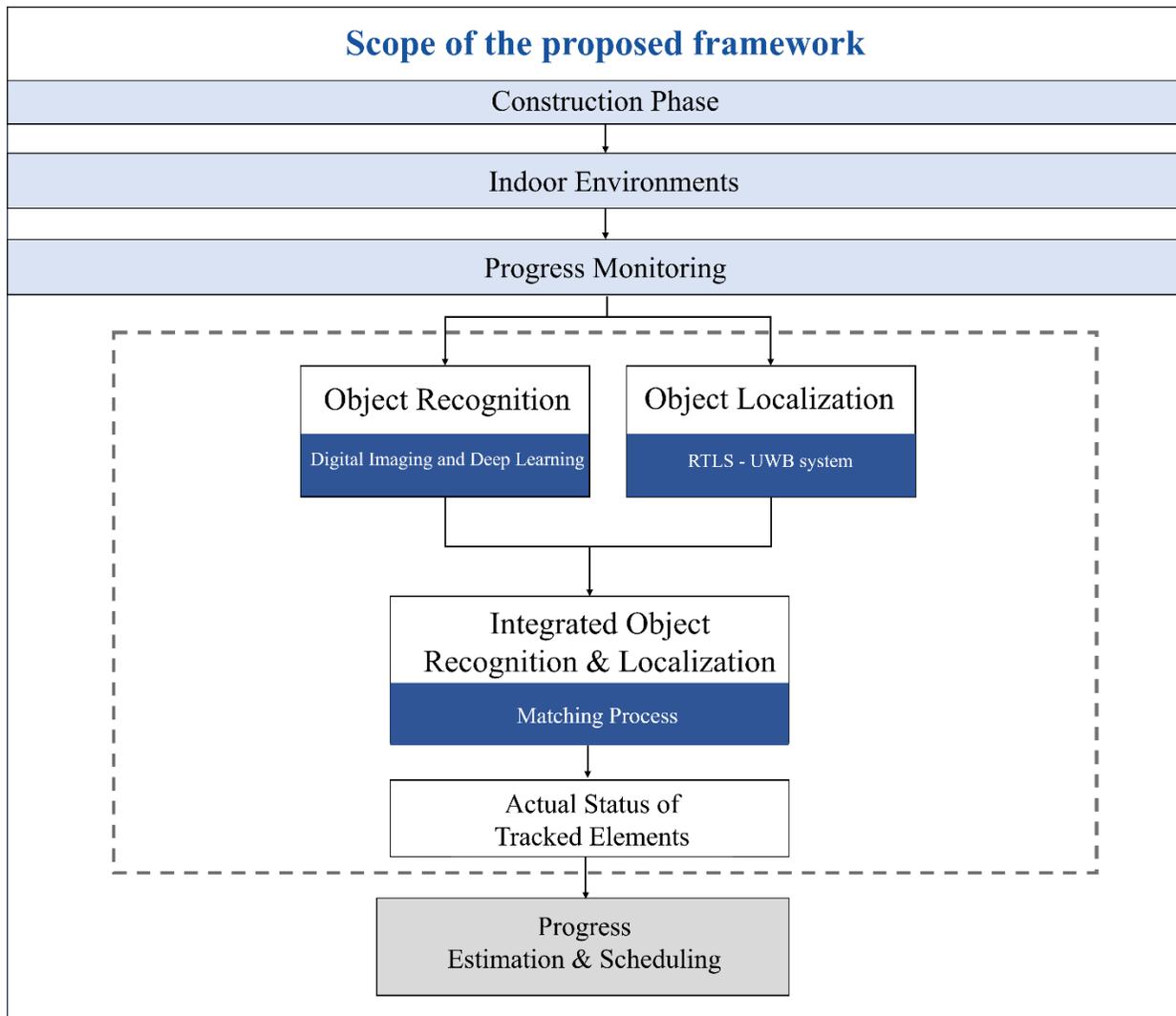


Figure 3-1. Overview of the research methodology.

The developed method encompasses three main modules:

- 1) Object recognition: which trains deep learning-based instance segmentation algorithms to automatically process images of tracked components in indoor environments. The outputs of deep learning models are 2D images wherein tracked components are classified, detected, and segmented.
- 2) Object localization: which identifies and localizes the corresponding tracked components in real-time using the UWB system. Its output is a text file containing each element's specific ID, its location in 3D coordinates and a timestamp.
- 3) Integrated object recognition and localization: it integrates the outputs of the two previous modules based on a matching process to automatically generate a comprehensive report about the status of tracked components in job sites. The data integration process enables the simultaneous recognition and localization of objects, which cannot be achieved with a single source. This method can enhance data management from different sources and facilitate understanding the actual project progress status.

The reasons for selecting these technologies for each module and the novelties of this research are elaborated in the following sections.

3.2 Object Recognition Module

Existing research on automated recognition in indoor environments typically focuses on simple-shaped objects and neglects real indoor challenges in construction job sites. However, the recognition of more complex, and detailed elements such as MEP components in challenging indoor environments, demands considerable attention. This module is developed to automatically recognize these components and it consists of different steps as shown in Figure 3-2. Namely:

- Image collection and labeling:

According to the literature, digital cameras were selected for visual data collection in this research since they are economical, easy to use, and accessible in most construction projects. In addition to real images captured at construction sites, virtual synthetic images were generated using BIM models. Image labeling was then performed by selecting and classifying regions of

objects in the images used in this development. Lastly, different mixes of synthetic and real images were created to find the optimum mix for model training.

- Model training and evaluation:

For the analysis of images, instance segmentation algorithms were chosen since they are the most comprehensive form of object recognition models. These algorithms are capable of object classification, detection, and segmentation simultaneously. In addition to detecting objects, they can apply pixel-wise segmentation for each instance separately. This segmentation allows predictions to be applied specifically to relevant regions of instances, providing more accurate recognition compared to using only bounding boxes. Furthermore, the models' abilities in both detection and segmentation are complementary tasks that enhance overall performance. For example, in instances with complex shapes where segmentation may not be ideal, object detection can still localize approximate areas of instances using bounding boxes.

This section's novelty lies in developing a methodology for the automated recognition of complex components in challenging indoor environments. It includes generating synthetic images using BIM models, which closely represent real indoor scenarios including lighting conditions, object complexity and scale, occlusion, clutter, and viewpoints to enhance dataset size and quality. In this research, two novel instance segmentation algorithms namely, Mask Region-based Convolutional Neural Network (Mask R-CNN) and You Only Look At CoefficientTs++ (YOLACT++) are employed. Particularly, YOLACT++ is a novel real-time instance segmentation algorithm in construction domain. It contains deformable convolutional neural networks version 2 (DCNv2), which enhances the model's ability to recognize objects with different scales, postures, rotations, and viewpoints in the images. This feature is essential for indoor construction environments due to the variety of components with different configurations.

To further enhance the models' performance in recognizing objects with different geometries and colour conditions, and mitigate overfitting problem, various data augmentation techniques were implemented. Lastly, different mixes of synthetic and real images were created to determine the optimum combination for effective model training.

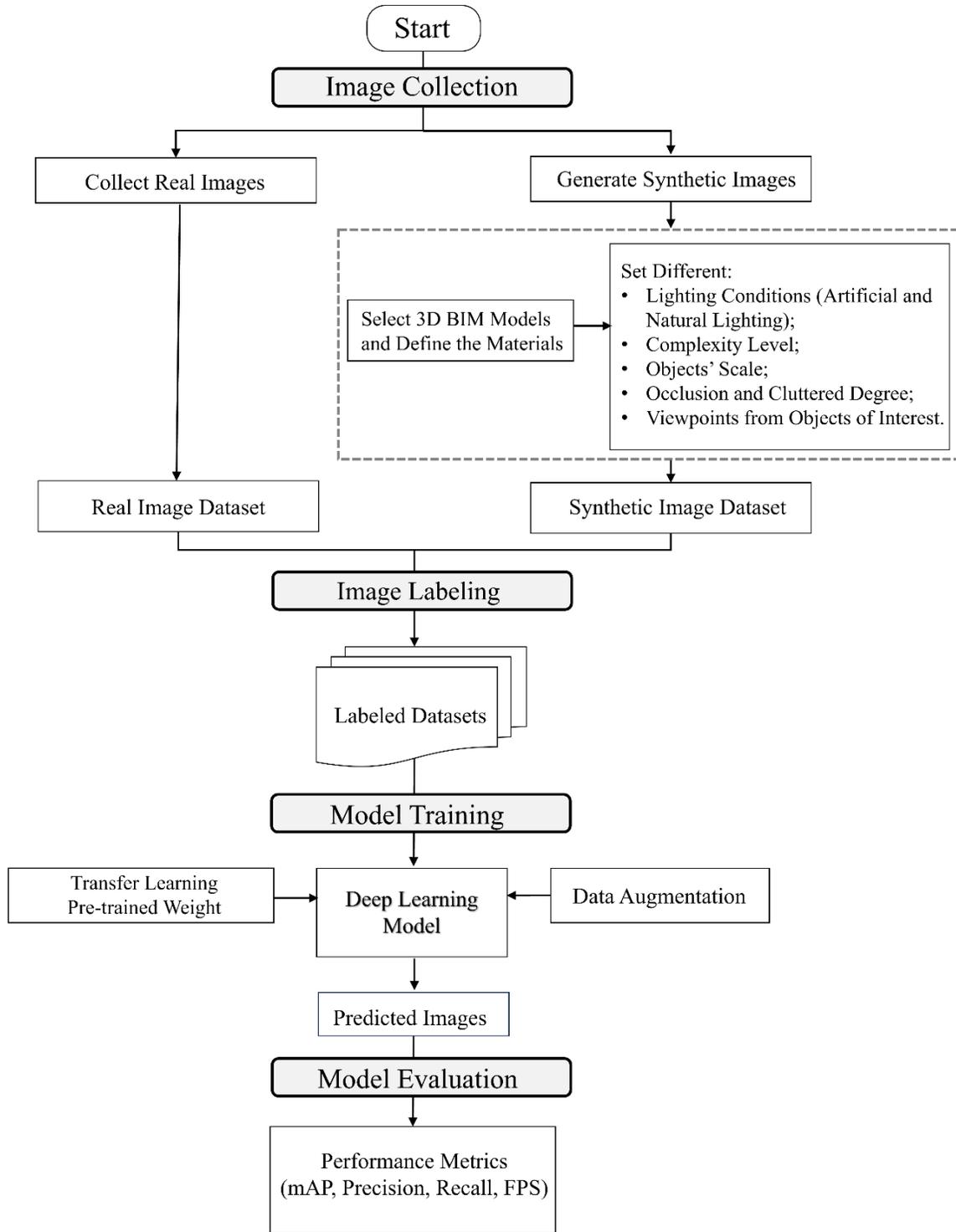


Figure 3-2. Overview of the object recognition module.

3.2.1 Image collection and labeling

For automated recognition of tracked objects using deep learning algorithms, a dataset containing a sufficient number of images from objects of interest is required. As a result, two datasets are created, one containing real images captured from construction sites and the other containing synthetic images. Similar to the recent research conducted in Golkhoo (2020), 3D BIM models are utilized to generate synthetic images. These synthetic images not only can generalize the datasets but also have the potential to support subsequent processes in construction progress monitoring.

Autodesk Revit 2019 is used as a BIM software to define properties of the building elements such as shape, material, texture, and dimensions. A rendering tool called, Enscape which is a plugin in the Autodesk Revit software is used for real-time synthetic image generation. The synthetic images were created with different modifications to improve deep learning models in recognizing elements under different indoor conditions, addressing the challenges of object recognition in such environments. These modifications cover aspects such as geometry, surrounding environment and lighting conditions, as well as camera viewpoints which are crucial challenges in indoor sites. Samples of these images are shown in Figure 3-3.

The considerations for image modifications are as follows.

- **Lighting conditions:** lighting adjustments were accounted for by adding or removing lighting sources including artificial and natural lighting and changing their intensities within the 3D models. The purpose was to create images with different lighting levels that are found in real indoor environments. Additionally, by changing the lighting intensities, the level of shadows and reflections on the objects were also altered.
- **Complexity of components:** this included changing the complexity of HVAC ducts and pipes in the images. This encompasses images with different network arrangements of HVAC ducts and piping. For instance, some images contain only a section of an HVAC duct or pipe while others encompass the entire complex network. Also, changes in objects' shapes contribute to this factor as well.

- Scale: the scale of HVAC ducts and pipes in the images was varied to address the challenge in recognizing objects of different sizes. This diversity included scales from small to large, could all appear in a single image to reflect the real-world array-like arrangement of HVAC ducts and pipes, or represented individually across different images.
- Occlusion and cluttered degree: this involved adding or removing elements within the BIM model, such as structural elements (e.g., beams and trusses), which are often located close to MEP components. By this consideration, we could mimic conditions where the HVAC ducts and pipes may be partially blocked or surrounded by other elements. These modifications are targeted to improve the model's ability to recognize these objects in different levels of occlusion and clutter.
- Viewpoints: images with different shooting angles of the objects were collected. In the 3D model, the viewpoints achievable in indoor locations were considered such as those from fixed cameras on walls or human perspectives. This approach enables the model to recognize objects from various viewpoints.

Next, combinations of real and synthetic images were explored to create high-quality image datasets and overcome the lack of available real images for training deep learning models. The advantages of determining the optimum combination of real and synthetic images in the training set include:

- Optimal model performance identification: this process facilitates identifying the model's peak performance for implementation in construction projects and works as a benchmark for future model implementations.
- Enhanced efficiency in data collection: this strategy improves the image collection process by reducing the time and cost of generating synthetic images and collecting proper real images through construction site visits and extensive web searches.
- Minimized data annotation effort: it can reduce manual annotation effort, which is time-consuming and costly, while still achieving comparable results.

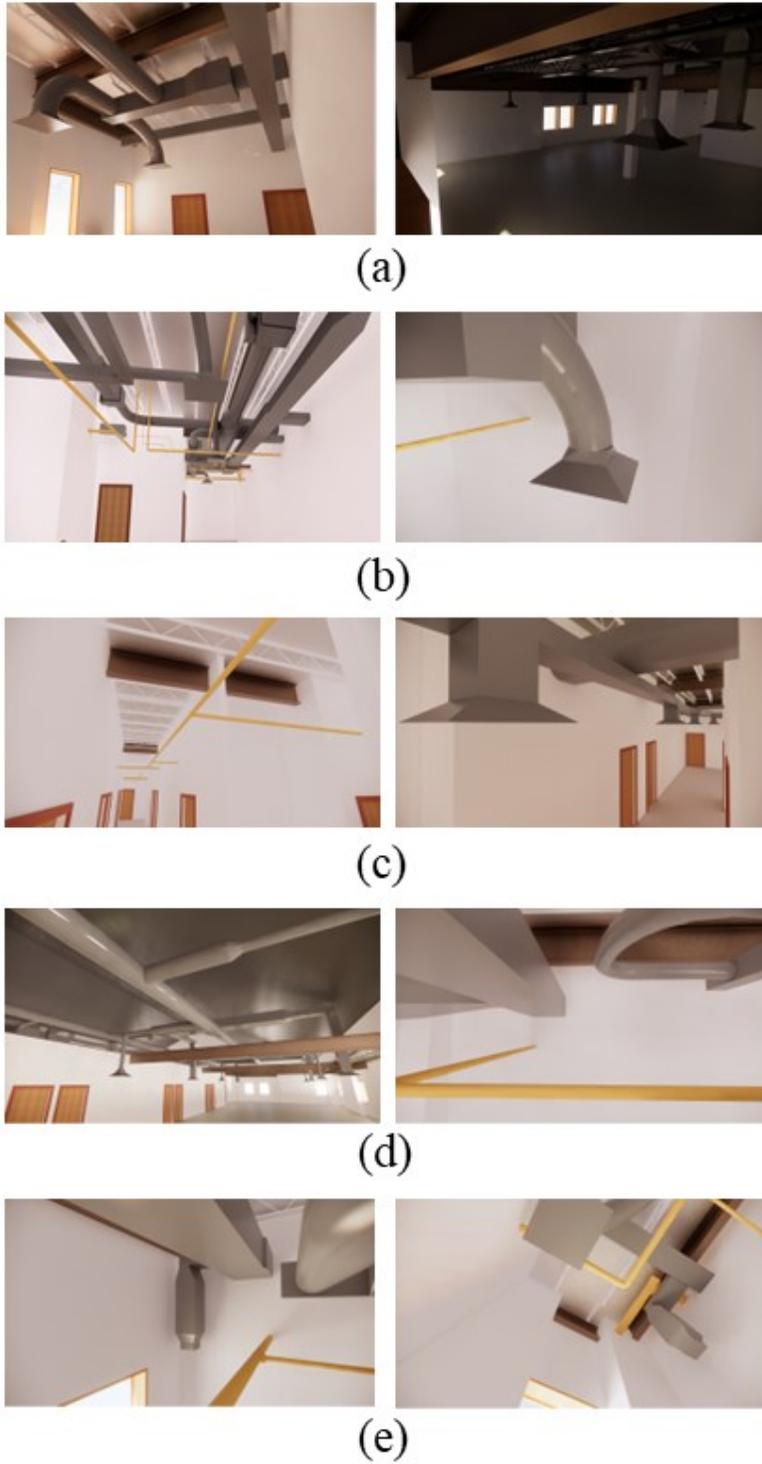


Figure 3-3. Generated synthetic images under different conditions: (a) Lighting condition; (b) Complexity level; (c) Scale level; (d) Occlusion degree; (e) Viewpoint level (Shamsollahi et al. 2024).

Consequently, eight different training datasets using both synthetic and real images were evaluated to find the optimum mix for effective model training. For these datasets, the number of real images was fixed, but the proportion of synthetic images was altered from 60% to 25% to the total images. In the final step, images within the datasets were labeled using VGG Image Annotator (VIA) web tool and labeling files were exported in JSON format for Mask R-CNN and COCO format for YOLACT++. In this research, polygon labeling was applied to images, and each object's boundary was meticulously determined. Polygon labeling has many advantages as compared to bounding box annotation, despite being more laborious. First, the predicted masks become accurate, closely aligning with the objects' region. Second, the labeling strategy can be employed for complex scenarios such as MEP networks, where components within the network can be identified as separate objects based on criteria such as major joints, material, component intersection or semantic distinctions. Consequently, the model successfully recognizes these objects as distinct components. This labeling strategy can be adjusted to be either more detailed and precise or more generalized, depending on project demands.

3.2.2 Model Training

According to the literature, deep learning-based object recognition algorithms have superior performance compared to traditional feature-based algorithms. Hence, for the automated recognition of tracked materials, two novel instance segmentation algorithms based on deep neural networks were selected. In this research, Mask R-CNN (He et al., 2017), a two-stage algorithm and You Only Look At CoefficientTs++ (YOLACT++) (Bolya et al., 2020), a one stage algorithm, were selected to evaluate both types of algorithms for object recognition.

To minimize the overfitting problem and to improve the model's generalization in recognizing instances with varying visual attributes (Liu et al. 2016), different sets of image augmentation techniques were implemented during the training of both algorithms. The augmentation techniques encompass geometric transformations such as flips, mirrors, and rotations as well as color transformations such as brightness and contrast. Geometric transformations were utilized to adapt the model to different geometry and spatial alternatives, while color transformations were applied to improve the model's recognition under different lighting and color conditions. Moreover, to further enhance the model's performance and avoid initiating training from the base level for learning image features (Nath & Behzadan, 2020; Zheng

et al., 2020) transfer learning was applied in both algorithms.

Mask R-CNN and YOLACT++ were implemented using cloud-based services, offering several advantages in this study. These included on-demand access to large image datasets and the deep learning model as well as resources such as GPU servers and memory which are essential for the training and testing of the model. Moreover, these services provide unified integration of input data with the model and eliminate the need for specific high-cost hardware and software systems. In this study, Google Drive was used to store the image datasets and then linked to Google Colaboratory (pro version) for model training with a Tesla P100/V100, 16GB GPU and Python3.

3.2.2.1 Mask R-CNN Network Architecture

Mask R-CNN is one of the foundational algorithms in the instance segmentation field. It is an extension of the Faster R-CNN model which is an object detection algorithm. In Mask R-CNN, as compared to Faster R-CNN, a mask prediction branch is added to the prediction network in parallel with the classification and localization branches. The detail of Mask R-CNN architecture is depicted in Figure 3-4.

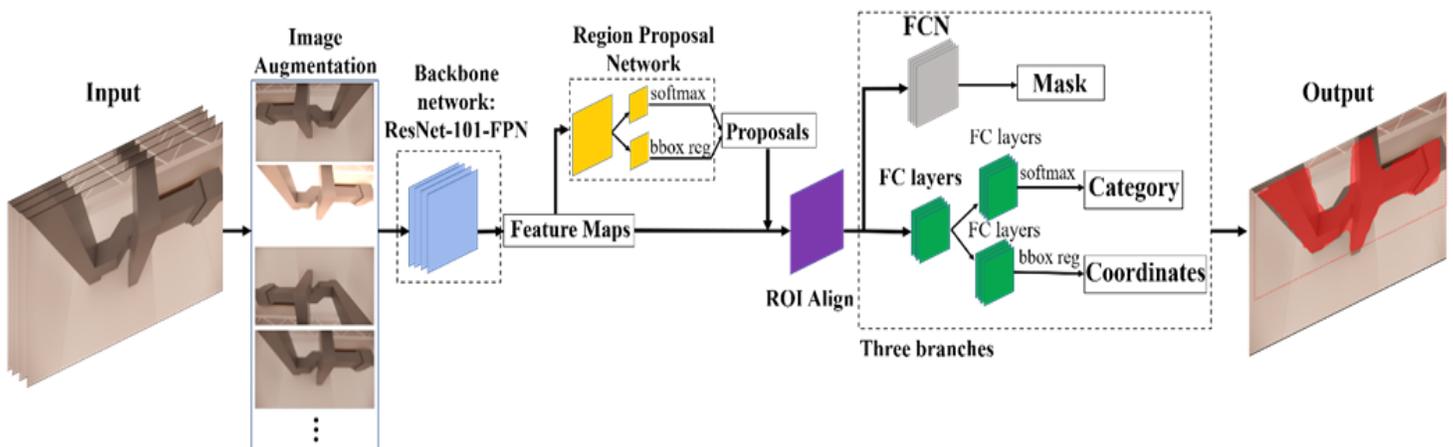


Figure 3-4. Mask R-CNN network architecture (Shamsollahi et al., 2021).

The model's architecture comprises a convolutional backbone network based on ResNet-101 and Feature Pyramid Network (FPN), for extracting feature maps from the input images. The feature maps are subsequently passed into the Region Proposal Network (RPN) to generate the Regions of Interest (RoIs) for the head network. Also, the Mask R-CNN is utilizing a quantization-

free layer, called RoI Align for extracting predefined size feature maps from each RoI. In the head network, three parallel branches exist for classification, bounding box regression, and mask prediction. Fully connected layers are used for object classification and bounding box regression for each RoI. In parallel a separate branch uses a fully convolutional network (FCN) to predict masks for each RoI and classifies each pixel in the image to a predefined object class.

The total loss function in the Mask R-CNN model is the sum of classification loss, the bounding-box regression loss, and the mask loss. The classification and bounding box loss functions are the same as the ones utilized in the Faster R-CNN model (Girshick, 2015), these loss functions, quantify the classification and detection errors of the model. In addition, the Mask loss measures pixel-level classification error which is described in He et al. (2017).

Training the Mask R-CNN is based on the Matterport’s implementation (Abdulla W., 2017) using the open-source libraries Keras and Tensorflow. The model training was initialized by utilizing pre-trained weights on the MS COCO dataset (Lin et al., 2014) rather than training the model from scratch.

The following hyperparameters were selected for training the model, as they achieved optimal performance after tuning and evaluating various values for each (a sample of trials is provided in Appendix D): (1) batch size = 2, (2) weight decay = 0.0001, (3) learning rate = 0.001, and (4) epoch = 90.

Different sets of image augmentation techniques such as Horizontal Flip, Vertical Flip, Rotation, Gaussian Blur and Brightness are investigated to create modified copies of the existing data. Details of this investigation are provided in Table 3-1. Samples of output images generated by these techniques are depicted in Figure 3-5. The imgaug library (Jung et al., 2020) was utilized to augment images during model training. The effect of augmentation techniques on Mask R-CNN performance was examined using two models: one trained with these techniques and one without.

Table 3-1. Data augmentation techniques and selected parameters for Mask R-CNN training.

Data Augmentation Technique	Parameters
Flip	Horizontal & Vertical

Rotation	One of $\Theta=90^\circ, 180^\circ, 270^\circ$
Brightness (Multiply)	(Adding value) (0.8,1.5)
Image smoothing (Gaussian blur)	(σ value of Gaussian kernel) (0.0,5.0)

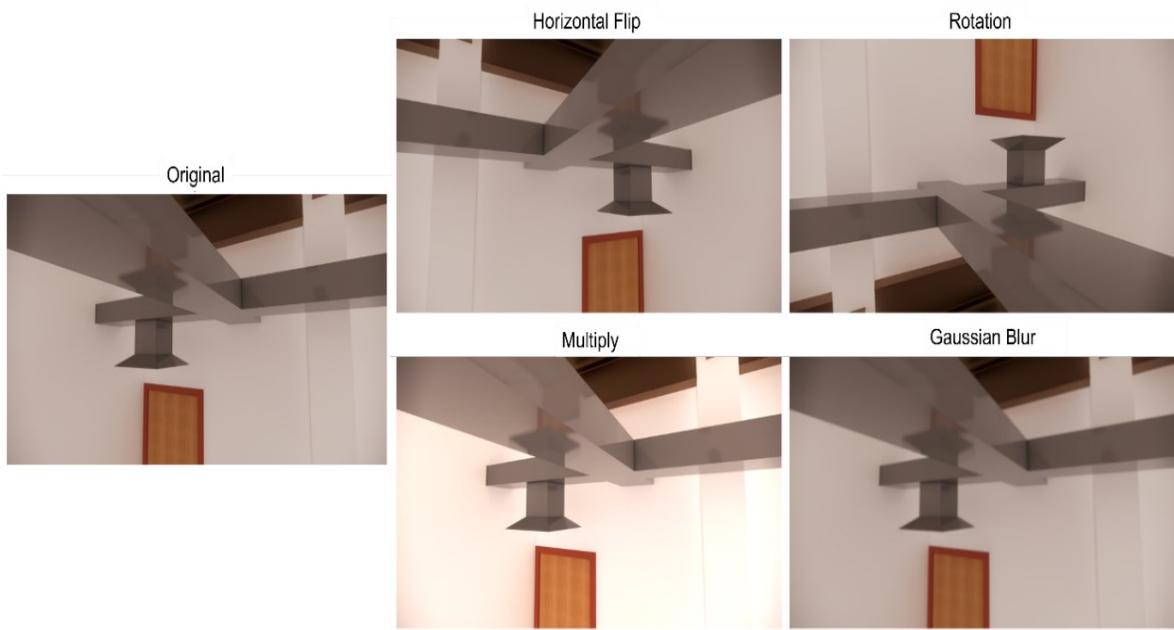


Figure 3-5. Samples of applied data augmentation techniques.

3.2.2.2 YOLACT++ Network Architecture

YOLACT++ is a one-stage instance segmentation algorithm, capable of classifying, detecting, and segmenting objects with high-quality masks in real-time (Bolya et al., 2020). It is an extended version of YOLACT (Bolya et al., 2019) with several improvements. The YOLACT architecture consists of a backbone network, Protonet, and a prediction head. The backbone network utilizes ResNET-50 /101 (He et al., 2016) and FPN (Lin et al., 2017) for feature extraction. For final mask prediction for each instance within the image, two complementary parallel processes are applied (1) generating k prototype masks via Protonet and (2) predicting mask coefficients by the prediction head. Prototype masks are generic full image feature maps which are not specific to any

individual instances. Hence, mask coefficients are utilized to specifically align the features of instances in the prototype with each detected instance in the image.

To produce k prototype masks, the Protonet, a FCN, with k channels in the last layer, is added to the network. It receives image feature maps from the deepest layer of FPN, the P_3 layer, and then applies upsampling to create high-quality predicted masks and improve the model's performance when dealing with small objects. The prediction head, attached to the FPN, consists of three parallel branches for object classification, bounding box regression, and mask coefficient prediction on the outputs of P_3 to P_7 layers of the FPN. The prediction head network is not complex and consists of a 3×3 convolutional layer used by all branches as well as a 3×3 convolutional layer within each branch to enable faster prediction.

The first two branches in the head network are for classification and bounding box regression tasks. The third branch generates k mask coefficients, each matching with one of the k prototypes from the Protonet. Fast NMS was used to select high-confidence detections according to IoU thresholds. Next, in the mask assembly process, the generated prototypes and mask coefficients were linearly combined to generate high-quality predicted masks. To clear predicted pixel masks that are out of the bounding box area and fit them into their bounding boxes, a cropping operation was applied to the final masks. During the training, this was obtained by using the ground truth bounding boxes, while during testing, the predicted bounding boxes were used. Following this, thresholding was applied to the cropped predicted masks to select which one should be kept as the final mask for each instance in the image.

YOLOACT++ enhances YOLOACT with some improvements. Fast Mask Re-Scoring is added with six convolutional layers after mask cropping. This improves segmentation reliability by realigning the classification confidence with the accuracy of the predicted masks. DCNv2 (Zhu et al., 2019) was also included in the ResNet architecture where convolutional layers of 3×3 dimensions in stages C3 to C5 are exchanged with deformable convolution layers of the same dimensions. Unlike standard convolution that use fixed grid structure, DCN (Dai et al., 2017) employs 2D offsets, allowing flexible and irregular sampling grids depending on input feature maps, as depicted in Figure 3-6. This improves the network adaptability to address instances with varying sizes, postures, and viewpoints leading to improved prediction performance. The

prediction head was also optimized by variations in anchor configurations, specifically scales, and sets of aspect ratios. The architecture of YOLACT++ is illustrated in Figure 3-7.

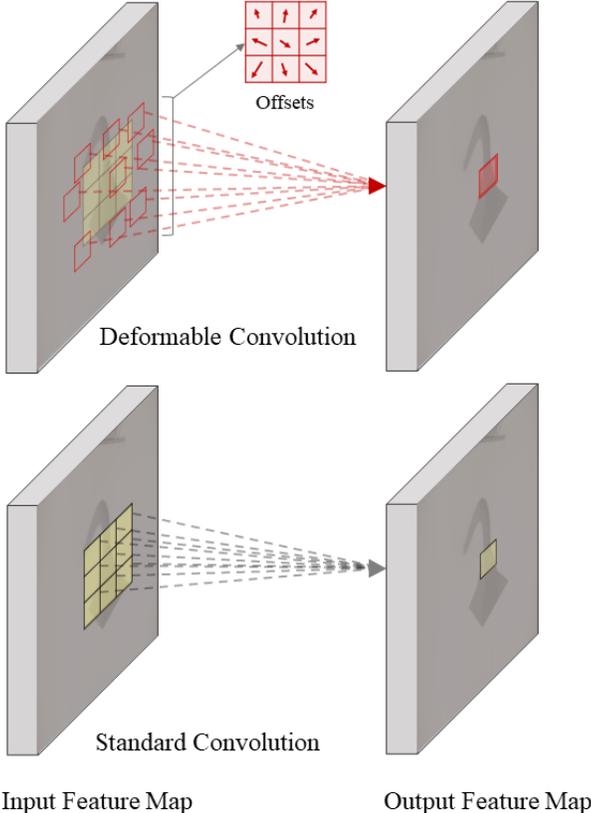


Figure 3-6. 3*3 Standard and deformable convolution architecture (Shamsollahi et al. 2024).

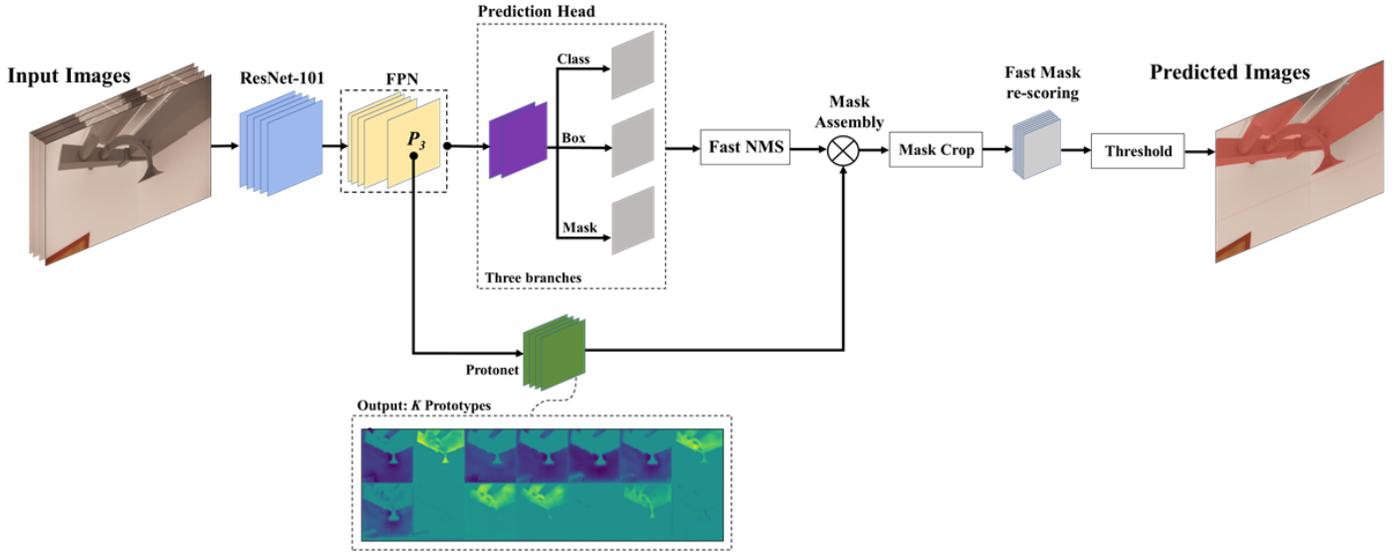


Figure 3-7. YOLACT++ network architecture (Shamsollahi et al. 2024).

As YOLACT++ performs multiple tasks, three loss functions were utilized. These include classification confidence loss (L_{cls}) for object classification, bounding box loss (L_{box}) for object localization as described in (Liu et al., 2016) and mask loss (L_{mask}) for instance segmentation as outlined in (Bolya et al., 2020). The model's total loss function (L_{Total}) is defined as follows:

$$L_{Total} = \omega_{cls}L_{cls} + \omega_{box}L_{box} + \omega_{mask}L_{mask} \quad (1)$$

The weight of each loss function is represented by ω_{cls} , ω_{box} , and ω_{mask} with values of 1, 1.5 and 6.125 respectively, as specified in the YOLACT++ paper (Bolya et al., 2020). For L_{mask} , pixel-level binary cross entropy (BCE) between predicted masks (M) and ground truth annotations (M_{gt}) was calculated as shown:

$$L_{mask} = BCE(M, M_{gt}) \quad (2)$$

The YOLACT++ model was built based on Dbolya's Implementation (Bolya et al. 2019) using the Pytorch framework (Paszke et al., 2019). ResNet-50 and FPN were used as the backbone network. The hyperparameters for training the model were set as the epochs of 90 and 116, batch sizes of 4 and 8, a learning rate of 0.001, momentum at 0.9, weight decay of 0.0005, and gamma at 0.1.

The aspect ratios were set at $[1, 1/2, 2]$ and the IoU threshold for both box and mask at 0.5 as a standard object recognition benchmark. Input images were resized to $550 * 550$ pixels, prototype masks to $138 * 138$ pixels, and k for prototypes was set to 32. Samples of the prototype masks developed by the Protonet are depicted in Figure 3-8, where yellow colors represent higher values, while blue signify lower ones (Bolya et al., 2020). The transfer learning technique was applied using pre-trained weights on the ImageNet dataset (Deng et al., 2010).

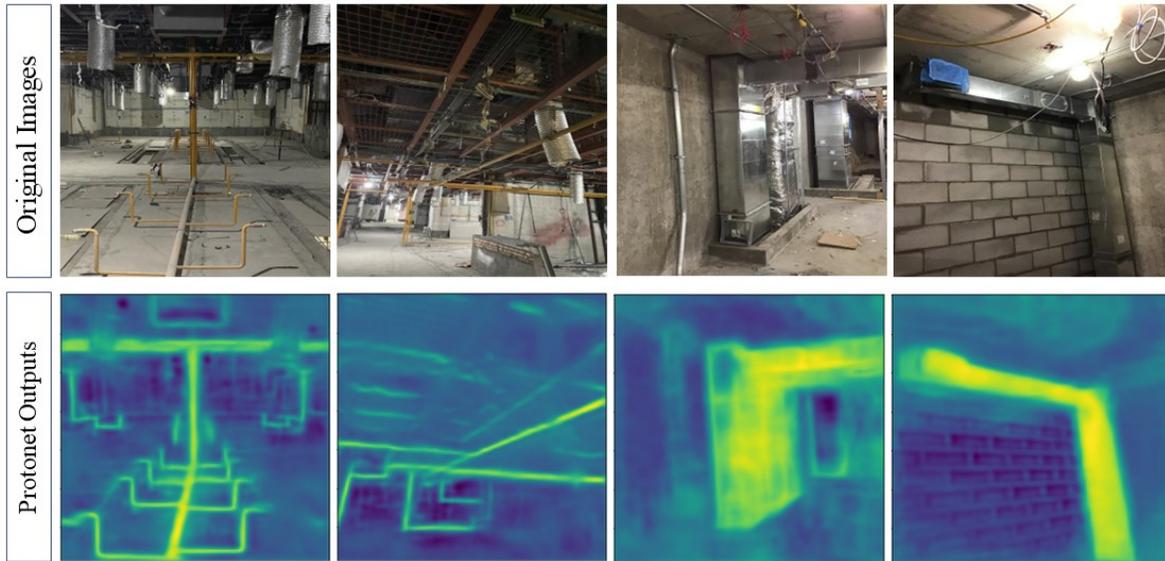


Figure 3-8. Original images and the Protonet outputs (Shamsollahi et al. 2024).

The data augmentation techniques utilized for training YOLACT++, along with their parameters, are included in Table 3-2. The selection of these techniques was based on Bolya et al. (2020) and Liu et al. (2016). It includes color transformations such as random contrast, brightness, lighting noise, hue, swap channel, and saturation and geometric transformations such as crop, resize, flip, mirror, and expand techniques. These techniques were implemented by using the Pytorch library (Paszke et al., 2019). Samples of these techniques are shown in Figure 3-9.

Table 3-2. Applied data augmentation techniques, their parameters, and values.

Augmentation Technique	Parameters	Values
Saturation	(Adding Value)	(0.5,1.5)
Hue	(Adding Value)	(-18,18)

Brightness	(Adding Value)	(-32,32)
Contrast	(Adding Value)	(0.5,1.5)
Lighting noise	Channel Permutation	(0, 1, 2), (0, 2, 1), (1, 0, 2), (1, 2, 0), (2, 0, 1), (2, 1, 0)
Sample Crop	Crop Size Range IoU Range	(0.3, 1) of Image Width x Height [None, 0.1, 0.3, 0.7, 0.9]
Expand	Expansion Ratio Range	(1,4)
Mirror	Horizontal	-
Flip	Horizontal	-
Rotation	Rotation angle	90,180,270

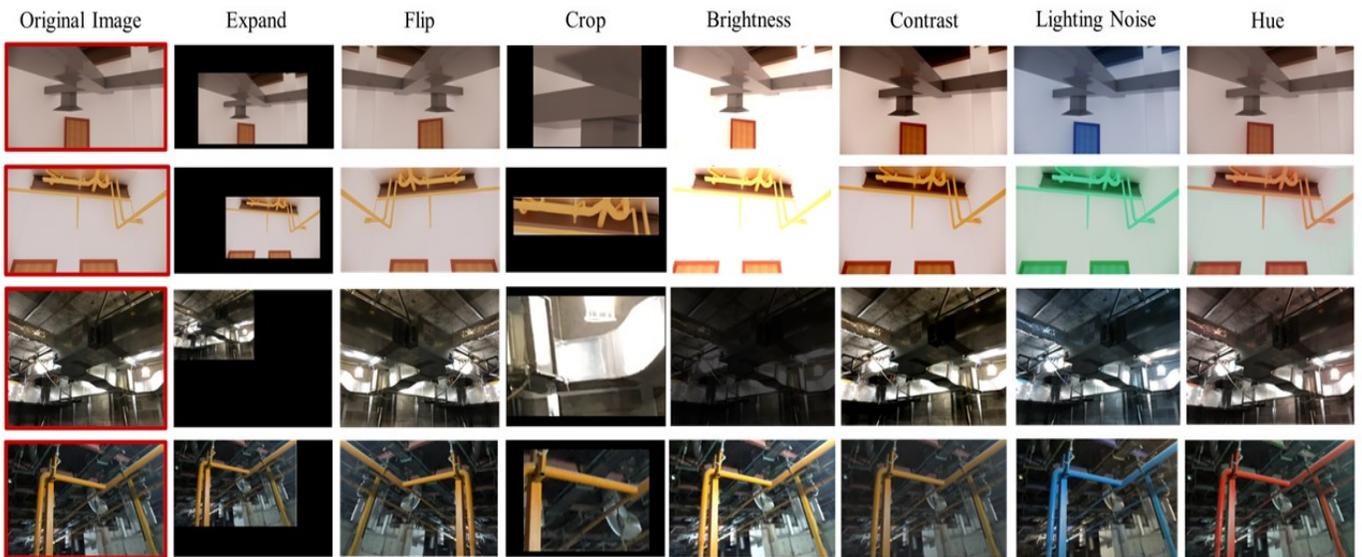


Figure 3-9. Sample of applied data augmentation techniques for the YOLACT++ model (Shamsollahi et al. 2024).

Overall, YOLACT++ offers a number of advantages that are appropriate for this research. First, YOLACT++ contains DCNv2 which enhance the model’s ability to recognize objects of varying sizes, poses, rotations and viewpoints. This feature is helpful in cluttered indoor

environments where building components can appear in numerous shapes and scales. Second, it is effective in handling small objects which is the limitation of many object recognition algorithms like Mask R-CNN and YOLO (Pal and Hsieh 2021; Yang et al. 2020). This is a great advantage for indoor environment since there are many small-scaled objects in indoor sites. Lastly, according to Bolya et al. (2020), YOLACT++ is the first real-time instance segmentation model capable of predicting at 30 frames per second (FPS) as evaluated on the MS COCO test dataset. While many real-time object detection algorithms exist, such as YOLO and SSD, however for instance segmentation achieving real-time prediction is more challenging. Speed is an important factor when integrating object recognition models with monitoring and reporting systems to ensure timely delivery of information to project managers.

3.2.3 Model Evaluation

Precision, Recall and F1-score were selected as three performance metrics to evaluate the performance of the object recognition models. Precision is calculated as the ratio of True Positives (TP) to the total positive predictions including TP and False Positives (FP). Recall is calculated as the ratio of TP to the total actual positive samples in the dataset including TP and False Negative (FN). The precision, recall and F1-score formulas are as follows:

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$F1\text{-score} = 2 \left(\frac{Precision \times Recall}{Precision + Recall} \right) \quad (5)$$

Precision identifies how many of the model's detections are true and recall demonstrates the model's ability in finding true positives of all predefined ground truths (Padilla et al., 2020; Zhang & Zhang, 2021). In the context of object recognition algorithms, mean Average Precision (mAP) is a metric that is used to evaluate the performance of bounding box and mask predictions across all object classes. It is calculated based on the mean of average precision values across all classes. According to the Pascal VOC2010–2012 benchmarks, for a given Intersection over Union (IoU)

threshold, AP is calculated as the area under the precision-recall curve, which ranges from 0 to 1 (Everingham et al., 2012; Padilla et al., 2020). mAP is calculated as follows:

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (6)$$

where AP_i equals the average precision of i^{th} object class and N represents the total number of object classes considered for evaluation. More information about the calculation of AP can be found in Padilla et al. (2020).

For calculating both mask mean Average Precision (mAP^{mask}) and bounding box mean Average Precision (mAP^{bbox}), Equation 6 is utilized with a key difference: for mAP^{mask} , Mask IoU is measured based on the overlap between the predicted segmented mask and the actual mask. For mAP^{bbox} , IoU measurement is based on the overlap of the predicted and actual bounding boxes. More details about the difference between Mask IoU and Box IoU are depicted in Figure 3-10.

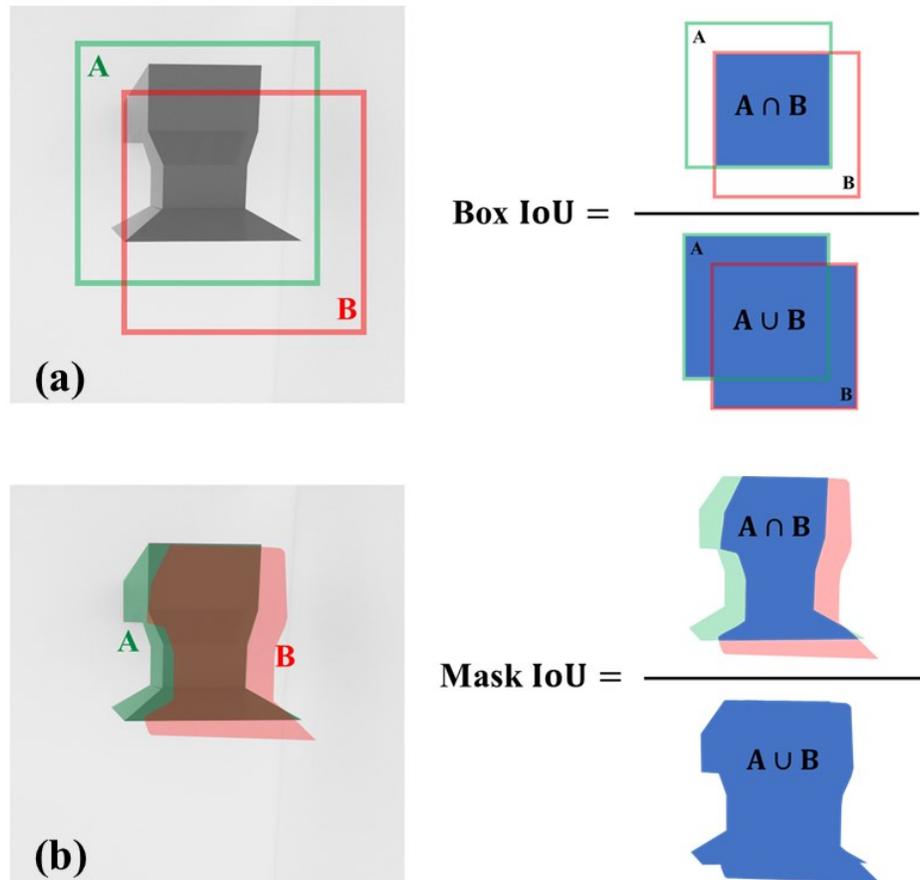


Figure 3-10. Box IoU and mask IoU (Shamsollahi et al. 2024).

3.3 Object Localization Module

This section focuses on the localization and identification of static materials in an indoor environment. According to the literature, the UWB system was selected as the RTLS for this module due to its high accuracy in 2D and 3D localization in indoor environments. This system has two main advantages including high multipath resolution and anti-interference potential (Zhang et al., 2020), which makes it an appropriate solution for complex, high multipath construction environments.

This research aims to evaluate the UWB system's performance within indoor construction environments for supporting automated progress monitoring and reporting. Previous studies have implemented the UWB system in indoor laboratory areas or under ideal conditions. In contrast, this research, in addition to laboratory experiments, considers challenges of actual indoor environments including multipath effects, signal blockage and occlusions. The study focuses on the 2D/3D localization of challenging objects such as metallic MEP components located at various positions and heights within indoor environments. Additionally, it identifies factors that can affect the UWB system's performance through indoor laboratory and field experiments.

3.3.1 The UWB System

The UWB system consists of three main components, namely:

- (i) UWB tags, which are attached to tracked elements to identify them with their specific ID and location in the site.
- (ii) Receivers, which are installed in fixed places around the test area.
- (iii) Software installed on a laptop for real-time data recording and visualization.

Figure 3-11 demonstrates the steps required for object localization using the UWB system. In the first step, it is important to identify the installation requirements of the UWB system such as the distances between receivers and walls or ceilings. Next, a layout for the placement of receivers needs to be designed. This layout should consider different factors related to the conditions of the specific test area including the presence of obstacles and the locations of tracked objects.

The number of receivers is determined according to the layout design. It is needed to install the receivers properly in their predefined locations since the accuracy of the UWB system highly depends on the positions of receivers. According to the type and quantity of the tracked elements in the selected site, the number of tags is identified. The installation location of tags and receivers during the construction process can be determined using several approaches. Existing building drawings, which illustrate the locations of tracked elements can be utilized to identify placements. Moreover, BIM models can simulate various layouts and placements in a 3D environment. These simulations support optimizing the locations to ensure maximum LOS between receivers and tags at different stages of construction. Site visits and consultation with the field managers allow us to refine these placements by considering site conditions. Another factor for the successful implementation of the UWB system is adjusting the receivers layout at various construction stages, based on congestion levels, accessibility, and the number of tracked elements. Attaching tags on components can vary depending on the construction stage. For instance, tags may be attached at the manufacturing company before shipment, upon delivery to the site or at a comfortable height prior to installation. If the objects are already installed, it is necessary to reach elements to attach tags.

After activation of tags and receivers within the site, data collection can be initiated using the software application installed on a laptop. The data collection process begins at time t_0 in order to record the 3D coordinates of the tags. After gathering sufficient data, the collection is stopped at $t_0 + \Delta t$ through the software application. Once data collection has been stopped, a report will automatically be generated. The report of the UWB system is a text file containing records of tags' information including tag ID, timestamp, 3D coordinates (x, y, z), and range measurements as depicted in Figure 3-12.

It is necessary to measure the true location of each tag and receiver within the site in order to assess the UWB system's error. In addition to the statistical analysis of the UWB data points, the 2D and 3D localization error of each tag is calculated as the Euclidean distance between the true location of each tag (X_{True} , Y_{True} , Z_{True}) and its average UWB location estimation (X_{Mean} , Y_{Mean} , Z_{Mean}) which are described below:

$$\mathbf{Error\ 2D} = \sqrt{(X_{True} - X_{Mean})^2 + (Y_{True} - Y_{Mean})^2} \quad (7)$$

$$\mathbf{Error\ 3D} = \sqrt{(X_{True} - X_{Mean})^2 + (Y_{True} - Y_{Mean})^2 + (Z_{True} - Z_{Mean})^2} \quad (8)$$

To measure the actual locations of tags and receivers, one receiver was set as the reference point. The locations of the remaining receivers and tags were measured based on this reference using a measuring tape or laser distance measurer with an accuracy level of within ± 0.0032 meters. To ensure the accuracy of measurements, the location measurements were repeated three times and then averaged. According to the findings presented by Maalek & Sadeghpour (2013), placing tags on top of components enhances the UWB system's accuracy in object localization due to a better LOS with receivers. Therefore, in this study, the preferred location of tags on the tracked components was on the top. If access to the top of component was restricted, the tag would be placed below the component.

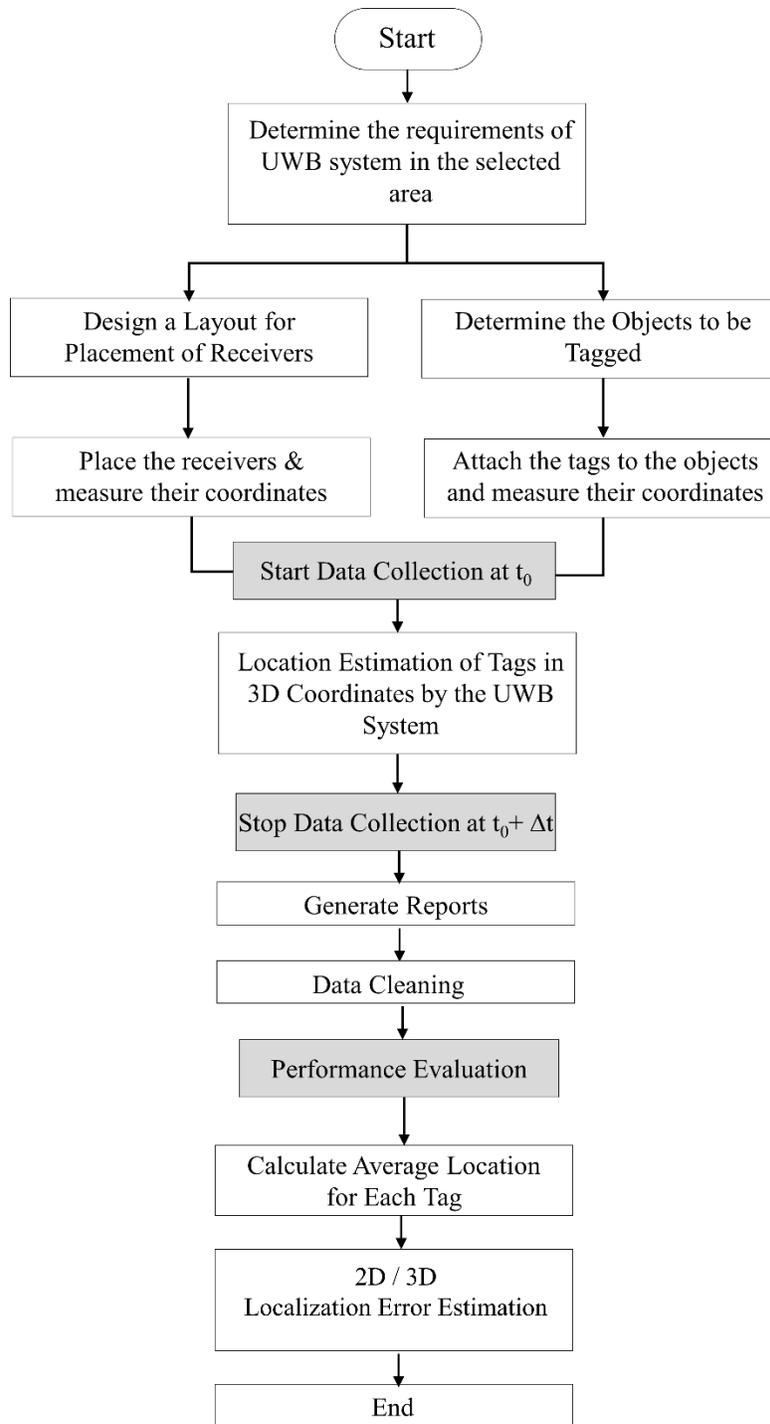


Figure 3-11. Process of implementing the UWB system for object localization.

```

20220725_125215RTLS_log-3 tags on- no filter - Notepad
File Edit View
Timestamp Tag ID
T:125215582 DecaRangeRTLS:LogFile:Ver. 2.10 TREK:Conf:Anchor0:0:Chan2
T:125215587 AP 0 0:0:2.31
T:125215587 AP 1 6.6:0:2.31
T:125215587 AP 2 6.6:5.4:2.31
T:125215587 AP 3 0:5.4:2.745
T:125215632 RR 0 0:5620:5620:118:55906
T:125215632 RR 0 1:4444:4444:118:55906
T:125215632 RR 0 2:3133:3133:118:55906
T:125215632 RR 0 3:4866:4866:118:55906
T:125215632 RM 0 15:118:55906
T:125215632 LE 0 17056:118:[4.19661,3.61976,1.37717] 5620:4444:3133:4866
T:125215632 TS 0 avx:4.19514 avy:3.61726 avz:1.47684 r95:0.0431733
T:125215660 RR 1 0:2692:2692:4:55910
T:125215660 RR 1 1:5217:5217:4:55910
T:125215660 RR 1 2:6013:6013:4:55910
T:125215660 RR 1 3:4510:4510:4:55910
T:125215660 RM 1 15:4:55910
T:125215662 LE 1 14941:4:[1.7871,1.87231,1.57001] 2692:5217:6013:4510
T:125215663 TS 1 avx:1.80917 avy:1.83319 avz:1.46003 r95:0.0444683
T:125215688 RR 2 0:5747:5747:234:55914
T:125215688 RR 2 1:2318:2318:234:55914
T:125215688 RR 2 2:3858:3858:234:55914
T:125215688 RR 2 3:6707:6707:234:55914
T:125215688 RM 2 15:234:55914
T:125215688 LE 2 16994:234:[5.39507,1.81935,1.52819] 5747:2318:3858:6707
Ln 1, Col 1 100% Windows (CRLF) UTF-8

```

Figure 3-12. Sample of the UWB output file.

To evaluate the UWB system, laboratory experiments are conducted in a controlled environment with LOS conditions. These experiments provide important information before the system’s implementation in real projects, helping to define baselines, installation requirements, and the layout of the receivers. Moreover, they validate the system’s hardware and software capabilities and identify parameters that might affect the UWB system’s performance. This leads to enhanced functionality, cost and time savings and improved accuracy when the system is deployed in real job sites.

The objectives of the laboratory experiments are:

- Assessing the impact of tags' heights: This includes evaluating the impact of tag placements at different heights on the system’s performance. This experiment helps to assess the applicability of the system in real construction job sites, specifically, when tracked elements are placed at different heights. These elements can be located at floor level, either awaiting installation or already installed at elevated heights.

- Exploring the effect of tag proximity to receivers: It includes assessing the UWB performance when the tags are located close to the receivers.

The field experiments are implemented to evaluate the UWB system's performance under both LOS and NLOS conditions in real indoor job sites and high multipath environments. Following are the objectives of the field experiments:

- Localizing metallic objects such as MEP components in indoor environments.
- Assessing the UWB system performance in challenging scenarios such as congested indoor environments filled with metallic objects which can cause signal blockage or reflection.
- Studying the effect of LOS/NLOS conditions between tags and receivers on the system performance.
- Determining the localization accuracy of tags located out of the enclosure area created by the UWB receivers.
- Examining the number of active receivers on localization accuracy.
- Assessing the effect of the number of active tags on the UWB localization accuracy.

Each generated UWB report in a .txt format was imported in Google Colaboratory (pro version). Data cleaning, analysis and error assessment were conducted using Python and its frameworks including NumPy and Pandas. In data cleaning, unneeded information such as range measurements, receiver positions, and receiver-to-receiver distance reports was filtered out. The required information, such as the Tag ID, timestamp, and tag location in 3D coordinates, was retained for further analysis.

3.3.2 The UWB System's Configurations

Based on the performance assessments by Jiménez and Seco (2016) and Ruiz and Granja (2017), detailed in Section 2.2.2.1, the UWB system from Decawave (Qorvo) achieved better results compared to other available alternatives. Hence, in this research, the Trek1000 UWB Evaluation Kit from this company is used for the localization of tracked elements.

In the Trek1000 unit, the location of each tag is identified by using the trilateration technique which uses the measured distances between the tag and the receivers. These distances were measured through the two-way ranging (TWR) ToF technique. The receivers and tags were powered by portable power banks chargers and universal serial bus (USB) cables. The UWB

output report was automatically stored to a local computer which is connected to one of the receivers with a USB cable.

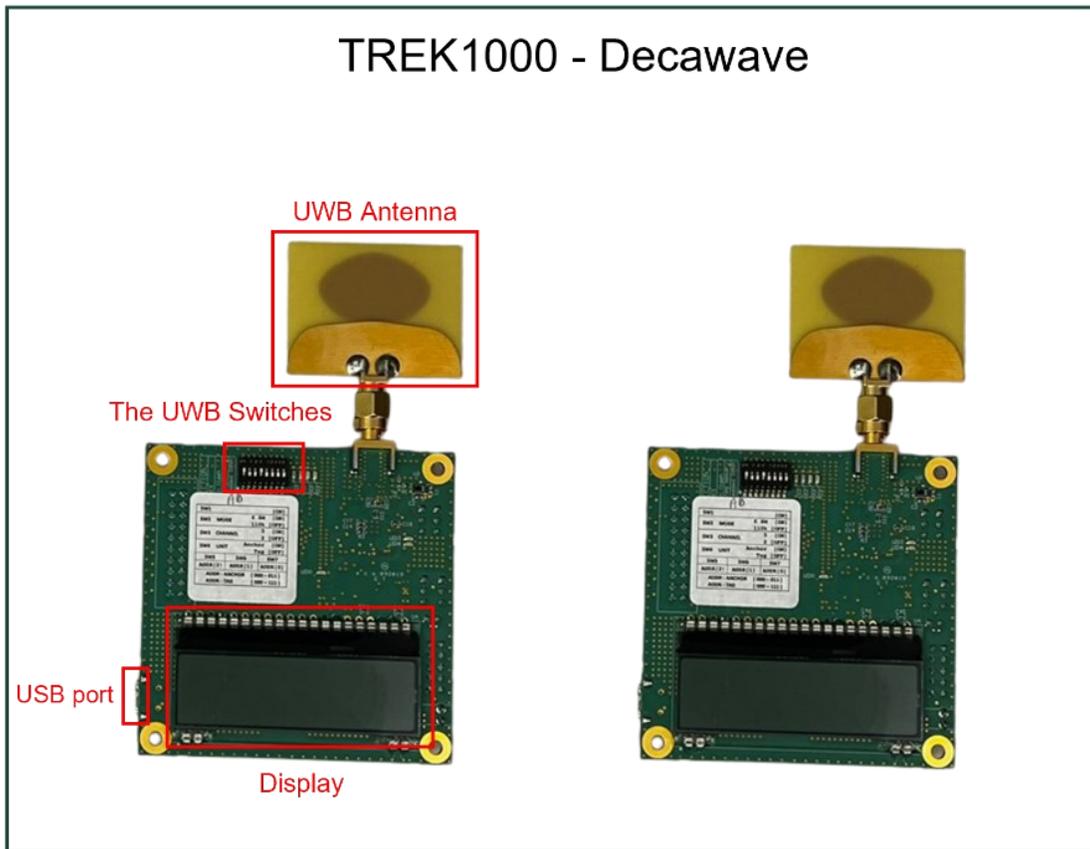


Figure 3-13. Trek1000 tags and receivers' components.

Each receiver or tag component consists of a Printed Circuit Board (PCB) containing a configuration switch and display screen as well as an antenna that is fixed on the PCB. Through the configuration switch, the type of the PCB unit can be configured as a Tag or Receiver. In addition, the type of the channels (Channel 2: 3.993 GHz or Channel 5: 6.489 GHz), data rate (6.8 Mbps or 110 kbps), and unit ID number can be set. After activation of the tag or receiver, the display screen demonstrates the unit type (tag or receiver) and its corresponding ID number. For these experiments, all tags and receivers were set to channel 2: 3.993 GHz with a 110-kbps data rate. The main components of the Trek1000 tags and receivers are depicted in Figure 3-13.

Figure 3-14. shows the user interface of the UWB application. The actual location of the receivers in 3D coordinates needs to be manually entered into the application. The next step is

selecting the application settings based on the objectives of the experiments. In all experiments conducted in this research, the tracking mode was activated, and auto-positioning of receivers was deactivated. The UWB application displays active tags and receivers in the testing area in real-time. The following recommendations from the manufacturer were considered for the experiments:

- Install four receivers in the test area to estimate the locations of tags in 3D planes.
- Install three receivers at the same height level and the fourth one higher, with the maximum possible distance.
- Locate the receivers in a way to maintain a clear line-of-sight with each other.
- Ensure that all receivers are installed at a minimum 15-centimeter distance from any wall or ceiling.

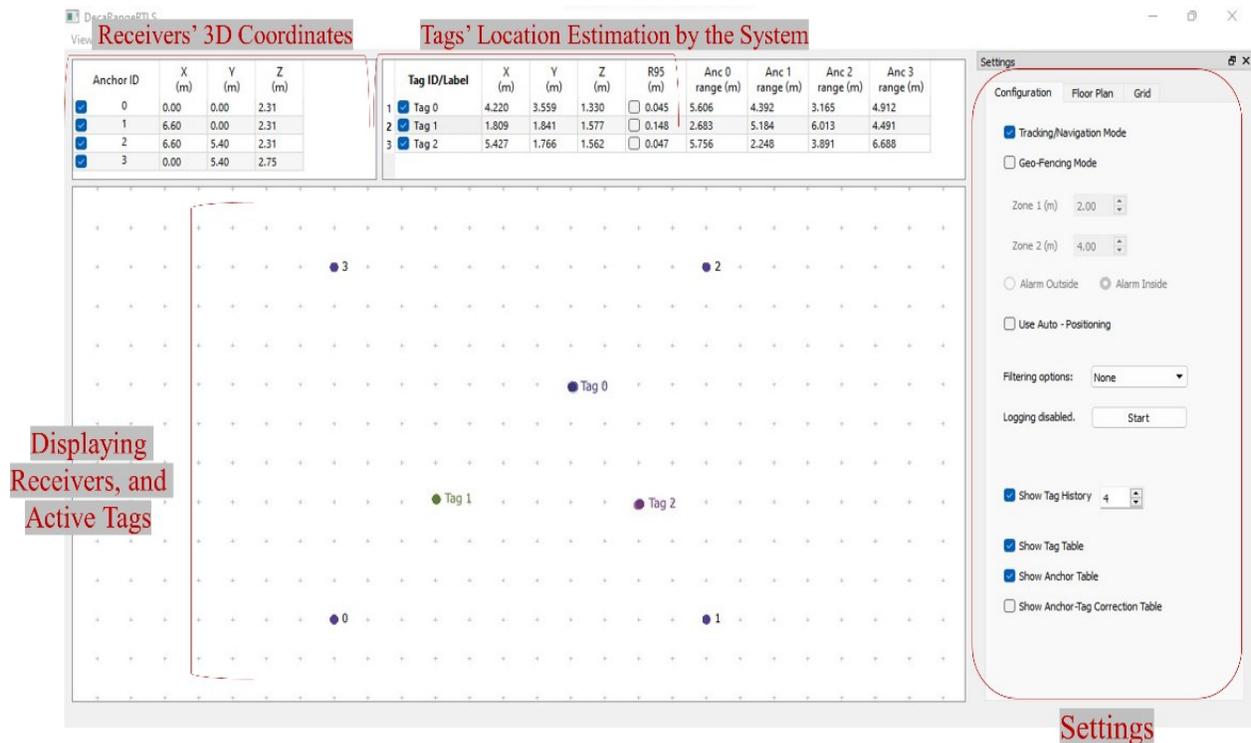


Figure 3-14. User interface of the UWB application.

3.4 Integrated Object Recognition and Localization

This section introduces an integrated method for automated recognition and localization of tracked components using digital cameras, deep learning models, and the UWB system. This method generates a comprehensive report about the status of tracked components in job sites, essential for

assessing their installation progress. It integrates data derived from YOLACT++ and the UWB system for each tracked element, then automatically reports each element's unique identifier (ID), location, visual data and capture time. The overview of the integrated method is depicted in Figure 3-15.

For this method, YOLACT++ was selected as the deep learning model due to its capabilities in recognizing elements under different visual configurations. Following the model's prediction results, each predicted image is stored into distinct folders on the local computer based on its identified object class. Although the YOLACT++ model is selected, other deep learning models could also be applied. Furthermore, the integrated model is not limited to UWB systems alone and can be generalized to different types of RTLS.

This method is designed to mitigate the limitations associated with each individual technique. Specifically, object recognition algorithms, while capable of recognizing building elements, fail to provide precise location information of the elements in construction sites. Conversely, RTLS can localize building elements but lacks the visual data needed to validate proper installation in jobsites. Consequently, the integrated method developed in this study provides visual and location information about tracked components for a more accurate and timely understanding of the project's status. As compared to previous studies that integrate the vision-based techniques and RTLS, this integrated method is not limited to fixed calibrated cameras with restricted fields of view and ideal conditions such as minimum obstruction and occlusions within job sites. Instead, it fully utilizes each system capability to capture data from different areas of construction sites under different conditions. This flexibility is crucial for progress monitoring in indoor environments where a variety of tracked components are installed throughout the building.

It should be emphasized that the proposed method can be applied at different construction stages, from initiation to the final stage where all components are installed. However, as indoor construction progresses, the number of installed components increases, resulting in greater complexity due to higher obstructions.

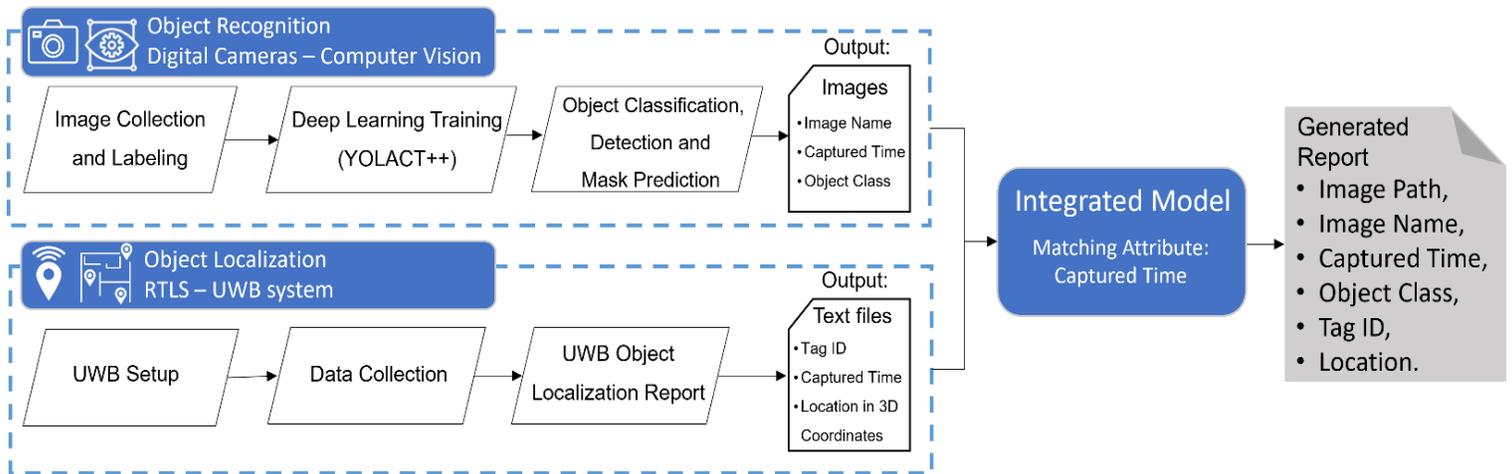


Figure 3-15. Overview of the integrated method.

3.4.1 The Integrated Model Process

The integrated model utilized an automated matching process to match each predicted image with the UWB data of the corresponding building component. The integration is based on the alignment of the “captured time” as the matching attribute derived from the timestamp of the UWB system (T_{tag}) and the “date taken” metadata from each image (T_{image}). After each successful matching process, all the required attributes from both sources related to each tracked component are extracted and gathered into a report. It is important to note that, for implementation of the integrated model, it is necessary to collect images and UWB data simultaneously. In addition, the tags from the same class could not be activated at once. Instead, once each component is installed in its location the tag will be activated for image capturing from the corresponding element and UWB data collection. Once image capturing was completed, the tag would be turned off, removed, renumbered, and attached to other building component that is set to be installed. This procedure is repeated for each subsequent component.

As illustrated in Figure 3-16, the model works based on three main input files including the predicted images by YOLOACT++, the UWB reports and a list of classified tags where each tag ID is assigned to a corresponding predefined YOLOACT++ class. The matching approach processes each object class individually and only tags and images associated with that class are examined. After the first class is selected, the first image within the class is identified, and its captured time is determined. The model then searches the UWB report for the timestamp of Tag ID associated

with the same class. It looks for the closest timestamp match between the image's and tag's captured time based on the absolute value of the time difference within a 90-second margin. If a match is found, the model collects the required attributes from both sources for that element for the final report. The 90-second margin is considered for the duration the tag remains active for data collection before it is deactivated. This margin ensures that the data collection of the active tag aligns with the image capture time and adds a limit on the model's search range within the UWB report. The margin value is adjustable based on the site conditions or UWB system configurations.

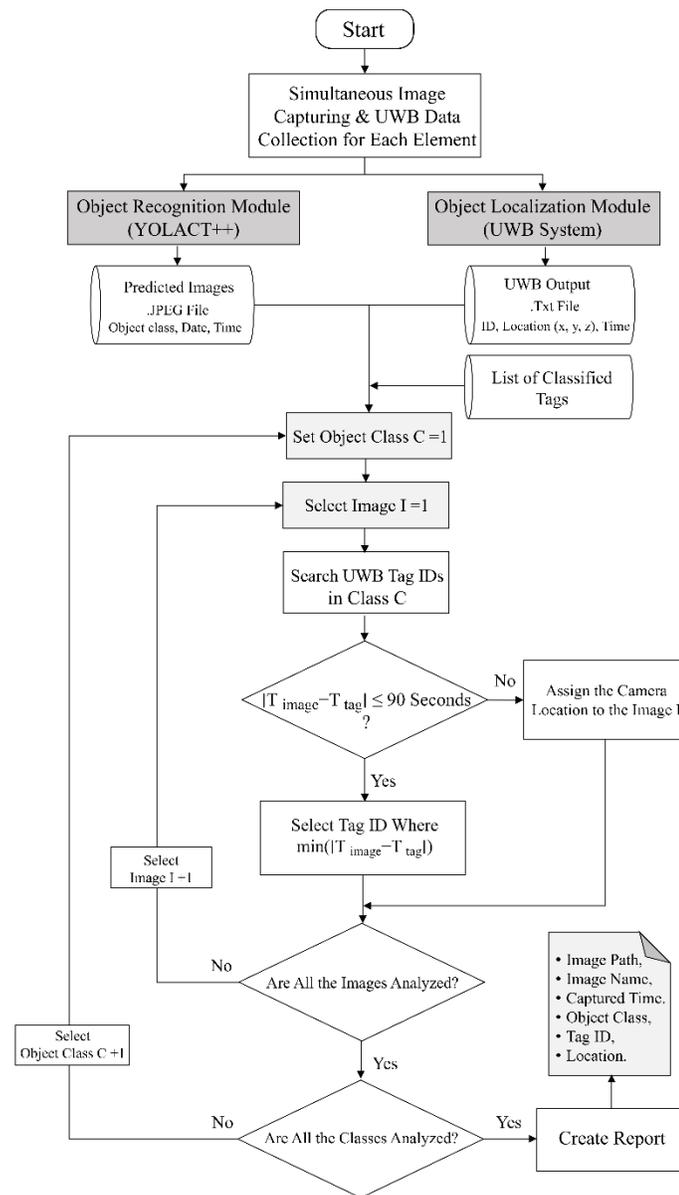


Figure 3-16. Steps of the integrated framework.

Furthermore, the model was considered for an optional complementary solution when no valid tag ID match is found in the UWB output. During the experiments, a tag can be fixed to the camera which captures images. If the system failed to find any valid tag ID corresponding to an image's captured time, the camera's location at that time would be used as an approximate location of the captured image. Following the analysis of all images and UWB datapoints of the same object class, the process will be repeated with the next object class until the last class has been analyzed to generate the final report. The report contains required information about the status of each tracked element. The information from each image includes its image name and its path on the local computer to enable faster access to image, its captured date and time and for the UWB tag include its specific ID, timestamp and 3D coordinates. These reports provide project managers with timely updates that can facilitate the decision making process.

Accurate visual and 3D location data enable precise tracking of elements at construction sites which enhances automated construction progress monitoring and reduces errors associated with manual data collection, correlation and reporting at jobsites. It also improves resource management and minimizes delays by timely identifying installed elements and any misplacements or incorrect installations. Additionally, the integration process does not require complex procedures, extensive mathematical computations, or specific requirements for image capturing, such as fixed calibrated cameras. These capabilities make the method easily applicable in construction projects, reducing manual efforts and the need for professional expertise.

3.5 Summary and Conclusions

In this chapter, an overview of the research methodology was elaborated to achieve the objectives of this research study. In summary, the main purpose of this study is the development of a method for automated construction progress monitoring systems in indoor environments. This method is capable of collecting, analyzing, and reporting essential information from tracked elements accurately. This helps project managers to identify the current status of the project and apply correct decisions in a timely manner. Unlike most research studies that focus on monitoring outdoor sites or indoor environments under ideal conditions, this research has considered challenging indoor environments containing complex elements installed across all the building.

The methodology includes three main modules. The first module is object recognition, which uses digital imaging and deep learning models. In this module, Mask R-CNN and YOLACT++

which are two novel instance segmentation algorithms were selected for model training. Instance segmentation algorithms are the most comprehensive algorithms in object recognition domain, due to their capability to perform classification, detection, and mask prediction of instances simultaneously. These algorithms were considered to automatically obtain visual data from tracked elements and recognize them from images collected at indoor job sites.

The second module is object localization which the UWB system was selected for the automated identification and localization of tracked elements. This system is capable of accurate 2D/3D object localization in real-time, even in high-multipath indoor environments. In this study, different factors that may affect the UWB system are considered during laboratory and field experiments. These include metallic objects, number of tags and receivers, LOS/NLOS conditions, Tags' heights, etc. The last module is the integrated object recognition and localization, in which the data derived from the YOLACT++ and the UWB system for each tracked element are integrated based on a matching process. After the completion of the integration process, a report is automatically generated in a comprehensive and structured manner. This report provides all the required information from tracked elements that cannot be obtained from a single technology.

These three modules are implemented and validated in Chapter 4.

CHAPTER 4: MODEL IMPLEMENTATION and VALIDATION

In this chapter, the performance of the methods outlined in Chapter 3 is evaluated through different laboratory and field experiments to determine their applicability for automated progress monitoring in construction projects. This chapter is divided into three main sections. The first section evaluates the performance of two deep learning-based instance segmentation algorithms, Mask R-CNN and YOLACT++, through different datasets and model configurations. The second section validates the performance of the UWB system for object localization in indoor environments and identifies the factors that affect its localization accuracy. Finally, the third section evaluates the developed integrated model in providing comprehensive information about each tracked object through a detailed report.

This research focuses on the Mechanical, Electrical, and Plumbing (MEP) components such as HVAC ducts and pipes in indoor construction environments to monitor their installation progress. MEP works are important indoor activities that can considerably affect project delays and cost overruns (Akhil & Das, 2019; Shekhar et al., 2021). MEP systems contain complex components, such as HVAC ducts and pipes, which are typically installed in confined spaces in built facilities, often at ceiling-level heights and in close vicinity to architectural and structural components. The installation of these systems is difficult and time-consuming resulting in reworks and project delays (Korman et al., 2003; Teo et al., 2022). Hence, efficient monitoring of MEP activities is essential to timely identify deviations between planned and as-built states, allowing the implementation of remedial measures prior to costly and complicated reworks (Bosché et al., 2015; Navon & Shpatnitsky, 2005).

4.1 Automated Recognition of MEP Components in Indoor Job Sites.

4.1.1 Overview

To create high-quality datasets, both real and virtual synthetic images were used for model training. The real images were manually captured by smartphones at two indoor construction sites, both hospital projects, in Montreal, Canada, and in Tehran, Iran. The set of Montreal based images were collected earlier by Golkhoo (2020). These sites were mostly illuminated by artificial lighting and in a few cases by indirect natural light. The image resolutions were 3024 * 4032 and 2448 * 3264 pixels, containing HVAC ducts and pipes with different shapes, sizes, and orientations. Virtual synthetic images were also generated using 3D BIM models, containing MEP components.

Two publicly available 3D BIM models with rich existing elements obtained from the National Institute of Building Sciences (NIBS) and Tools (2017) were used in this study. The synthetic image sizes were set to high-quality resolution with 1920*1080 pixels.

Figure 4-1 demonstrates the two downloaded BIM models containing MEP components. These models contain a rich network of HVAC ducts and pipes with different sizes and shapes. In the BIM model, unneeded components such as lighting fixtures and furniture were removed, and for remaining components, properties such as material specifications were selected. For generating synthetic images, the modifications mentioned in Section 3.2.1 were considered. The images were generated under different conditions including various lighting, complexity, scale, occlusion, clutter, and viewpoints. The Enscape plugin in Revit software was used to generate these images in real-time.

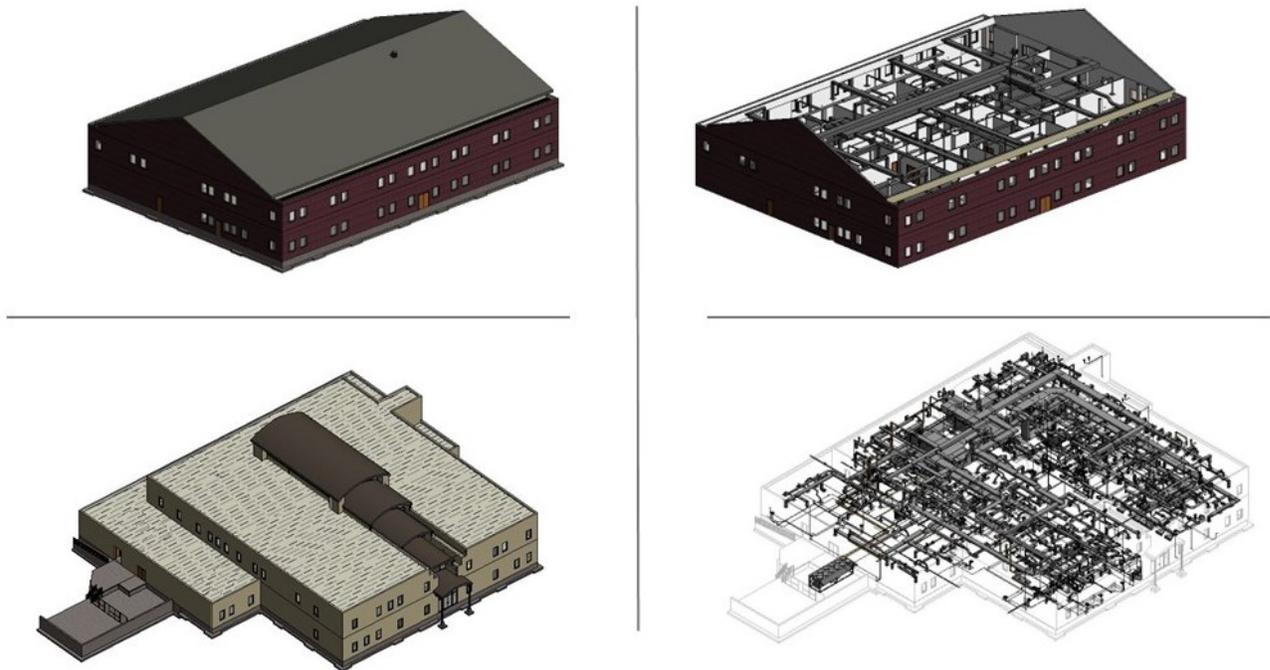


Figure 4-1. 3D BIM models and their MEP networks.

4.1.2 Object Recognition with Synthetic Image Datasets

In this section, to initially assess the object recognition model's performance, and the impact of data augmentation techniques, only synthetic images were used for model training. For this purpose, only HVAC ducts were selected to be monitored and selected as the only class.

A total of 1,143 synthetic images were generated and used for network training. The dataset consisted of 1,887 HVAC duct instances across the images. The training set distribution shows that from 1,143 images, 56% of images contain only one HVAC duct in each image, 32% have two ducts, 9% three ducts, 2% four ducts, and 1% five ducts. 172 synthetic images were randomly selected for testing and validation purposes. The test set data follows nearly the same distribution of the training; 54% of images having one HVAC duct, 44% having two ducts, and 2% having three ducts. After data labeling and specifying the regions of HVAC ducts within the images, the training and testing datasets with data labeling files were fed into the deep learning model.

In this section, Mask R-CNN which is a fundamental instance segmentation algorithm was selected with the model’s configurations stated in Section 3.2.2.1. To assess the impact of data augmentation techniques on the model’s performance, two experiments were conducted. In the first experiment (Experiment #1), the model was trained without any data augmentation techniques, and in the second experiment (Experiment #2), data augmentation techniques were applied. Details of these techniques are described in Table 3-1. The results of the experiments are summarized in Table 4-1. The performance of Experiment #2 with a precision value of 80.87% and a mAP score of 90.6% is better than the Experiment #1 with a precision value of 75.08% and mAP value of 88.69%. The F-1 score also indicates the same, confirming that Experiment #2 is superior.

In Experiment #1, overfitting was observed, but this issue was not observed in Experiment #2. Figure 4-2 illustrates the downward trend of the loss function during the training process in the Experiment #2 which shows the success of the model in preventing overfitting since there is a desired convergence of the training and validation errors. It can be stated that data augmentation improved the model’s performance and mitigated the overfitting problem. The output images of Mask R-CNN are depicted in Figure 4-3.

Table 4-1. Results of HVAC duct recognition with synthetic test dataset.

Training dataset	TP	FP	FN	Precision (%)	Recall (%)	F1-score (%)	mAP ^{bbox} (%)
Experiment #1	223	74	32	75.08	87.45	80.79	88.69

Experiment #2	224	53	31	80.87	87.84	84.21	90.60
---------------	-----	----	----	-------	-------	-------	-------

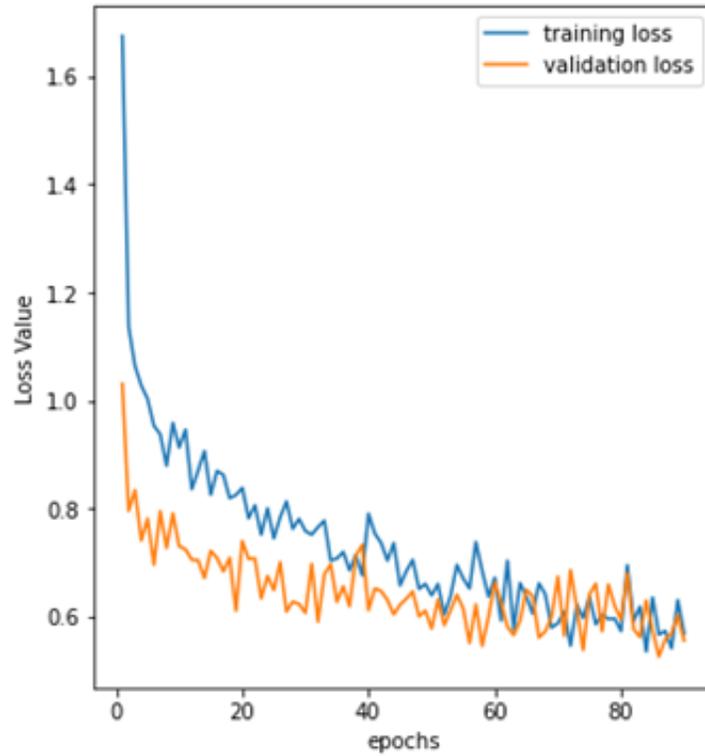


Figure 4-2. Training and validation losses during 90 epochs in the Experiment #2 (Shamsollahi et al., 2021)

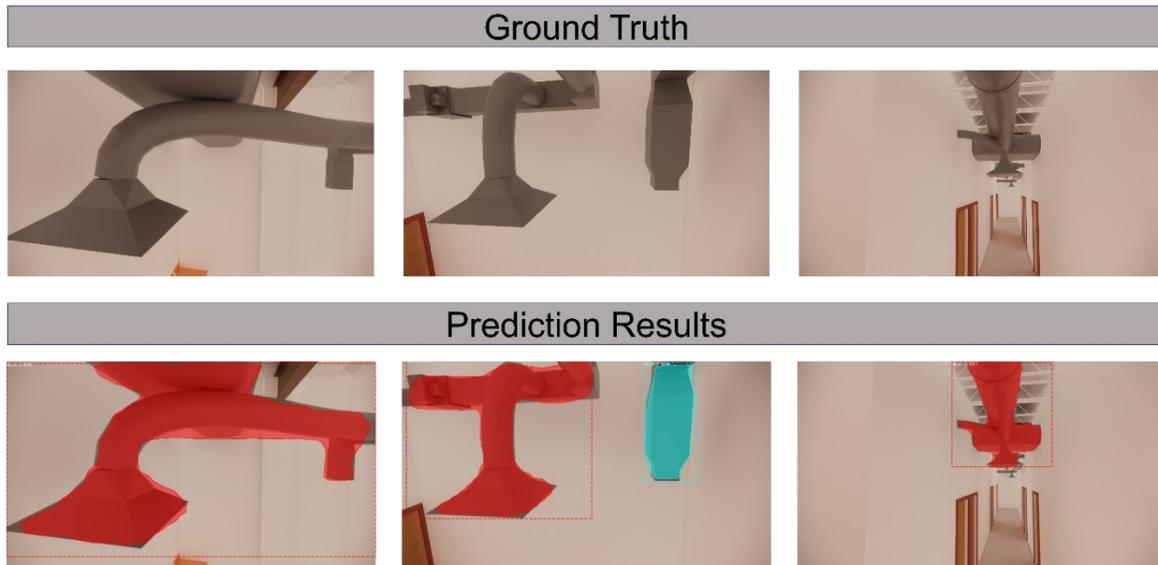


Figure 4-3. Results of the Mask R-CNN prediction.

To determine whether a model trained solely on synthetic images can accurately predict real images, 168 real images were added to the test dataset, bringing the total to 340 images. The model's performance significantly decreased, with a precision of 42.26%, a recall of 68.42% and a mAP^{bbox} of 48.0%. Based on the results, it can be concluded that training with synthetic images alone is insufficient for recognizing real HVAC ducts and pipes. To improve the recognition capabilities, real images from construction sites were added to the training dataset in the next section.

4.1.3 Object Recognition Using Mixed Synthetic and Real Images with Two Deep Learning Models

The objectives of this subsection are:

- Determine the optimum mix of synthetic and real images in a dataset for training models.
- Implement and validate Mask R-CNN and YOLACT++ and compare their performance to determine their applicability in construction projects.

In this study, 782 real images with the size of 3024*4032 pixels that contained HVAC ducts were added to the synthetic training dataset described in the previous subsection. In total, the dataset reached 1,925 images with 3,011 HVAC duct instances. The dataset distribution is as follows: 60% of the images contain only one duct, 29% have two ducts, 8% have three ducts, while

2% and 1% contain four ducts and five ducts, respectively. Figure 4-4 demonstrates sample of images within the dataset containing different numbers of HVAC duct instances. The total number of images in the test set is 340, consisting of 168 real images and 172 synthetic images.



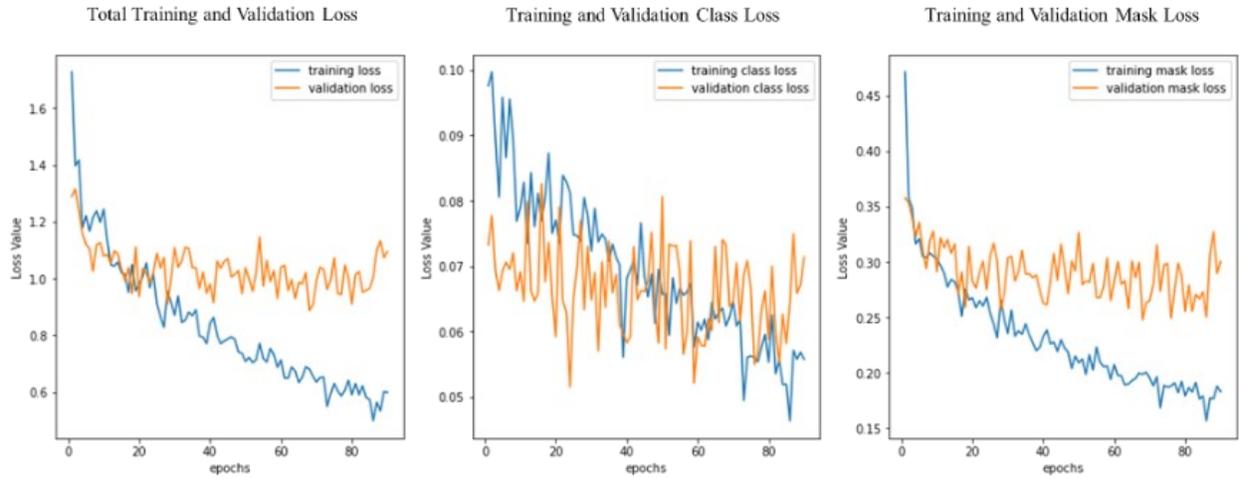
Figure 4-4. The presence of a different number of HVAC ducts in each image.

Eight datasets using both synthetic and real images were evaluated as detailed in Table 4-2. Across all the eight datasets, the number of real images remained fixed at 782 images, while the number of synthetic images decreased sequentially starting from the first dataset (Dataset #1).

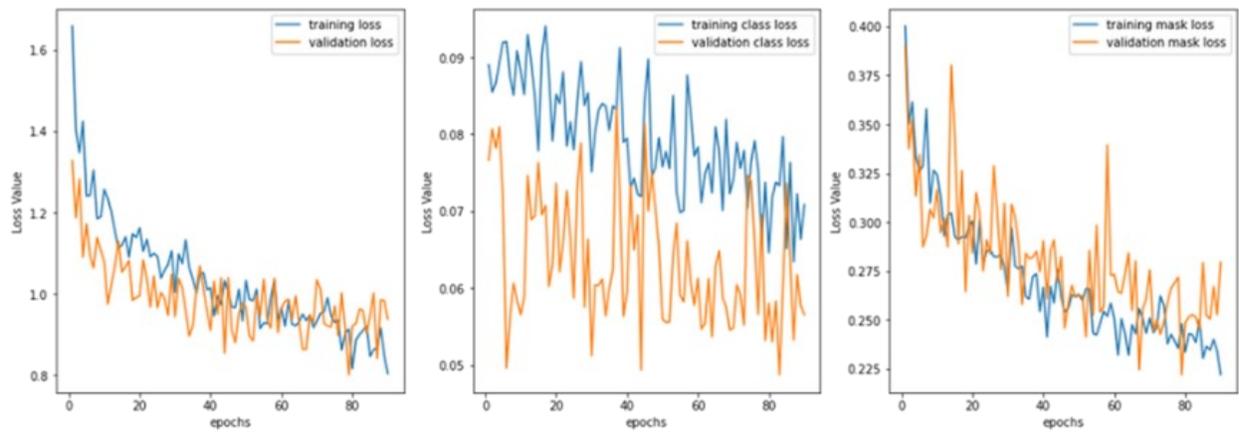
The evaluation of deep learning models began with Mask R-CNN. The model was evaluated with and without data augmentation techniques using Dataset #1. Similar to the results in the previous subsection, training the model without data augmentation techniques resulted in overfitting (Figure 4-5.a). This issue was resolved by using data augmentation techniques (Figure 4-5.b). Consequently, these techniques were applied across all eight datasets during training.

According to Table 4-2 the optimal result of the Mask R-CNN model was achieved using Dataset #6 containing 35% synthetic images and 65% real images. The model achieved a precision of 65.72% and a recall of 78.65%. This dataset performed better than Dataset #1, which had the highest number of synthetic images. This indicates that increasing the number of synthetic images in the dataset does not necessarily improve the model's performance.

The results from the Mask R-CNN model were not satisfactory for accurately recognizing HVAC ducts, especially in real images captured from construction sites. This was due to high occlusion levels, low lighting conditions, and different shapes and sizes of HVAC ducts. Specifically, the model had issues such as false positive and false negative detections, and inaccurate mask predictions as demonstrated in Figure 4-6.



(a)



(b)

Figure 4-5. Loss curves of Mask R-CNN: (a) Without data augmentation (b) With data augmentation.

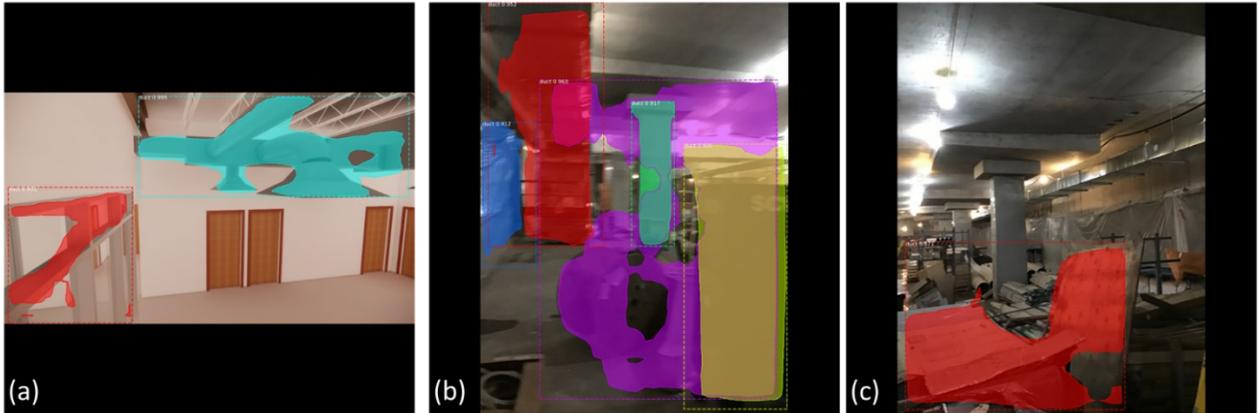


Figure 4-6. Examples of inaccurate recognition of objects by Mask R-CNN.

Consequently, YOLACT++ was trained with the same datasets. The model's configurations and the utilized data augmentation techniques are explained in Section 3.2.2.2. Similar to Mask R-CNN, YOLACT++ was trained up to epoch 90. According to Table 4-2, the best results for YOLACT++ were achieved from Dataset #5, which contained 40% synthetic images and achieved a precision of 81.25% and a recall of 82.20%. In terms of precision, it achieved 15.53% better performance compared to Mask R-CNN and its recall was 3.55% higher.

Regarding computational time, training the Mask R-CNN model took approximately 4 to 8 hours, while training YOLACT++ took between 21 to 24 hours, depending on the dataset's size. Also, the average prediction speed for 340 images was 558.20 seconds for Mask R-CNN and 68.14 seconds for YOLACT++. The prediction speed of YOLACT++ was eight times faster than Mask R-CNN in recognizing HVAC ducts, demonstrating its capability for near real-time applications.

Samples of predicted images by Mask R-CNN and YOLACT++ for HVAC duct recognition are presented in Figure 4-7. Through a comparative analysis, it was identified that YOLACT++ demonstrated better performance than Mask R-CNN in terms of object detection and mask prediction. YOLACT++ could detect HVAC ducts that Mask R-CNN failed to detect. Moreover, YOLACT++ generated predicted masks that are more aligned to the ground truth.

Table 4-2. The results of Mask R-CNN and YOLACT++ algorithms for HVAC duct recognition.

#	Training Dataset	Algorithm	Precision (%)	Recall (%)	F1-score (%)
1	782 Real images + 1,143 Synthetic images (Synthetic/Total \approx 60%)	Mask R-CNN	58.05	78.95	66.91
		YOLACT++	80.57	76.04	78.24
2	782 Real images + 956 Synthetic images (Synthetic/Total \approx 55%)	Mask R-CNN	52.01	76.74	62.00
		YOLACT++	78.44	76.63	77.52
3	782 Real images + 782 Synthetic images (Synthetic/Total \approx 50%)	Mask R-CNN	56.58	79.07	65.96
		YOLACT++	75.32	72.82	74.05
4	782 Real images + 640 Synthetic images (Synthetic/Total \approx 45%)	Mask R-CNN	63.25	78.06	69.88
		YOLACT++	77.36	72.22	74.70
5	782 Real images + 522 Synthetic images (Synthetic/Total \approx 40%)	Mask R-CNN	60.46	78.06	68.14
		YOLACT++	81.25	82.20	81.72
6	782 Real images + 422 Synthetic images (Synthetic/Total \approx 35%)	Mask R-CNN	65.72	78.65	71.61
		YOLACT++	77.77	73.29	71.61
7	782 Real images + 336 Synthetic images (Synthetic/Total \approx 30%)	Mask R-CNN	60.10	75.16	66.79
		YOLACT++	76.38	75.73	76.05
8	782 Real images + 260 Synthetic images (Synthetic/Total \approx 25%)	Mask R-CNN	64.77	76.79	70.27
		YOLACT++	72.25	70.92	71.58

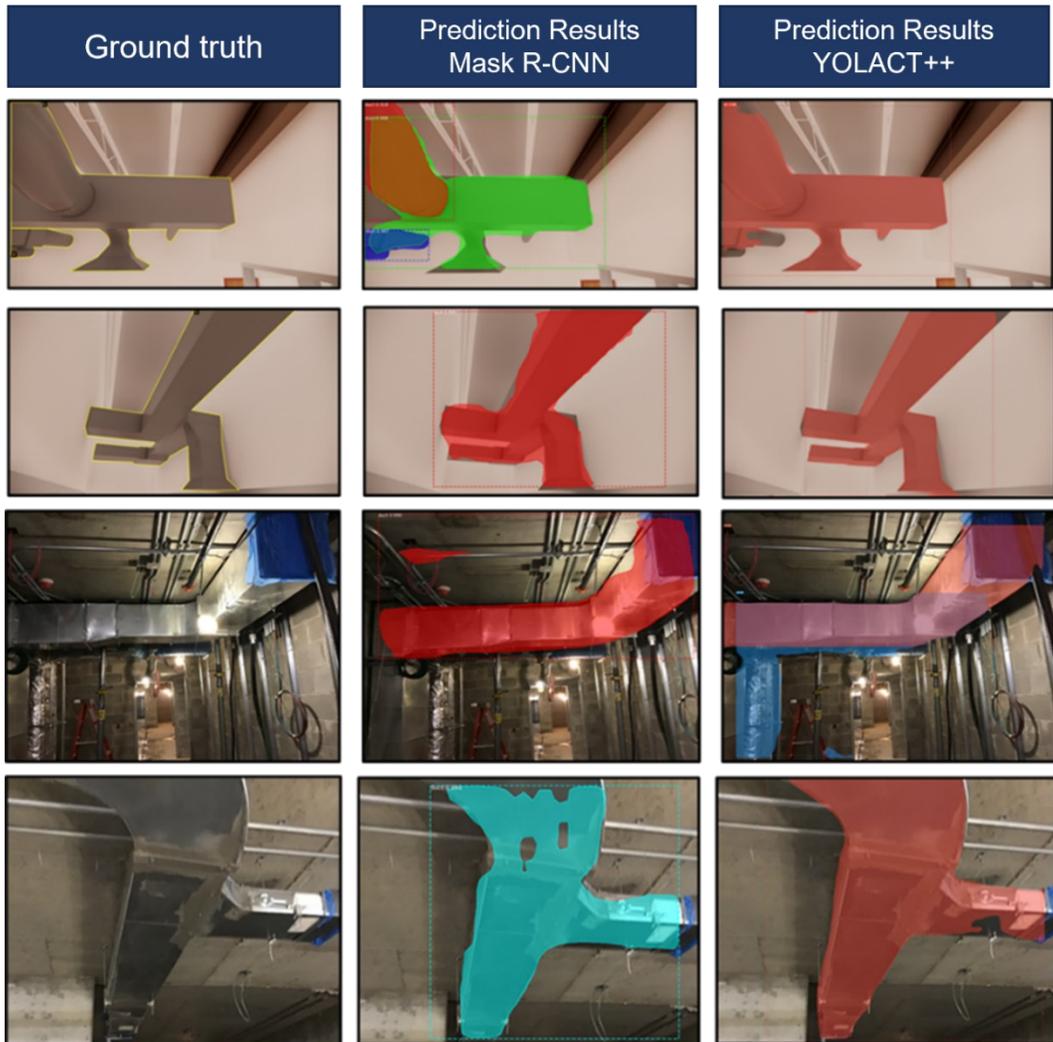


Figure 4-7. Output images from Mask R-CNN and YOLACT++ models.

4.1.4 MEP Recognition using YOLACT++

This subsection is a modified version of “Automated Detection and Segmentation of Mechanical, Electrical, and Plumbing (MEP) Components in Indoor Environments by Using the YOLACT++ Architecture” published by *Journal of Construction Engineering and Management* (Shamsollahi et al. 2024).

The objective of this section is to enhance YOLACT++'s generalizability for recognizing more building elements after identifying its superior performance over Mask R-CNN. In order to incorporate more MEP components, the number of object classes has been set to two, including

HVAC ducts and pipes. The number of real and synthetic images in the dataset increased to include both pipes and HVAC ducts. The training dataset reached 3,135 images containing 1,881 synthetic and 1,254 real images. From 1,881 synthetic images, 1,143 included only HVAC ducts, 451 included only pipes, and 287 images had both pipes and HVAC ducts. Also, from 1,254 real images 782 contained HVAC ducts and 472 images had pipes. Samples of real and synthetic images are demonstrated in Figure 4-8.

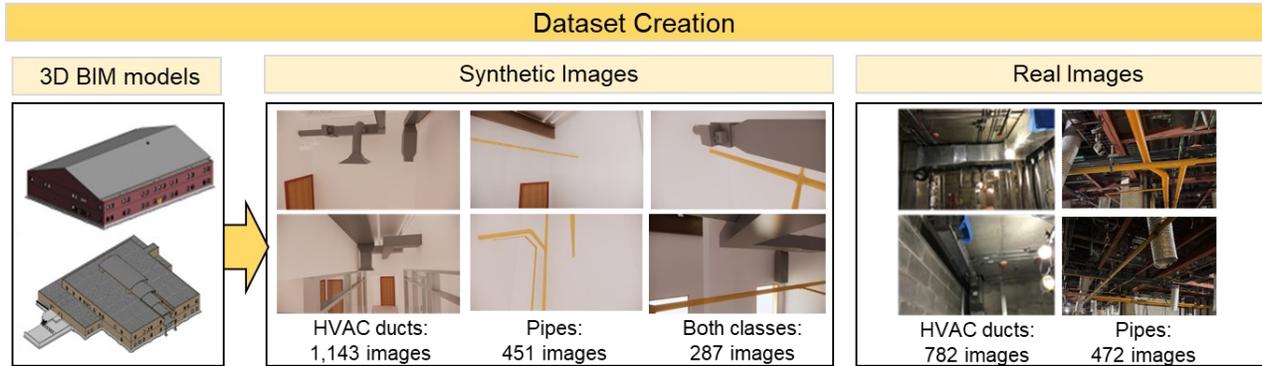


Figure 4-8. The types and numbers of images within the dataset.

Similar to the previous subsection, eight training sets were created by decreasing the number of synthetic images from 60% of total images to 25%, while the number of real images was fixed. Table 4-3 presents the number of synthetic and real images along with the number of instances in each dataset. The validation set included 497 images including 229 synthetic images and 268 real images. The configuration of YOLACT++ was set to epoch 116 and batch size of 8.

Table 4-3. Distribution of synthetic and real images in each dataset and the number of instances.

Dataset	Synthetic/Total (\approx)	Number of Synthetic Images	Number of Real Images	Number of Instances
Dataset #1	60%	1881	1254	8533
Dataset #2	55%	1533	1254	7351
Dataset #3	50%	1254	1254	6674
Dataset #4	45%	1026	1254	6019
Dataset #5	40%	836	1254	5562
Dataset #6	35%	676	1254	5158
Dataset #7	30%	538	1254	4812
Dataset #8	25%	418	1254	4507

YOLACT++ was trained using the eight datasets and its performance was assessed with the test dataset. The results are summarized in Table 4-4. The overall performance of the YOLACT++ in object detection and mask prediction across these datasets is shown in Figure 4-9.

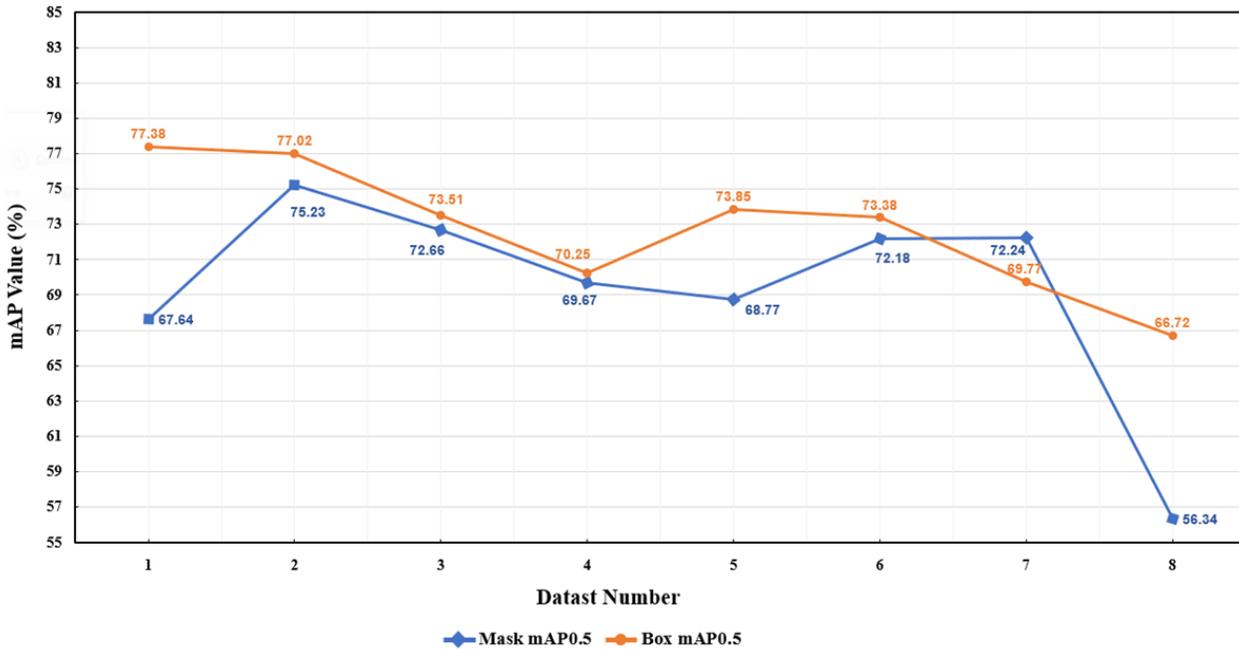


Figure 4-9. Comparison of bounding box and mask mAP values across different datasets (Shamsollahi et al. 2024).

Analysis of the results reveals that in most cases (7 out of 8 datasets), the mAP^{mask} was lower than the mAP^{bbox} , highlighting a greater challenge in predicting the instance regions compared to bounding box prediction. Dataset #1, with the highest number of synthetic images (60%) including 6501 HVAC duct and 2032 pipe instances achieved the highest performance in object detection, with a mAP^{bbox} of 77.37%. Specifically, in detecting HVAC ducts, the model achieved a precision of 84.80% and a recall of 85.58%. For pipes, precision and recall were 86.87% and 73.93%, respectively.

However, Dataset #1 could not achieve the best performance in mask prediction. Dataset #2, with 55% synthetic images, showed slightly lower detection performance with a mAP^{bbox} of 77.03% but outperformed in mask prediction with a mAP^{mask} of 75.23%. In terms of per-class performance, Dataset #2 also achieved the highest AP^{mask} at 69.81% for pipe class. For HVAC ducts, its AP^{mask} was close to the highest at 80.65%.

Considering both detection and mask prediction criteria, Dataset #2 was identified as the most preferred, despite of Dataset #1's higher detection rates. Conversely, Dataset #8 demonstrated the lowest performance in both detection and mask prediction. Particularly, for HVAC duct detection, it obtained the lowest precision and recall rates at 72.49% and 79.18%, respectively. In pipe detection, this dataset had the lowest recall at 63.35%, and its precision, while not the lowest, was still comparatively low. Moreover, for mask prediction, Dataset #8 had the lowest performance among the others with mAP^{mask} at 56.34%.

The YOLACT++ model had superior performance in both detection and mask prediction of HVAC ducts compared to pipes with a better balance between precision and recall. This lower performance in pipe recognition can be attributed to the fewer pipe instances in the training datasets relative to HVAC duct instances. Moreover, the unique geometric properties of pipes being of slenderical shapes may present challenges for the model in accurately recognizing pipes. A notable observation from Figure 4-9 is the overall increase in the mAP^{bbox} values corresponding with the rise in the number of synthetic images within the datasets.

The above suggests that increasing the synthetic images enhances the model's object detection ability. However, this increase in synthetic images does not have the same positive impact on mask prediction. This may be attributed to the inherent complexity in determining the region of instances in mask prediction which requires further investigation. Regarding the prediction speed of YOLACT++, for processing 497 images, the average frames per second (FPS) was between 4.69 to 4.91 across different datasets.

This study assessed the effect of synthetic image quantity on recognition of objects from real images. Hence, Dataset #2 with 55% synthetic images and Dataset #8, with the least synthetic images at a rate of 25%, were evaluated, with the sample results depicted in Figure 4-10. The model, trained with Dataset #2, performed better in challenging indoor conditions such as low lighting, clutter, and occlusion, showing enhanced recognition of instances across various scales, shapes, and poses. For instance, Figure 4-10(c-1) and (c-2) demonstrate the model's improved recognition of HVAC ducts with different scales trained on Dataset #2 which was not achieved with Dataset #8 Figure 4-10(b-1) and (b-2). Figure 4-10(c-3) also showcased that using Dataset #2, the model could accurately recognize HVAC ducts with different shapes and scales in a cluttered and low-lit scene. Conversely, the model trained with Dataset #8 Figure 4-10(b-3),

displayed weaker performance, inaccurately predicting the bounding box and a mask that covered nearly the entire image. The model's ability enhanced with Dataset #2 in distinguishing target instances from similar non-target ones. For example, in Figure 4-10(c-4), it accurately differentiated HVAC ducts from a similarly shaped and sized column, unlike Dataset #8, Figure 4-10(b-4), which misidentified a column as a duct.

Moreover, Dataset #2's training resulted in enhanced recognition of slender objects, as shown in Figure 4-10(c-5), where it accurately detected and segmented a pipe even under low light conditions. In contrast, Dataset #8 had incorrectness, including incomplete masks, and a false positive for the non-present duct Figure 4-10(b-5). Lastly, Dataset #2 Figure 4-10(c-6) improved the identification of multiple varied-scale instances from both classes in high-density, cluttered scenes. However, Dataset #8 Figure 4-10(b-6) had false negatives, missing some HVAC ducts and pipes.

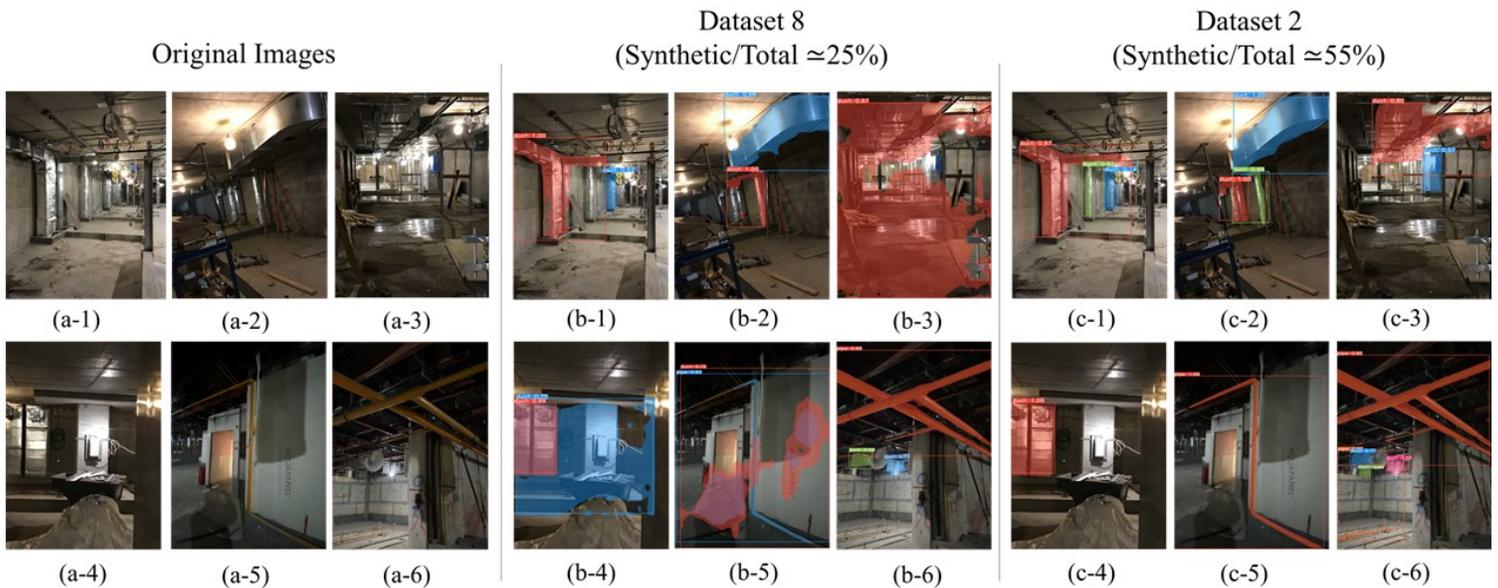


Figure 4-10. Output results of YOLACT++ trained on two different datasets (Shamsollahi et al. 2024).

YOLACT++ efficiently performs detection and segmentation, offering faster predictions than algorithms like Mask R-CNN (Bolya et al., 2020). The resulting higher prediction speed makes YOLACT++ more suitable for integration with monitoring and reporting systems in order to provide project managers with timely needed information. Moreover, YOLACT++ backbone

network utilizes a DCNv2, which improves the model’s ability to handle objects of different scales and poses through non-grid sampling capability. This feature is aligned with the objectives of this research, which involves instances with different geometric variations. The outcomes of this section illustrate that the YOLACT++ model, when trained with a balanced dataset of synthetic and real images achieved promising performance in the recognition of HVAC ducts and pipes in indoor construction sites, especially in challenging locations where distinguishing them from surrounding objects is difficult.

The above highlights the significance of generating synthetic images that closely reflect real indoor construction sites to enhance the model’s performance abilities. Analysis of the results reveals that the balanced dataset can increase mAP^{bbox} by over 10% and mAP^{mask} by more than 18%. However, it is important to note that a dataset with the highest proportion of synthetic images do not necessarily has the optimal results in both mask prediction and object detection. The results highlight the need for further research to identify the ideal balance of synthetic and real images in training datasets to maximize performance across all evaluation criteria.

Table 4-4. The results of YOLACT++ with two object classes

Dataset	Object Class	Precision	Recall	F1-score	AP_{50}^{mask}	AP_{50}^{bbox}	FPS
Dataset #1	HVAC Duct	84.80	85.58	85.19	78.82	82.78	4.69
	Pipe	86.87	73.93	79.88	56.46	71.97	
Dataset #2	HVAC Duct	81.79	85.65	83.68	80.65	82.21	4.72
	Pipe	85.98	73.77	79.41	69.81	71.83	
Dataset #3	HVAC Duct	82.25	83.15	82.70	80.34	81.17	4.89
	Pipe	75.58	68.06	71.62	64.99	65.85	
Dataset #4	HVAC Duct	81.85	82.75	82.30	76.53	80.30	4.82
	Pipe	70.78	63.63	67.01	62.80	60.20	
Dataset #5	HVAC Duct	79.10	86.09	82.45	82.30	82.27	4.91
	Pipe	82.89	67.02	74.11	55.24	65.43	
Dataset #6	HVAC Duct	79.74	85.00	82.29	80.84	82.41	4.73
	Pipe	76.92	67.01	71.62	63.53	64.35	
Dataset #7	HVAC Duct	78.87	81.95	80.38	78.92	78.79	4.80

	Pipe	73.37	63.91	68.31	65.56	60.76	
Dataset #8	HVAC Duct	72.49	79.18	75.69	65.19	72.89	4.70
	Pipe	75.15	63.35	68.75	56.34	60.56	

While the YOLACT++ model successfully recognized HVAC ducts and pipes in many images, it also had some failures that need to be improved in the future. For instance, in Figure 4-11(a), the model failed to recognize all the HVAC ducts present in the image. In Figure 4-11(b), despite accurately detecting the pipes, the model was inefficient in segmenting all the pixels assigned to the pipe class. In Figure 4-11(c), the model could not differentiate between HVAC ducts and areas of the ceiling and walls that had similar color and texture. The model's prediction speed reached around 5 FPS, lower than the over 30 FPS, stated in the original paper (Bolya et al., 2020) with the COCO test datasets. This reduced speed is primarily due to the absence of high-performance computing systems like GPUs, often not provided by cloud computing systems.

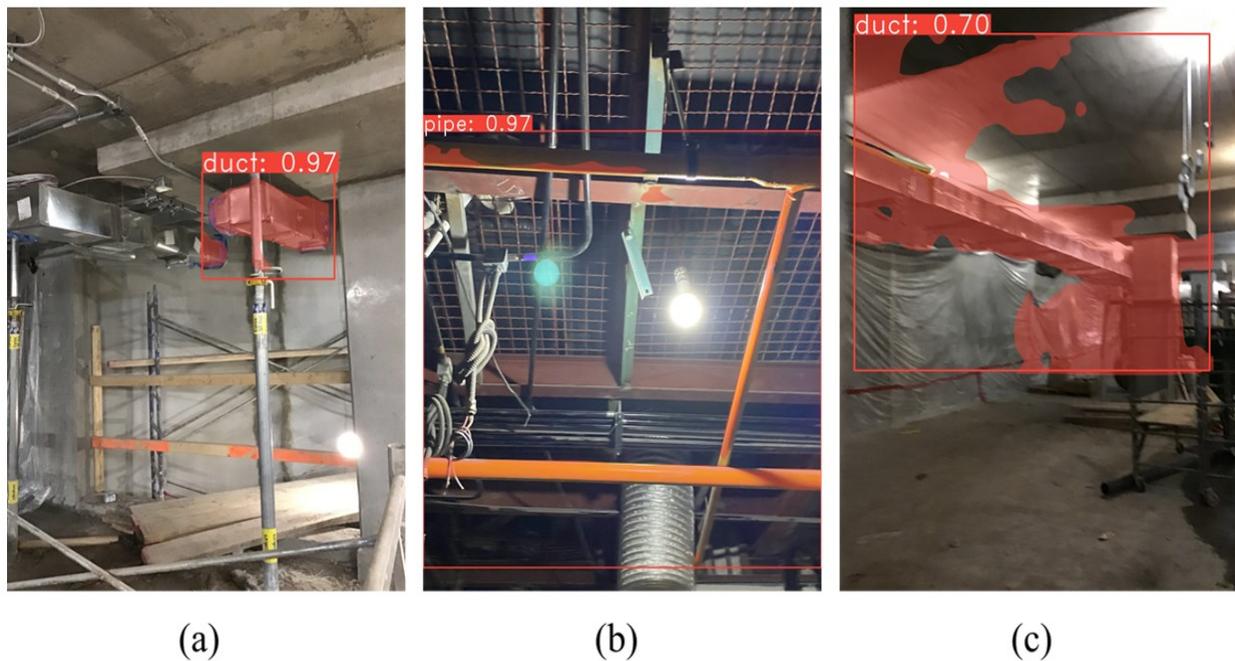


Figure 4-11. Examples of failures in the MEP recognition using YOLACT++ (Shamsollahi et al. 2024).

4.1.5 Summary and Conclusions

Indoor construction environments, with their high density and detailed components and low-lit conditions, are complex areas for identifying the installation statuses of components for progress monitoring and reporting. Addressing this challenge requires the automated and timely recognition of these components within indoor environments.

This section implemented a novel method for monitoring of construction operations, employing digital imaging and deep learning-based instance segmentation algorithms to automatically recognize MEP components in challenging indoor settings. Instance segmentation algorithms, in addition to detection, further specify regions of each instance by classifying pixels within the images (Athira & Khan, 2020; Hao et al., 2020; Shamsollahi et al., 2022; Wu et al., 2020). The model's capabilities for detecting and segmenting tracked elements play a complementary role in enhancing the overall performance.

To enhance the dataset quality for model training and overcome the lack of real images from construction sites, this research generated virtual synthetic images using BIM models and a rendering plugin. To reflect real indoor scenarios, different image modifications were considered during image generation to improve the model's prediction performance. The image modifications included different lighting conditions, complexity of components, scale, degrees of occlusion and clutter, and viewpoints. Synthetic images were mixed with real images at various ratios across eight datasets to determine the optimum mix for model training. For these datasets, the number of real images was fixed, but the proportion of synthetic images varied from 60% to 25% of the total images. This integration provides a wide range of images with different visual conditions. To improve the models' generalizability and adaptability, data augmentation techniques and transfer learning were applied.

The selection of an appropriate instance segmentation algorithm initially involved evaluating the Mask R-CNN through various tests for recognizing HVAC ducts. This study investigated the effect of data augmentation on model performance and found that it improved performance and resolved the overfitting problem. However, the Mask R-CNN model did not achieve acceptable recognition performance. The best result for the model was a precision of 65.72% and a recall of 78.65%. Consequently, YOLACT++ was chosen for this study and was trained on the same eight

developed datasets. Through a comparative analysis, it was determined that YOLACT++ demonstrated better performance than Mask R-CNN in terms of object recognition and prediction speed. The best performance of YOLACT++ was achieved with a precision of 81.25% and a recall of 82.20%. Additionally, YOLACT++ predicted 340 images in 68.14 seconds, which is eight times faster than Mask R-CNN.

Since YOLACT++ demonstrated better results, it was used for the final validation process. It was trained using two object classes, including both HVAC ducts and pipes. Similar to previous tests, mixes of synthetic and real datasets were created for training. A total of 497 images that included HVAC ducts and pipes were utilized for model evaluation. The results illustrate that the dataset with 55% synthetic images, exhibited a more balanced performance when evaluating both detection and mask prediction metrics. The detection performance showed 81.79% precision and 85.65% recall for HVAC ducts; 85.98% precision and 73.77% recall for pipes, and with a mAP^{mask} of 75.23%. The results demonstrated that YOLACT++ can recognize complex objects in challenging indoor environments.

4.2 Object Localization using the UWB System

4.2.1 Overview

To evaluate the performance of the UWB system for 2D and 3D localization of tracked MEP objects within indoor job sites and understand the applicability of the system for progress monitoring and reporting, four laboratory and three field experiments were conducted. More details about each experiment are described in the following subsections.

This section and Section 4.3 are extended versions of “Data Integration Method Using a Deep Learning Algorithm and Real-Time Locating System (RTLS) for Automated Construction Progress Monitoring and Reporting” published by *Automation in Construction* (Shamsollahi, et al., 2024).

4.2.2 Laboratory Experiments under Line-of-Sight (LOS) Conditions

For laboratory experiments, Construction Automation Lab on the 11th floor of the EV building at Concordia University was selected to implement the UWB system. For the layout of the system, a rectangular shape was selected and a test area with a dimension of 6.60*5.40 meters was determined to perform the experiments in the test area. Based on the manufacturer’s

recommendations, in these experiments four receivers were installed, three at the same height level and the fourth one higher than the others. Figure 4-12 illustrates the experiment area, the layout plan with its dimensions, and the 3D coordinates of each receiver.

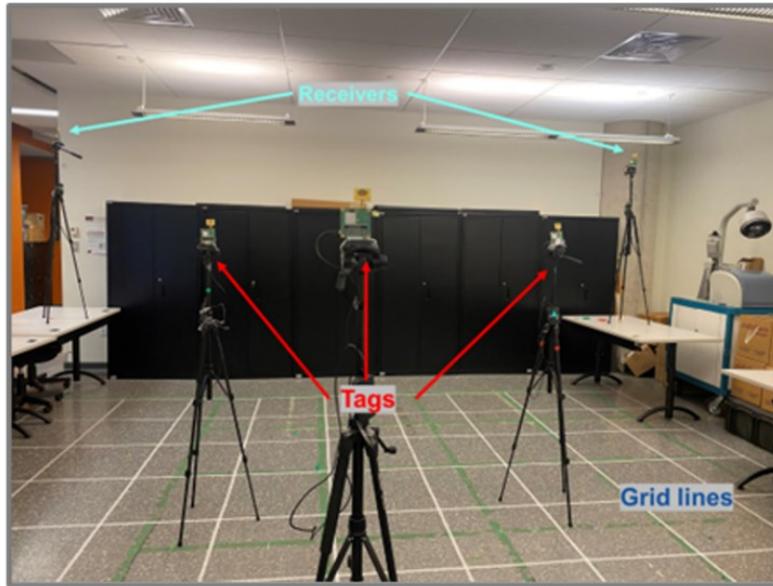
The experiments were conducted in a controlled test area under line-of-sight conditions. In this area, no obstacle existed between the tags and receivers to obstruct the direct path between them. These experiments enabled to find the optimum performance of the UWB system for 2D and 3D localization. Furthermore, they allow for accurate observation and analysis of the UWB localization performance before implementation in real work environments.

For the system installation, four receivers were fixed to tripods using tapes and fixed in the corners of the rectangular area as depicted in Figure 4-12. Instead of placing the tripods on the ground, they were placed on tables to achieve the highest possible position for the receivers. The first three receivers (A₀, A₁, A₂) had an equal height of 2.31 meters and the fourth one (A₃) had a height of 2.75 meters. For the fourth receiver, 2.75 meters was the maximum height to maintain 0.15 meters with the ceiling. To identify the actual locations of receivers and tags within the test area, a grid layout procedure was implemented. This procedure involved creating grid lines on the floor of the test area using adhesive tape, with the grid lines set at 0.6 meters apart.

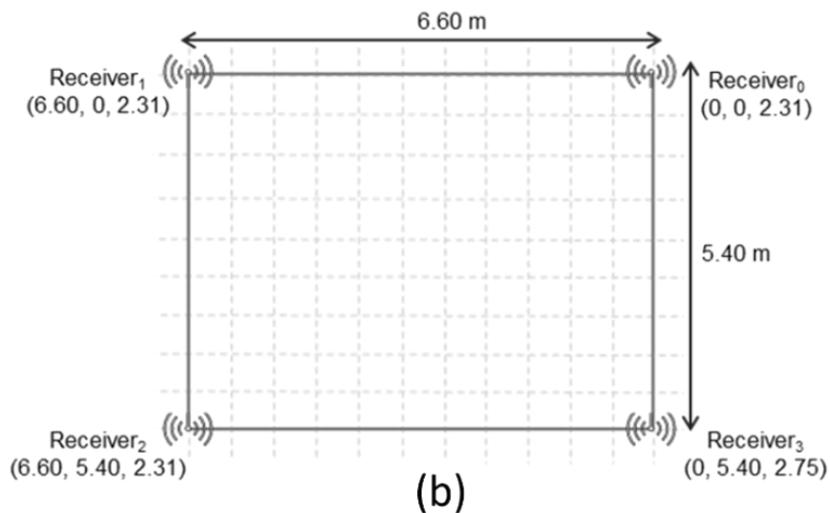
Four laboratory experiments were conducted in the test area under different scenarios to assess the impact of tags' height and tag proximity to receivers on the UWB system's performance. The following elements were repeated in all of the four experiments:

- Three static tags were used within the test area for data collection.
- All the tags were fixed on tripods for the tags' safety and easy height adjustment.
- The duration of each experiment was set to five minutes, ensuring uniformity in the collected data.
- The locations of the receivers were the same in the experiments.
- The first receiver (A₀) was assigned as the origin point and set to (0,0) in 2D coordinates.
- The first receiver (A₀) was connected to the laptop by a USB cable for powering and execution of the UWB application for data acquisition.
- The remaining receivers and tags are powered by USB batteries.

Table 4-5 demonstrates the total number of data points collected in each experiment, along with the number of data points captured from each individual tag. The number of data points in the four experiments was close to each other and ranged between 3,054 and 3,130 data points. Table 4-6 provides each tag's true location, its UWB averaged location, and the corresponding standard deviations in meters.



(a)



(b)

Figure 4-12. (a) Lab area (b) Layout plan and location of the receivers.

Table 4-5. Collected datapoints in each experiment.

	Experiment #1	Experiment #2	Experiment #3	Experiment #4
Tag-0	1,067	1,088	1,080	1,073
Tag-1	920	942	971	963
Tag-2	1,067	1,090	1,079	1,068
Total Data points	3,054	3,120	3,130	3,104

Table 4-6. True location of tags, mean and standard deviation of UWB data.

Experiment	Tag's name	True Position (m)			UWB Average Position (m)			Standard Deviation (m)		
		X (m)	Y (m)	Z (m)	X (m)	Y (m)	Z (m)	X (m)	Y (m)	Z (m)
1	Tag-0	4.20	3.60	1.57	4.19	3.61	1.47	0.018	0.017	0.017
	Tag-1	1.80	1.80	1.57	1.80	1.83	1.46	0.015	0.021	0.021
	Tag-2	5.40	1.80	1.57	5.42	1.82	1.48	0.021	0.020	0.020
2	Tag-0	4.20	3.60	0.50	4.29	3.68	0.31	0.024	0.016	0.184
	Tag-1	1.80	1.80	1.57	1.83	1.87	1.46	0.012	0.034	0.056
	Tag-2	5.40	1.80	2.50	5.39	1.86	2.67	0.037	0.035	0.282
3	Tag-0	1.20	4.20	1.57	1.25	4.13	3.36	0.019	0.306	1.103
	Tag-1	5.40	4.20	1.57	5.48	4.24	1.43	0.050	0.220	0.098
	Tag-2	1.20	1.20	1.57	1.21	1.24	1.59	0.015	0.029	0.054
4	Tag-0	1.20	4.20	0.50	1.29	4.26	0.40	0.034	0.041	0.211
	Tag-1	5.40	4.20	1.57	5.51	4.21	1.65	0.028	0.011	0.678
	Tag-2	1.20	1.20	2.50	1.23	1.28	3.16	0.015	0.026	0.089

- Experiment #1 – Tags at the same height level.

In this experiment, it was aimed to design an ideal scenario to be able to find the optimum performance of the UWB system. This experiment setup serves as a baseline for comparative analysis with the following experiments (Experiment #2 – Experiment #4). To implement this experiment, all three tags were placed at a height of 1.57 meters. The height selection corresponded to an intermediate level between the floor and the height of receivers as visually demonstrated in Figure 4-13.a. Therefore, these tags were placed at a lower height compared to the four receivers. The tags were positioned within the enclosure’s rectangular area created by the four receivers such that there was a clear Line-of-Sight (LOS) between each tag and the receivers. Table 4-6 provides the actual location of each tag within the test area. According to Table 4-5, 3,054 data points were collected during Experiment #1. During the 5-minute experiment, Tag-0 and Tag-2 each collected 1,067 data points, while Tag-1 collected slightly fewer data points of 920.

Based on the results presented in Table 4-6, it was shown that the standard deviation of all tags along the x, y, and z axes were remarkably small, each equal to or less than 0.02 meters. This demonstrates a clustered distribution of UWB points around their respective average for all three tags. This also can be visualized in Figure 4-14.a. which demonstrates true locations, UWB datapoints and UWB averaged locations for the tags in 2D and 3D planes. The average error for all three tags in this experiment was 0.018 meters in the 2D plane and 0.058 meters in the 3D plane. In addition, the individual error for each tag was calculated and represented in Table 4-7. This table shows that all three tags had relatively low errors. The 2D error values ranged from 0.012 to 0.022 meters, while the 3D errors were between 0.052 to 0.064 meters.

A centimeter-level error in UWB localization in 2D and 3D planes within an indoor laboratory area demonstrates the precision and reliability of the UWB system. This experiment serves as an ideal scenario with tags at the same height level and not close to the ground or ceiling.

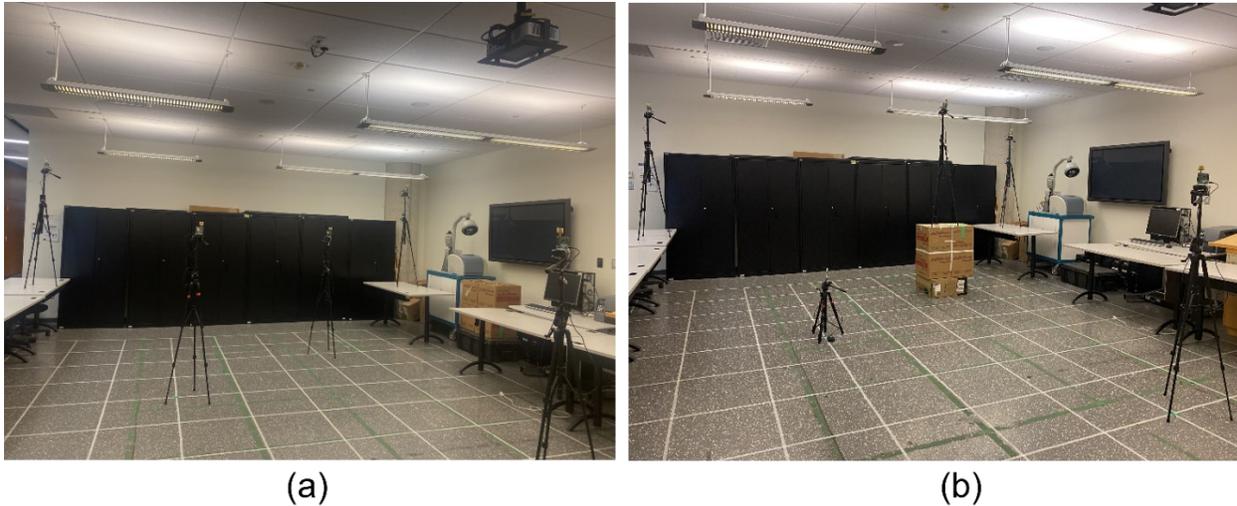
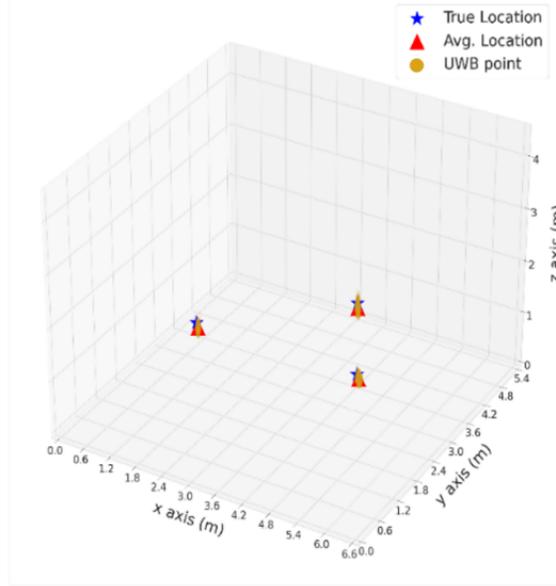
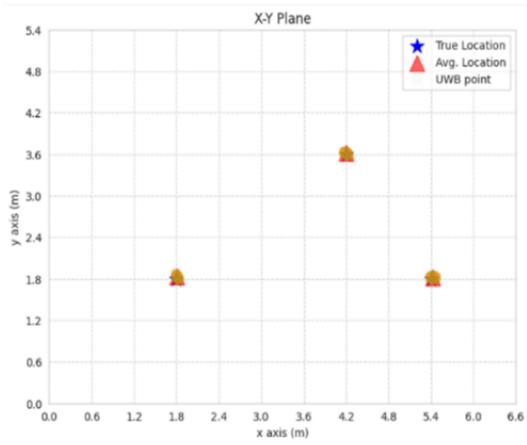


Figure 4-13. Placement of the tags in the test area: a) Experiment #1 and b) Experiment #2.

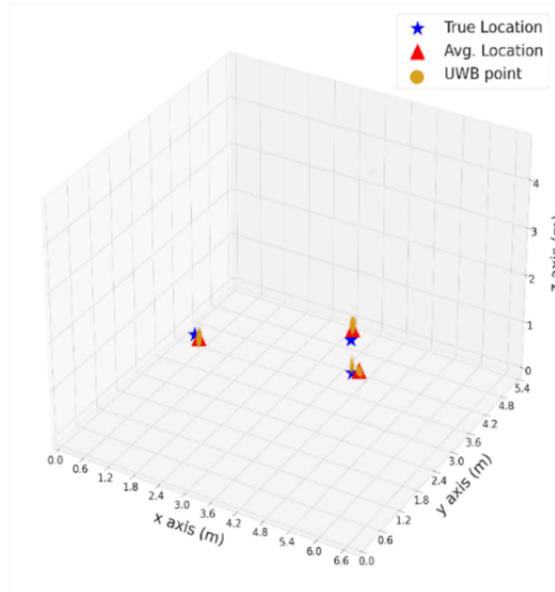
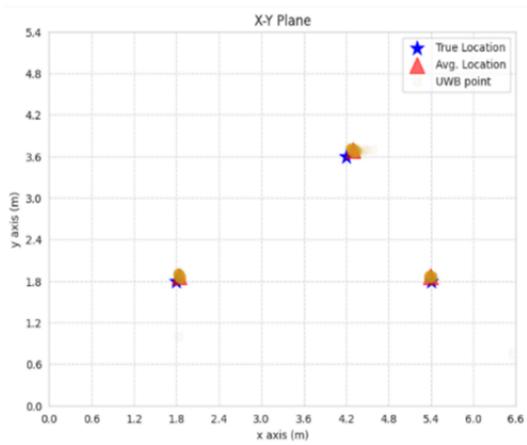
- Experiment #2 – Tags at different height levels.

This experiment aims to investigate the UWB system's performance when the tags are positioned at varied heights instead of a uniform intermediate level. In this experiment, compared to Experiment #1, the only alteration is changing the adjustments of the tags' heights, while the placement of tags in the 2D plane remained the same.

In Experiment #2, Tag-0's height was adjusted close to the floor by changing the tripod's height to a level of 0.5 meters. Tag-1's height remained at 1.57 meters while Tag-2 was elevated to the height of 2.50 meters. The placement of tags is depicted in Figure 4-13.b. Due to tripod height limitations, Tag-2's tripod needed to be elevated further, so it was placed on a cardboard box to reach 2.50 meters, making it higher than the three receivers. Based on Table 4-5, the total collected data points in this experiment were 3,120 from which 1,088 points were from Tag-0, 942 from Tag-1, and 1,090 from Tag-2. The UWB data points had higher standard deviations than those in Experiment #1 as shown in Table 4-6. The most increased standard deviation values were observed in Tag-0 and Tag-2, especially in the z-axis, which were located at very low and very high heights, respectively. This highlights the impact of the tags' heights on the performance of the UWB system.



(a)



(b)

Figure 4-14. UWB data readings for tags on 2D and 3D planes in (a) Experiment #1 and (b) Experiment #2.

In this experiment, an average error of 0.064 meters was obtained in 2D and 0.102 meters in 3D, showing an increase in error as compared to Experiment #1. Specifically, in Experiment #2 compared to Experiment #1, for Tag-0 which was located near the ground level, the error increased by 0.08 meters in 2D and 0.07 meters in 3D. For Tag-2, situated close to the ceiling, the error increased by 0.02 meters in 2D and 0.053 meters in 3D. In addition, for Tag-1, which remained at the same height in both experiments, the error increased in Experiment #2, but not significantly as illustrated in Figure 4-15.

The analysis of standard deviation values and errors indicates that changing the height of tags, whether located very close to the floor or higher than receivers can reduce the performance of the UWB system. In general, placing the tags at floor level or ceiling level might increase the likelihood of signal diffraction, and reflection by other objects present in the area. Despite the increased errors in this experiment, the localization error of each tag remained below 0.15 meters even on a 3D plane. It was demonstrated that the UWB system maintained its reliability and accuracy within this experiment in which the tags' heights varied.

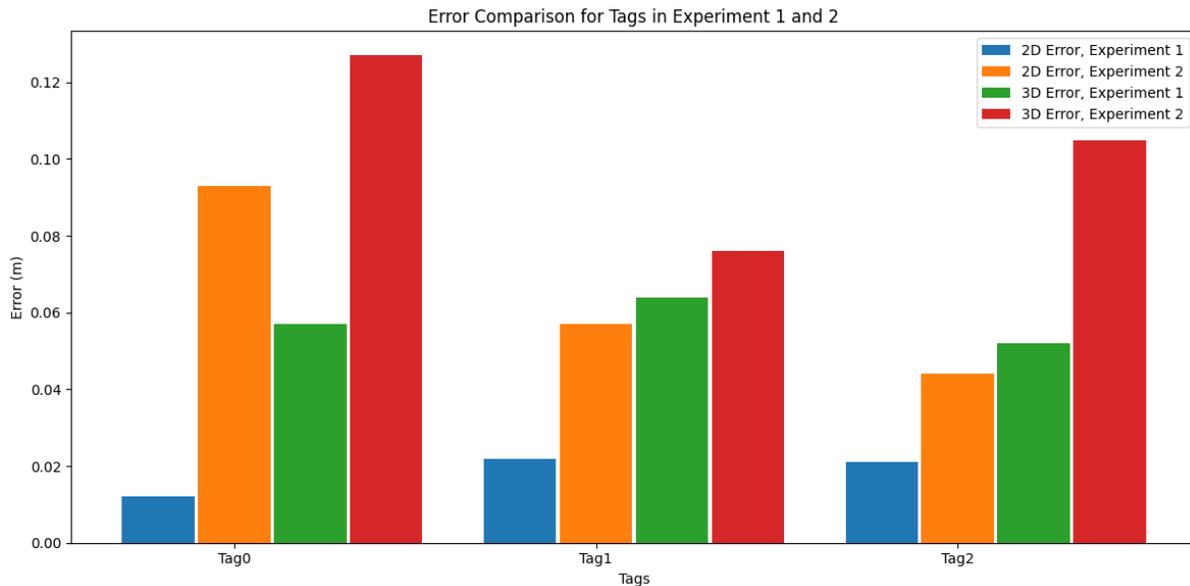


Figure 4-15. 2D and 3D errors of the tags in Experiment #1 and Experiment #2.

- Experiment #3 – Tags at the same height level with proximity to receivers.

Experiment #3 was conducted to determine the performance of the system when tags are close to receivers. All three tags were in the proximity of 1.2 meters to one of the receivers (Receivers 0, 2, and 3) as depicted in Figure 4-16.a. In this experiment, all the tags' heights were equal to 1.57 meters. A total of 3,130 data points were collected and distributed as: 1,080 for Tag-0, 971 for Tag-1, and 1,079 for Tag-2. According to the standard deviation values in Table 4-6, the UWB data points are more scattered compared to Experiment #1, where the tags also had the same height of 1.57 meters. The highest standard deviation values were observed for Tag-0 on the z-axis and y-axis as presented in Table 4-6.

In this experiment, the average error was 0.053 meters in 2D and 0.388 meters in 3D. The most significant difference in error compared to Experiment #1 was observed with Tag-0, where the error increased to 1.040 meters in 3D. The other tags showed slight differences in their 3D errors.



Figure 4-16. Placement of the tags in the test area: a) Experiment #3 and b) Experiment #4.

- Experiment #4 – Tags at different height levels with proximity to receivers.

In this experiment, the location of the tags within a 2D plane remained the same as in Experiment #3. However, the heights of the tags were adjusted to reflect the changes made in Experiment #2. This included one tag close to the ground at 0.5 meters, another at 1.57 meters, and the third one at 2.5 meters.

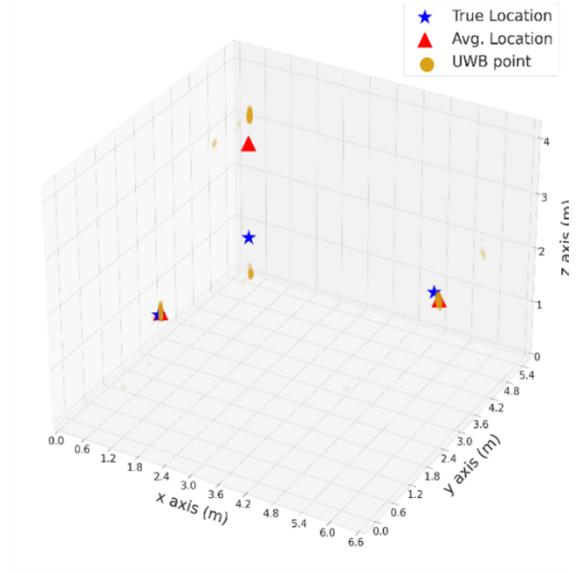
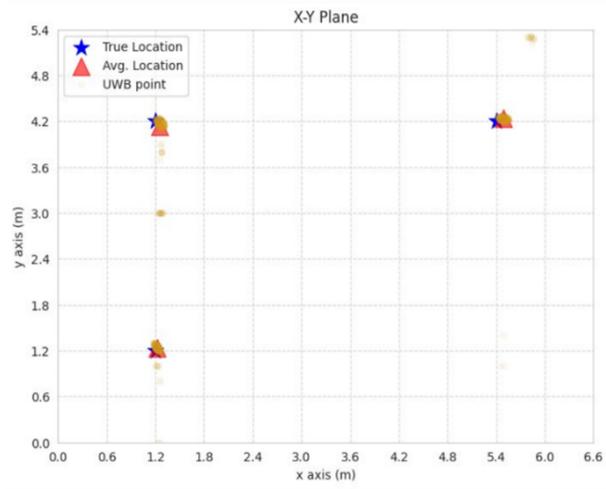
In this experiment, according to Table 4-5, 3,104 data points were collected across all three tags. Analysis of the data from Table 4-6, demonstrates that the standard deviation of all points in the x and y axis was less than 0.05 meters. This demonstrates that the points are closely clustered in a 2D plane which is depicted in Figure 4-17.b. However, the standard deviation values on the z-axis are higher, showing a more dispersed distribution of data points in this dimension.

In Experiment #4, The alteration in tags' heights resulted in a small increase in the 2D average error of all three tags compared to Experiment #3. This highlights the robustness of the UWB system for the 2D localization of materials in indoor job sites, even under scenarios where the tag heights are different from the ground floor to the ceiling.

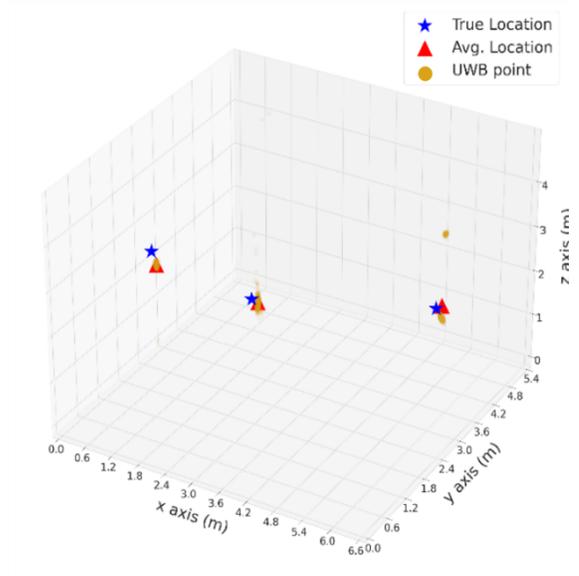
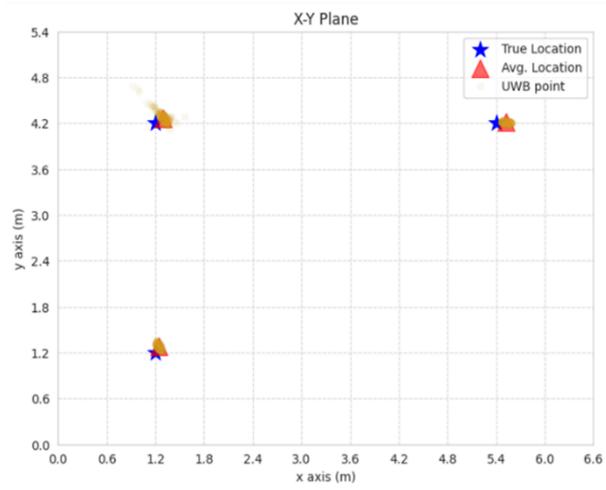
In this experiment, the average 3D error across all tags was 0.185 meters. When the height of Tag-2 was increased to 2.5 meters, the 3D error increased to 0.385 meters. This shows that increasing the height of tags above the receivers can negatively affect the performance of the UWB system, as also observed in Experiment #2. For Tag-1, the 3D errors remained relatively consistent compared to Experiment #3. However, the 3D error of Tag-0 in this experiment was significantly reduced to 0.086 meters when its height was lowered to floor level. This suggests that Tag-0 communicates better with the receivers at floor level resulting in better localization estimation.

Table 4-7. UWB localization errors on 2D and 3D planes.

Experiment	Tag's name	Error in 2D (m)	Error in 3D (m)
1	Tag-0	0.012	0.057
	Tag-1	0.022	0.064
	Tag-2	0.021	0.052
2	Tag-0	0.093	0.127
	Tag-1	0.057	0.076
	Tag-2	0.044	0.105
3	Tag-0	0.060	1.040
	Tag-1	0.068	0.094
	Tag-2	0.031	0.030
4	Tag-0	0.084	0.086
	Tag-1	0.083	0.085
	Tag-2	0.064	0.385



(a)



(b)

Figure 4-17. UWB data readings for three tags on 2D and 3D planes in (a) Experiment #3 and (b) Experiment #4.

4.2.3 Field Experiments under Line-of-Sight (LOS) and Non-Line-of-Sight (NLOS) Conditions

Field experiments were conducted under both LOS and NLOS conditions in an indoor high multipath environment. Mechanical rooms are challenging workplaces due to the presence of numerous objects in a congested environment that can negatively affect signal communications in RTLS. For this study, the mechanical room, located on the 16th floor of the John Molson building at Concordia University was selected to evaluate the UWB system's performance under the specified conditions. This mechanical room, with an area of 223.621 square meters, was occupied with metallic MEP components, including HVAC ducts, pipes, lighting, and fire components. Additionally, structural elements such as steel columns and beams, and other objects such as steel ladders, rails, and hangers were present in the room as shown in Figure 4-18.



Figure 4-18. Mechanical room at Concordia University.

The Trek1000 evaluation kit was selected for this experiment in order to localize HVAC ducts and pipes within the mechanical room. In a congested workplace like a mechanical room, selecting the location for receivers is essential to ensure accessibility, and maintain proper LOS with tags. The position of receivers and tags in the mechanical room was determined through site visits prior to setting up the experiments, available 2D drawings of the room and consultations with the facility management of Concordia University. The receivers were positioned to ensure

proper signal communication between each other, accessibility to their locations, and to maintain a minimum distance of 15 centimeters from any surrounding objects. Four receivers were mounted on tripods and taped for stability. The first three receivers were installed at the same height of 4.90 meters while the fourth receiver was installed higher, at 5.40 meters, to ensure achieving more accurate results. Figure 4-19. shows the layout of the mechanical room and the locations of the receivers. In this study, UWB tags are assigned to critical components that require tracking. If additional tags are available, they can be placed at different locations on MEP components to identify variations and ensure alignment within the network.

The experiments were conducted with all tags installed lower than the receivers based on the lab experiments in Section 4.2.2 where it was observed that tags located higher than the receivers had increased localization errors. In order to power and activate the receivers, Receiver-0 was connected to the laptop, while the other three receivers were powered by external USB batteries. Receiver locations were fixed for all the experiments conducted in the mechanical room.

As a prerequisite for running the UWB application installed in the laptop, the 3D coordinates of the receivers in the mechanical room had to be measured and imported manually. Accurate measurement of the receivers' locations has a direct impact on the accurate localization of the tags by the UWB system. Hence, a systematic measurement approach was followed. Receiver-0 was set as the reference point (0,0,4.90) within a three-dimensional plane. The locations of the remaining receivers were then measured using a laser distance measurer. To validate the accuracy of the receiver positions in the mechanical room, the location measurements were repeated three times, and the results were averaged.

In total, eight locations were chosen for tagging components, with two tags attached to the HVAC ducts and the remaining tags attached to pipes. Figure 4-20. Provides a visual representation of the eight tags attached to components in the mechanical room. The tags were securely placed on the stands and fixed on the components using tape. They powered by USB batteries with accompanying USB cables.

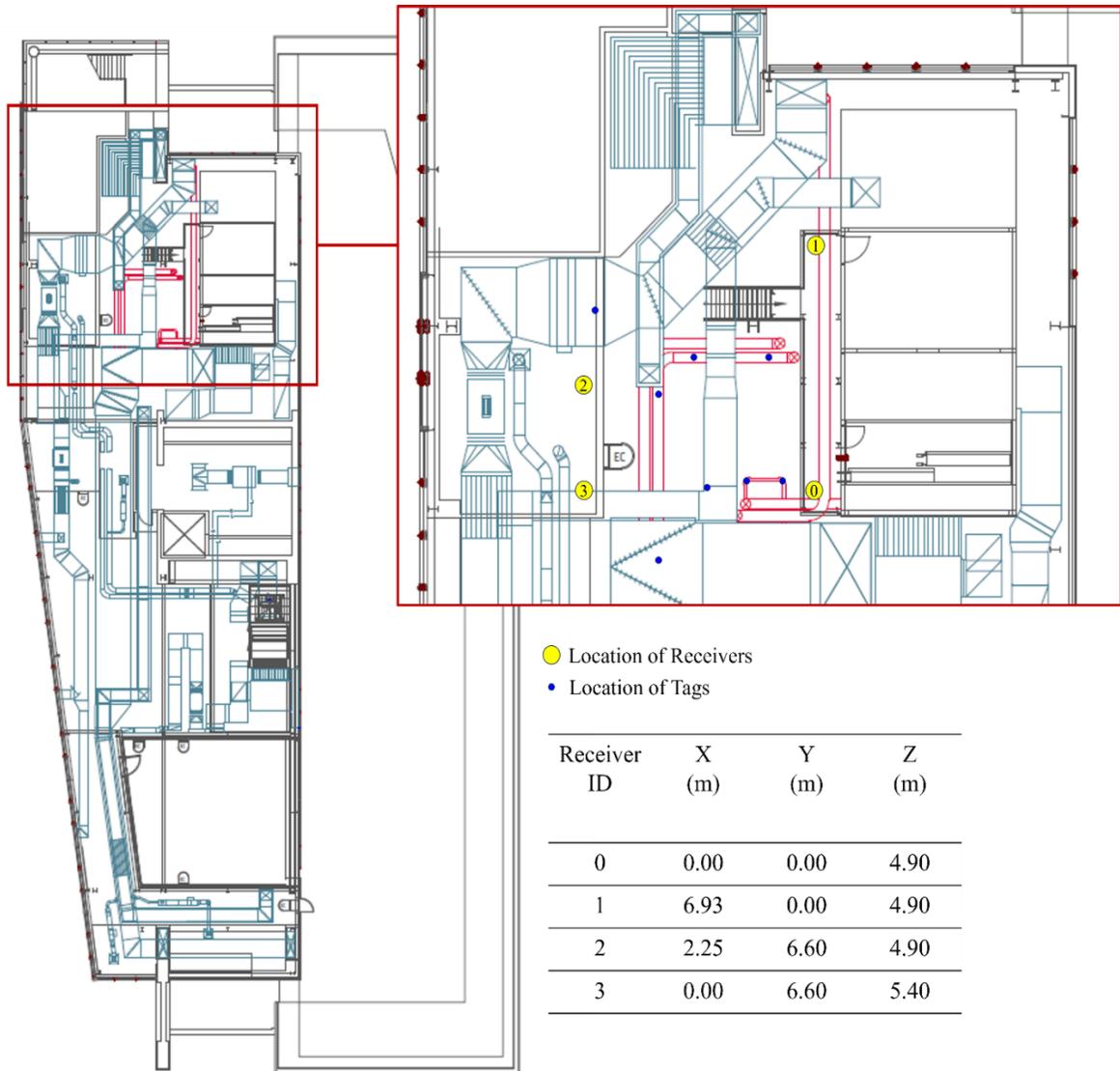


Figure 4-19. Mechanical room plan and the locations of the receivers in the room.

According to Maalek & Sadeghpour, (2013), most of the tags in this study were placed on top of HVAC ducts and pipes to obtain better LOS conditions. However, as a result of restricted access to the top surface of one HVAC duct, one tag (Tag-4) was placed below the element. Due to the limited number of available tags (four), a sequential approach was employed. Initially, the four tags were attached to four predefined locations. After sufficient data collection from the tags, they were detached and relocated to the remaining four locations, and the data collection process was repeated. The tags were first numbered from 0 to 3 using switches located on the PCB of the tags. Then the tags were detached, renumbered from 4 to 7 and placed in other four locations. The

tag locations on tracked objects within the mechanical room were selected to evaluate the UWB system under different indoor conditions that commonly exist in job sites. These conditions include the presence of metallic objects between the tags and receivers, the tags being within an enclosure area, LOS, and NLOS conditions.

Table 4-8 describes the information on each tracked object class, the unique ID of the attached tag, the tagging location on the objects (either on top or below the object), whether the tag is located inside or outside the enclosure area formed by the four receivers and LOS/NLOS conditions with the receivers. Figure 4-21 demonstrates samples of different LOS/NLOS conditions that can happen between each tag and receiver. Figure 4-21(a), shows a clear LOS between a tag and a receiver, which is the ideal condition. However, in some conditions LOS exists but metallic objects are in close proximity to the tag or receiver, which may cause signal reflections as depicted in Figure 4-21(b). Additionally, high obstructions by numerous obstacles such as ladders or rails, can cause NLOS condition as depicted in Figure 4-21(c). NLOS conditions cause increased error in localization performance (Liu et al., 2022) since the signals pass through a longer distance between the tag and receiver as compared to the direct LOS path (Kristensen et al., 2019).

The setup of the UWB system and the experiments were completed over three days in June 2023. The tags' locations were determined using the same measurement process as receiver locations. Table 4-9 illustrates the actual locations of the tags within the mechanical room and Table 4-10 displays the distances between each tag and receiver. In the mechanical room, three tests were conducted following the installation and activation of tags and receivers which are described in the following subsections.

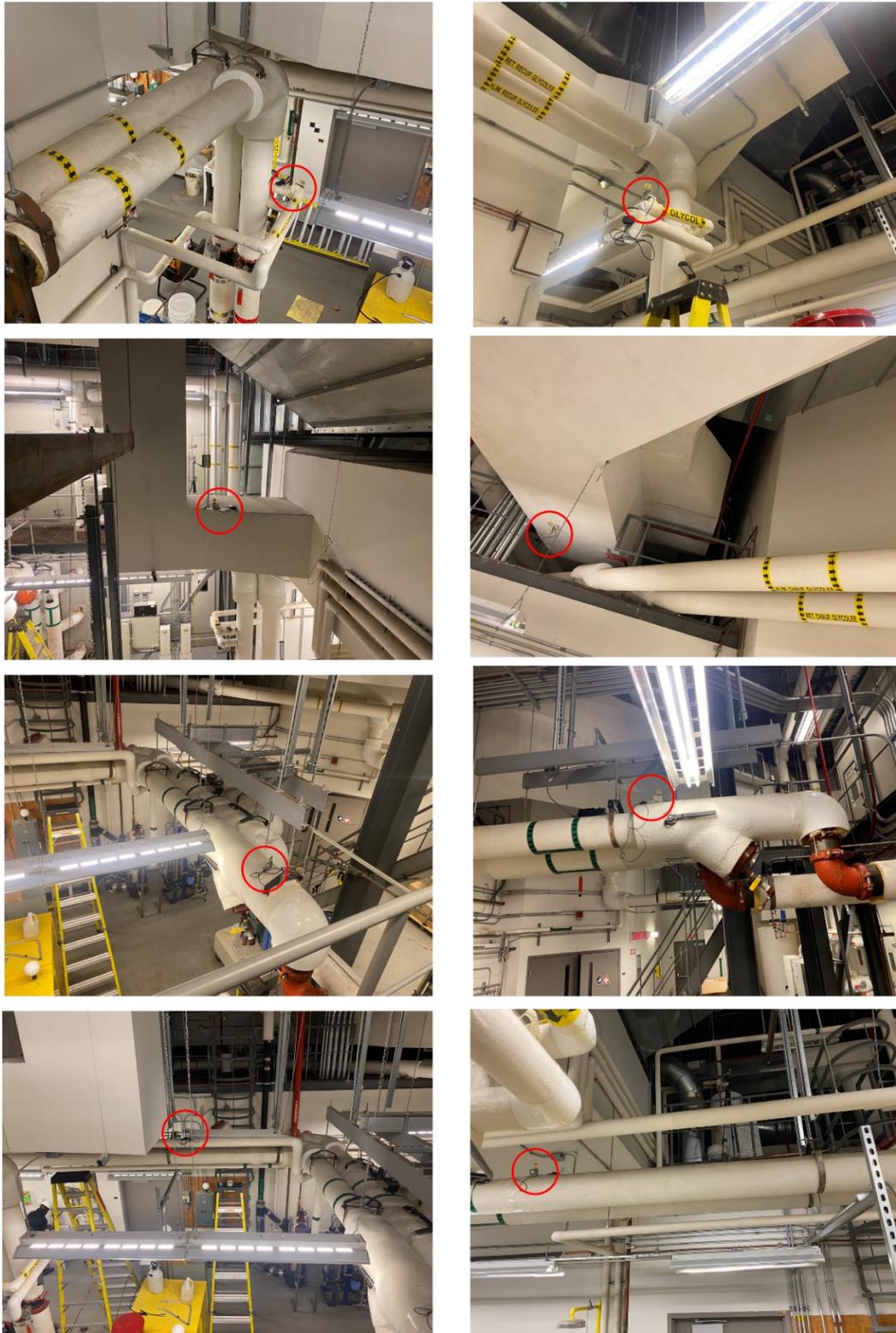


Figure 4-20. Distribution of the tags in the mechanical room.

Table 4-8. Tagging details of each tracked component.

Object class	Tag ID	Tagging location	Enclosure Area	LOS/NLOS condition between the tag and each receiver
Pipe	0	Top	Inside	Clear LOS with Receiver-0 and Receiver-3. LOS with Receiver-1 and Receiver-2, but metallic objects are in proximity.
HVAC duct	1	Top	Inside	Clear LOS with Receiver-0 and Receiver-3. Blocked LOS with Receiver-1 due to complete blockage by the other section of the HVAC duct. NLOS with Receiver-2 due to obstructions from metallic components.
Pipe	2	Top	Inside	Clear LOS with Receiver-2 and Receiver-0. LOS with Receiver-1 but metallic objects are in proximity. NLOS with Receiver-3 due to obstructions from metallic components.
Pipe	3	Top	Inside	Clear LOS with Receiver-3. Blocked LOS with Receiver-0 due to complete blockage by the HVAC duct. NLOS with Receiver-2 and Receiver-1 due to obstructions from metallic components.
HVAC duct	4	Below	Outside	Clear LOS with Receiver-0 and Receiver-2. NLOS with Receiver-1 and Receiver-3 due to obstructions from metallic components.
Pipe	5	Top	Outside	Clear LOS with Receiver-2 and Receiver-1. NLOS with Receiver-0 due to metallic components obstructions. Blocked LOS with Receiver-3 due to complete blockage by the HVAC duct.
Pipe	6	Top	Outside	Clear LOS with Receiver-0. LOS with Receiver-1, but metallic objects are in proximity. NLOS with Receiver-2 and Receiver-3 due to obstructions from metallic components.
Pipe	7	Top	Inside	Clear LOS with Receiver-0, Receiver-2 and Receiver-3. LOS with Receiver-1 with a few metallic objects in proximity.

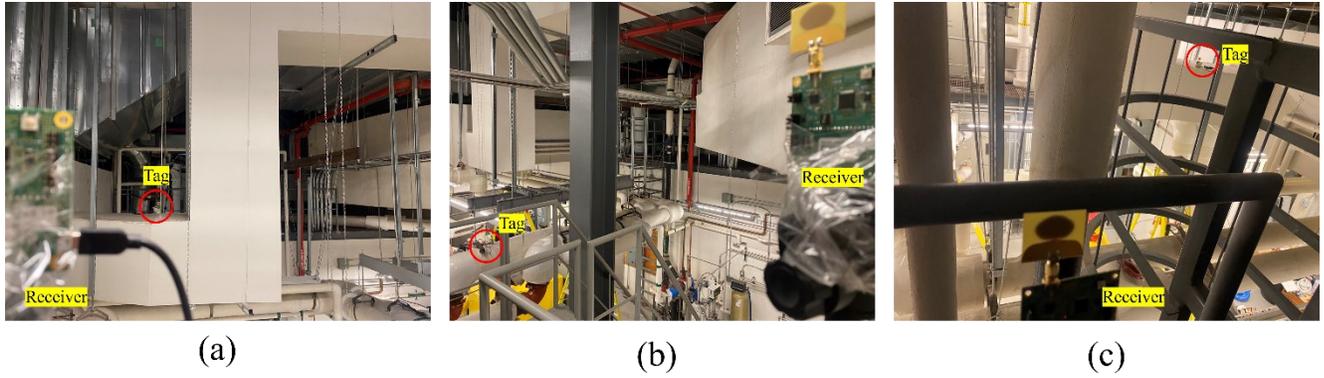


Figure 4-21. Examples of conditions between tags and receivers.

Table 4-9. Actual locations of tags within the mechanical room.

Tag's name	X (m)	Y (m)	Z (m)
Tag-0	0.197	2.925	3.061
Tag-1	0.014	3.401	4.744
Tag-2	2.870	1.620	3.381
Tag-3	0.792	4.281	3.300
Tag-4	4.087	6.521	4.584
Tag-5	-1.296	4.472	3.300
Tag-6	-0.197	2.255	3.056
Tag-7	2.870	2.572	3.313

Table 4-10. Distances between the tags and receivers.

	Tag-0	Tag-1	Tag-2	Tag-3	Tag-4	Tag-5	Tag-6	Tag-7
Receiver-0	3.460	3.404	3.628	4.638	7.702	4.923	2.919	4.167
Receiver-1	7.567	7.708	4.627	7.652	7.120	9.498	7.699	5.061
Receiver-2	4.593	3.906	5.243	3.172	1.865	4.434	5.316	4.373
Receiver-3	4.360	3.265	6.092	3.227	4.168	3.258	4.940	5.368

4.2.3.1 Test #1 and #2 – Performance Analysis of the UWB System.

To evaluate the UWB system with different number of receivers in indoor, high multipath environments two tests were conducted. Test #1 included four active receivers based on the manufacturer's recommendations for accurate 3D object localization. In Test #2, three receivers were activated, and the fourth receiver (Receiver-3) was deactivated. This test aimed to evaluate the system's ability in scenarios where one receiver may fail to function or when installing additional receivers is challenging in congested rooms.

For each Test #1 and Test #2, data collection time was set to 40 minutes, with 20 minutes allocated to each set of four tags. A total of 29,210 and 30,208 data points were collected for Test #1 and Test #2, respectively as detailed in Table 4-11. Except Tag-1, the remaining tags in both tests obtained relatively constant number of data points between 3,749 and 4,592. However, Tag-1 recorded significantly lower datapoints, with 112 in Test #1 and only 2 in Test #2, indicating communication issues due to a high level of obstructions between Tag-1 and two receivers, as mentioned in Table 4-8. The removal of the fourth receiver (Receiver-3) in Test #2, which had a clear LOS with the Tag-1, further weakened its communication with the UWB system. However, this removal did not impact on the data consistency for other tags.

Table 4-11. Captured datapoints in Test #1 and Test #2.

Test	Total data points	Tag-0	Tag-1	Tag-2	Tag-3	Tag-4	Tag-5	Tag-6	Tag-7
Test #1	29,210	4,285	112	4,253	3,749	4,329	4,343	3,811	4,328
Test #2	30,208	4,592	2	4,580	3,800	4,422	4,424	3,978	4,410

The individual analysis of each tag's performance is crucial due to the specific location and conditions between each tag and receivers. This analysis allows for future planning and deployment of UWB system in similar indoor areas. Table 4-12 presents the average location of UWB data points for each tag, along with their standard deviation in 3D coordinates and the UWB localization error in both 2D and 3D planes for Test #1 and Test #2. Figure 4-22 demonstrates the scatter of UWB data points, the average location of UWB datapoints and the actual location of each tag for both tests.

The analysis of Table 4-12 highlighted the significant impact of LOS conditions on the UWB localization performance. Tags with LOS to receivers, such as Tag-7 and Tag-2, achieved considerably lower localization errors as compared to those affected by obstacles which caused NLOS conditions. This was evident in both 2D and 3D localization errors, with 3D errors being more affected. According to Figure 4-22 in both tests, tags with proper signal communication have more clustered UWB data points, which indicates higher reliability of the UWB localization. In contrast, tags with NLOS conditions with the receivers such as Tag-3 and Tag-6, had more dispersed data points.

In Test #1, the majority of tags (7 out of 8) had 2D localization errors between 0.036 and 0.407 meters, with one tag which reached to 1 meter error. In Test #2, The 2D error range was between 0.034 to 0.388 meters with two tags exceeding a 1-meter error. For 3D localization in Test #1, errors were from 0.150 to 0.939 meters with one tag reaching to 1.277 meters. In Test #2, errors for six tags, had an error range of 0.151 and 0.916 meters, with two tags having errors above 1 meter.

The standard deviations of UWB points in Test #1, excluding Tag-1, on the x and y axes ranged from 0.041 to 0.378 meters. On the z-axis, standard deviations for five tags varied from 0.049 to 0.602 meters excluding Tag-6 and Tag-5 which had deviations of 2.422 and 3.018 meters, respectively. In Test #2, without considering Tag-1, the standard deviations were smaller across all axes ranging from 0.019 to 0.30 meters. This shows that UWB points were more clustered with three receivers in Test #2 as compared to Test #1 as illustrated in Figure 4-22. The deactivation of the fourth receiver may have reduced multipath effects or simplified the system complexity that results in better clustering of datapoints.

A comparative analysis between Test #1 and Test #2 revealed key observations. First, 2D localization errors for each tag in Test #1 were closely similar to those in Test #2, except for Tag-1, which encountered data loss in Test #2. This suggests that adding a fourth receiver did not significantly reduce 2D errors. Thus, for 2D object localization even in complex environments, three receivers are sufficient for the UWB system's reliable performance. This approach can mitigate extra costs and deployment challenges, provided that proper signal communication exists between tags and the receivers to prevent data loss. Second, tags with clear LOS to the first three receivers, including Tag-2 and Tag-7, showed no significant change in localization errors with the

removal of the fourth receiver. A comparison of error distributions of the UWB data points collected for Tag-7 in both 2D and 3D is shown in Fig 4-23(a), indicating similar UWB localization for the tag in both tests.

In addition, the tags with NLOS or blocked conditions with Receiver-3 (Tag-2, Tag-4, and Tag-6), had reduced 3D errors by its deactivation in Test #2. This improvement is mainly due to the reduced multipath propagation, resulting in better localization in Test #2. This phenomenon is more observed in 3D errors than in 2D. Fig 4-23(b) demonstrates 2D and 3D error distribution of Tag-6 in both tests. Despite minimal differences in 2D errors, the 3D error distribution in Test #1 showed an additional cluster of errors in a higher range. Similar additional clusters of errors were seen for Tag-2 and Tag-4, but the range and occurrence of these errors varied among the tags.

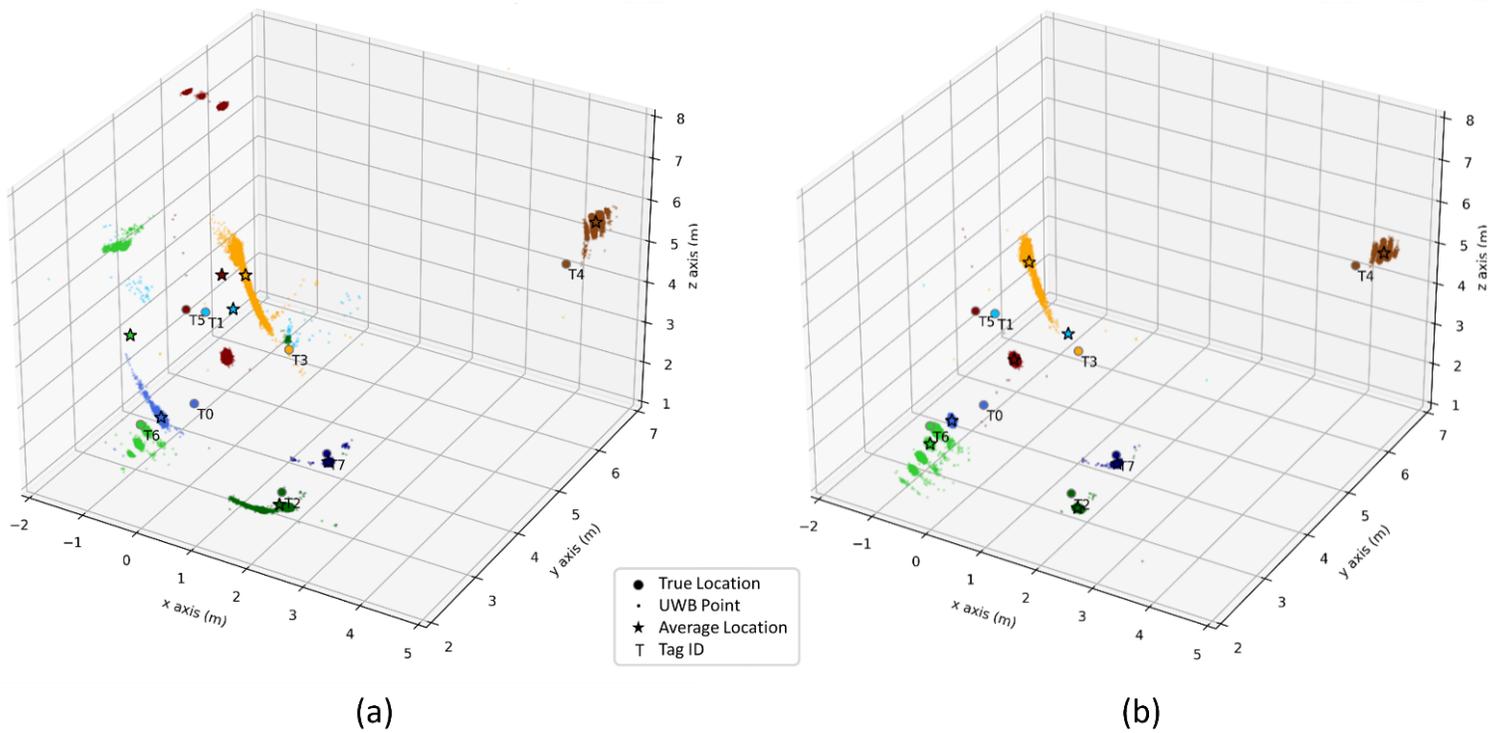


Figure 4-22. The tags' actual locations, UWB average locations, and data points on a 3D Plane in (a) Test #1 and (b) Test #2. Each color corresponds to a different tag ID.

Table 4-12. Mean, standard deviation and error of the UWB system for Test #1 and Test #2.

	UWB Average Position			UWB Standard Deviation			UWB Error		
	Tag ID	X (m)	Y (m)	Z (m)	X (m)	Y (m)	Z (m)	2D (m)	3D (m)
Test#1	0	-0.376	2.869	2.544	0.100	0.046	0.081	0.407	0.446
	1	0.238	3.736	4.589	1.564	0.120	0.374	0.285	0.249
	2	2.695	1.779	2.839	0.378	0.135	0.602	0.167	0.341
	3	-0.507	4.871	4.079	0.354	0.143	0.470	1.009	0.939
	4	4.277	6.977	5.263	0.041	0.076	0.149	0.349	0.485
	5	-0.930	4.845	3.955	0.052	0.197	3.018	0.369	0.483
	6	-0.246	2.158	5.266	0.046	0.220	2.422	0.076	1.277
	7	2.860	2.622	3.057	0.059	0.043	0.049	0.036	0.150
Test#2	0	-0.337	2.854	2.525	0.026	0.019	0.031	0.381	0.438
	1	-0.099	5.122	2.579	2.147	1.941	2.694	1.219	1.598
	2	2.917	1.698	2.972	0.025	0.039	0.034	0.064	0.241
	3	-0.664	4.928	4.329	0.160	0.078	0.300	1.126	1.095
	4	4.270	6.984	4.532	0.043	0.080	0.131	0.352	0.289
	5	-0.944	4.894	1.810	0.043	0.030	0.072	0.388	0.916
	6	-0.238	2.293	2.559	0.052	0.187	0.197	0.040	0.288
	7	2.857	2.618	3.054	0.051	0.025	0.064	0.034	0.151

The results exhibited the UWB system’s capability to localize tags located outside the enclosure area created by the receivers, including Tag-4, Tag-5, and Tag-6. A deeper investigation is needed to determine the maximum distance from the tag to the enclosure area at which the UWB system can provide accurate 2D and 3D localization. This factor is crucial in construction sites as the tracked components may be located outside the area where receivers are placed, yet identifying their accurate location is necessary. Also, the placement of tags on the objects can be determined based on the installation location of the tracked component. For instance, for components installed at high heights and close to the ceilings, attaching the tags below them may improve the LOS between them and receivers, such as Tag-4 which is attached below the HVAC duct.

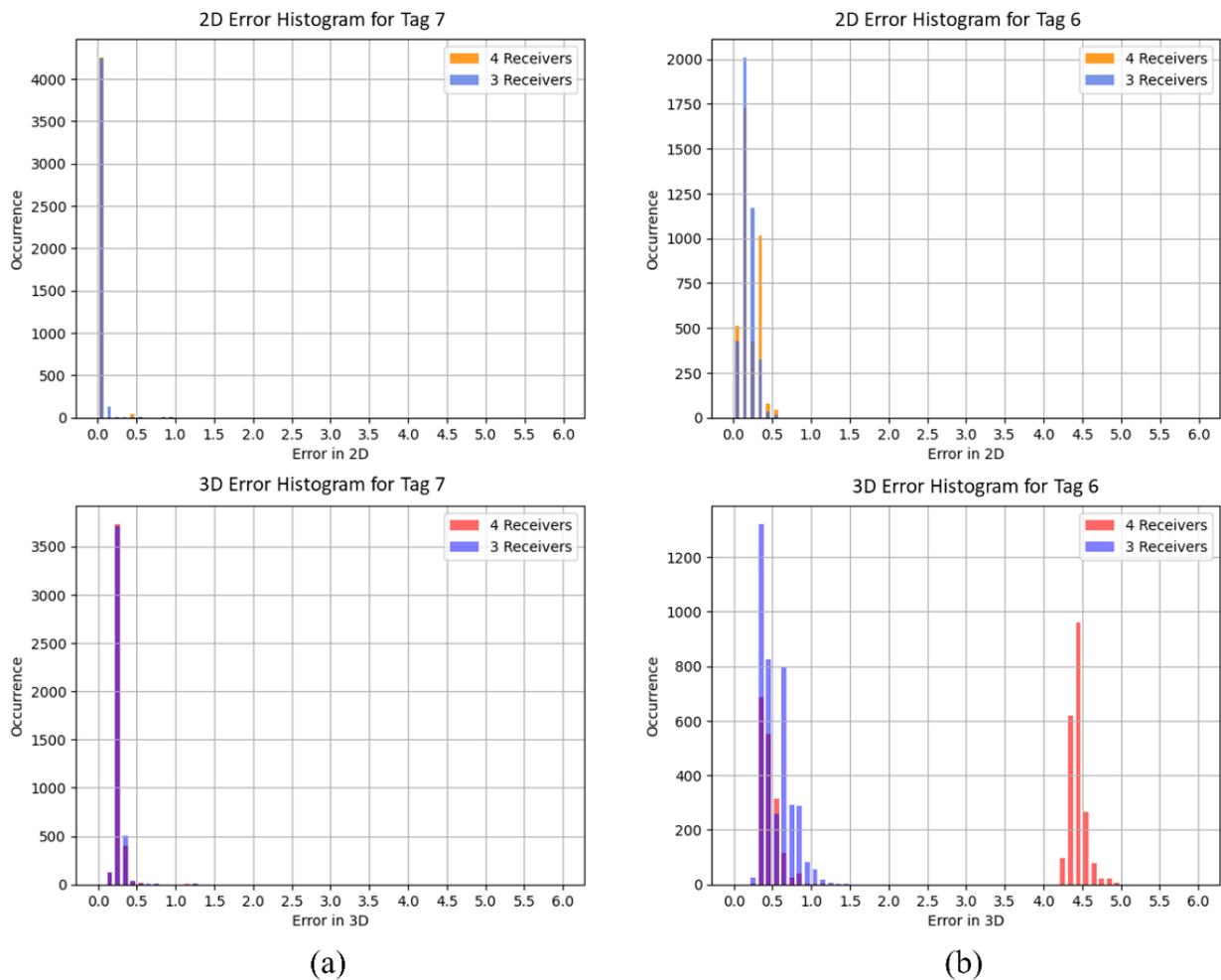


Fig 4-23. Histogram of 2D and 3D errors of the UWB tags: (a) Tag-7 and (b) Tag-6.

4.2.3.2 Test #3 - Impact of the Number of Tags on UWB System Performance

This section evaluates the effect of the number of active tags on the UWB system's localization performance at real job sites. For this purpose, a series of four experiments were conducted sequentially. In Experiment #1, a single tag (Tag-4) was activated for data collection. In Experiment #2, a second tag (Tag-7) was added and data from both tags were recorded. This process continued until four tags were activated in Experiment #4. The locations and numbering of the tags (Tag-4 to Tag-7) align with those described in Table 4-9. The data collection time for each experiment was set to 20 minutes. In these experiments, the four active receivers were maintained at the same locations as shown in Figure 4-19.

Despite this factor being investigated in previous studies, the results were different. In one study, Maalek and Sadeghpour (2013) reported that increasing the number of tags within the test area led to decreased localization accuracy. In contrast, Jin et al. (2019) found that the quantity of tags does not affect the system performance. Hence, in this section, the effect of the number of tags on the performance of the UWB was reevaluated. Table 4-13 provides the number of collected data points for each tag in each experiment. The table also provides the UWB average position for each tag, as well as the UWB standard deviation along x, y and z axes. Figure 4-24 also illustrates the 2D and 3D errors of each tag in each experiment in meters.

According to Table 4-13 and Figure 4-24, there is no significant difference in the performance of the UWB system for the localization of Tag-4 across all four experiments. Tag-7 and Tag-5 also demonstrated consistent performance. These results suggest that the quantity of active tags does not influence the performance of the UWB system.

Possible discrepancies in the results of the mentioned research studies could be due to the use of different commercial UWB products. Therefore, it is recommended that researchers evaluate the effect of the number of active tags on their specific UWB system prior to field implementation.

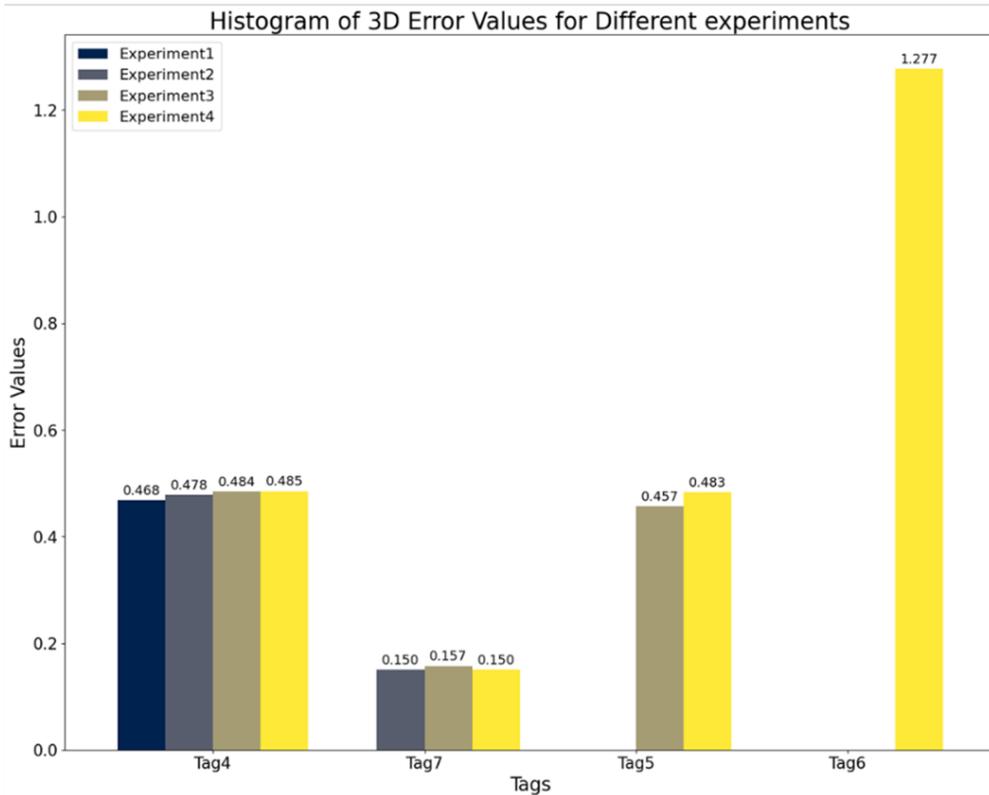
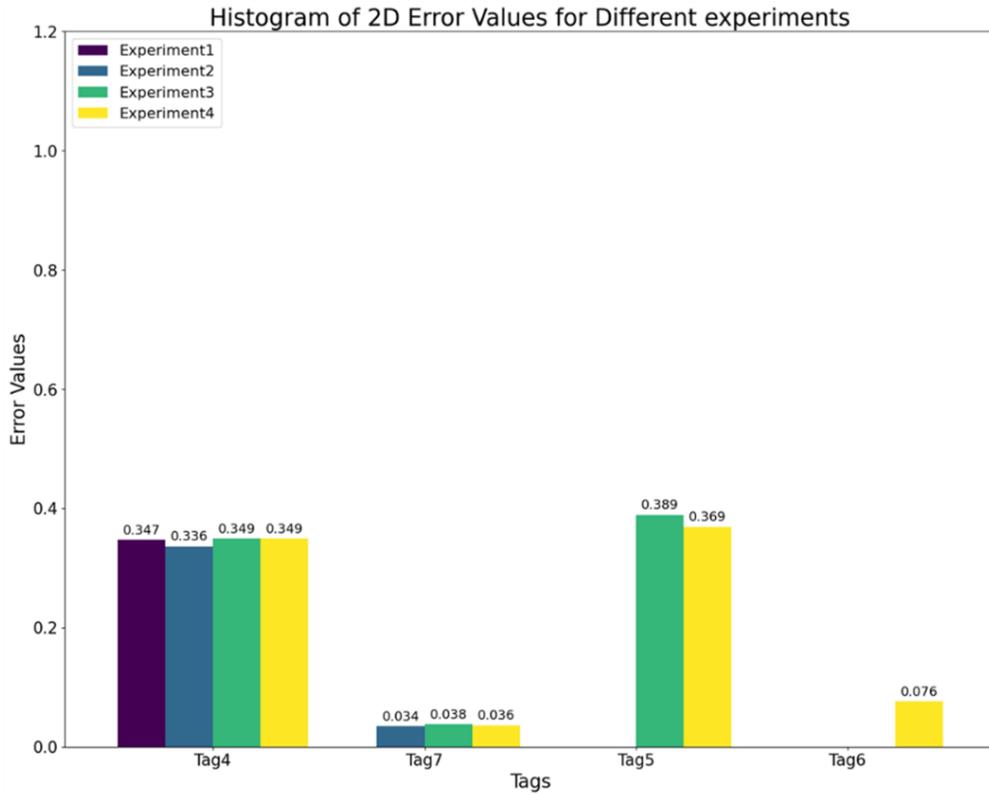


Figure 4-24. Histogram of 2D and 3D error values for Experiments #1- Experiments #4.

Table 4-13. The tags' information in each experiment.

Experiment	Tag's name	Datapoints	UWB Average Position (m)			UWB Standard Deviation (m)		
			X (m)	Y (m)	Z (m)	X (m)	Y (m)	Z (m)
Experiment #1	Tag-4	4,266	4.276	6.974	5.230	0.043	0.137	0.218
Experiment #2	Tag-4	4,300	4.271	6.959	5.262	0.040	0.085	0.154
	Tag-7	4,280	2.858	2.619	3.057	0.067	0.027	0.052
Experiment #3	Tag-4	4,521	4.273	6.978	5.261	0.040	0.073	0.158
	Tag-7	4,519	2.849	2.622	3.044	0.080	0.032	0.061
	Tag-5	4,537	-0.9411	4.892	3.870	0.044	0.025	2.923
Experiment #4	Tag-4	4,329	4.277	6.977	5.263	0.041	0.076	0.149
	Tag-7	4,328	2.860	2.622	3.057	0.059	0.043	0.049
	Tag-5	4,343	-0.930	4.845	3.955	0.052	0.197	3.018
	Tag-6	3,811	-0.246	2.158	5.266	0.046	0.220	2.422

4.2.4 Summary and Conclusions

This section implemented and evaluated the UWB system for 2D/3D object localization through various laboratory and field experiments. These experiments aimed to assess the UWB system's applicability in construction sites to support automated progress monitoring and reporting systems. Moreover, these experiments considered different factors that may affect the system's performance in construction sites.

The Laboratory experiments were conducted under LOS conditions, which are ideal scenarios. Based on the results obtained from the four laboratory experiments, the UWB system demonstrated to be reliable and robust for 2D and 3D localization of objects in indoor environments. Across all experiments, all the tags had errors less than 0.10 meters for 2D localization. In addition, in terms of 3D localization, 9 out of the 12 tag locations had errors equal to or below 0.10 meters. However, it's worth noting that a 3D localization error of 1 meter was recorded when a tag was close to the receiver despite LOS conditions in an obstacle-free environment. Based on these experiments it was identified when the tags are located higher than

the receivers, both 2D and 3D localization performances are negatively affected. Consequently, for implementing the UWB system in real construction sites, it is recommended to avoid placing the tags higher than the receivers and to install receivers at the highest possible positions. In the future, additional laboratory experiments can be conducted to evaluate the UWB system under the following scenarios:

- 1) Evaluating the effect of environmental parameters such as humidity and weather on the UWB system performance.
- 2) Investigating the effects of surrounding metallic objects in the experimental area that may cause multipath effects.

For the field experiments, a mechanical room containing numerous metallic objects was selected. HVAC ducts and pipes were chosen for monitoring because mechanical, electrical, and plumbing (MEP) components are identified as having complex and expensive procedures, which make their effective progress monitoring crucial (Bosché et al., 2015; Song et al., 2006; Yarmohammadi & Ashuri, 2015).

In the field experiments, UWB localization errors ranged from 0.03 to 1.22 meters in 2D and from 0.15 to 1.6 meters in 3D. This range of errors was associated with the LOS condition between each tag and the receivers. Tags that have proper LOS with the receivers have maintained their localization error in the centimeter level in both 2D and 3D. However, the tags with the NLOS conditions with the receivers had higher localization errors.

Based on the results, the UWB system is capable of localizing objects in indoor environments. The UWB system offers a solution for automated progress monitoring and reporting systems by providing real-time location information of tracked objects in construction sites. This information can assist in determining their status on job sites, whether they are in the process of installation, have already been installed, or are located in storage areas.

4.3 Integrated Object Recognition and Localization

4.3.1 Overview

In this section, an application was developed to integrate the data from the deep learning-based object recognition model and the UWB system in an organized and unified format, providing

essential information about the tracked elements. The application's capabilities and its outcomes were then presented. In this section, the case study remains consistent with the one described in Section 4.2.3, focusing on the mechanical room at Concordia University. Similarly, the tracked objects remained HVAC ducts and pipes. The mechanical room represents the final stage of construction with the highest level of occlusion and clutter. All processes applied in the mechanical room can be carried out during earlier stages, which typically have fewer objects such as less piping and HVAC ducts. Consequently, image capturing, UWB installation and data collection become less challenging. Moreover, attaching tags is easier at comfortable heights prior to their final destination.

During the data collection period from the tracked HVAC ducts and pipes using the UWB system, images of these objects were also captured to be used in the developed application. Figure 4-25 displays the outcomes of the object recognition and localization modules which are then organized and fed into the developed application to generate the final report. Despite in this study YOLACT++ and the UWB system was selected, other object recognition models and RTLS technologies could also be applied.

Images from the HVAC ducts and pipes were manually captured using a smartphone with resolution of 2048 * 1536 pixels while the UWB system was active for data collection. To ensure comprehensive capture of all tracked HVAC ducts and pipe components, a sequential approach was employed, where images were taken in a series along the network of MEP components. Further detail was achieved by capturing images from various viewpoints, angles, and positions using different zoom levels at the job site. Despite the fact that the YOLACT++ model can recognize multiple components with different irregular shapes and sizes and account for the continuity and irregularity of MEP components within images. In this study, images were captured as close as possible to each component, ideally with one component per image or at most two components from different classes. This approach is developed and implemented for two main purposes: minimizing the effects of occlusion on object recognition performance and supporting the integration process which is based on a one-to-one matching basis.

After collecting a sufficient number of images from the HVAC ducts and pipes within the mechanical room, 208 of the images were randomly added to the real and synthetic images previously collected in Section 4.1.4. The dataset included 1,462 real images and 668 synthetic

images of HVAC ducts and pipes. It contained 3,245 HVAC duct and 2,196 pipe instances. To save computational time while achieving good performance, transfer learning was used. Instead of training the model from scratch, the pre-trained weight from YOLACT++ using Dataset #8, as detailed in Section 4.1.4, was used to begin model training.

For model evaluation, 328 real images of HVAC ducts and pipes including images captured from the mechanical room were randomly selected. The best performance was achieved with a weight obtained from epoch 268 as provided in Table 4-14. The model achieved a precision of 95.68% and recall 97.08% for HVAC duct detection and a precision of 90.9% and recall 81.98% for pipes. The YOLACT++ prediction speed averaged 6.13 FPS. Predicted images were then categorized by object classes and were stored in separate folders which in this study were HVAC ducts and pipes. If an image contains multiple object classes, it will be repeated in both relevant folders. This ensures that all objects are included in the generated report following the integration process. Sample predicted images are illustrated in Figure 4-26. A separate folder is also created for storing the generated UWB files.

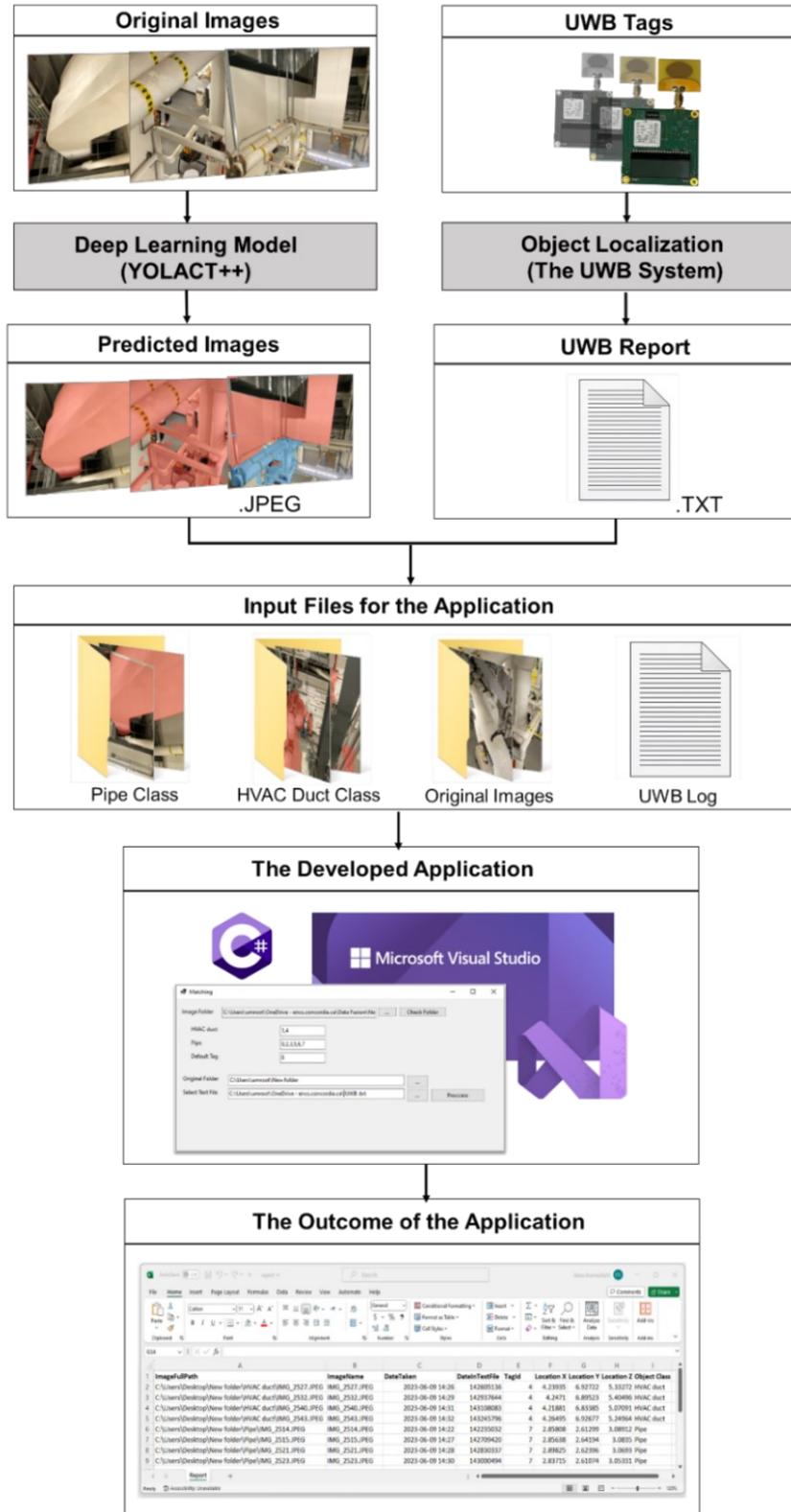


Figure 4-25. An overview of the integrated framework.

Table 4-14. YOLACT++ detection performance for HVAC ducts and pipes

Object class	TP	FP	FN	Precision (%)	Recall (%)
HVAC duct	266	12	8	95.68	97.08
Pipe	141	14	31	90.97	81.98

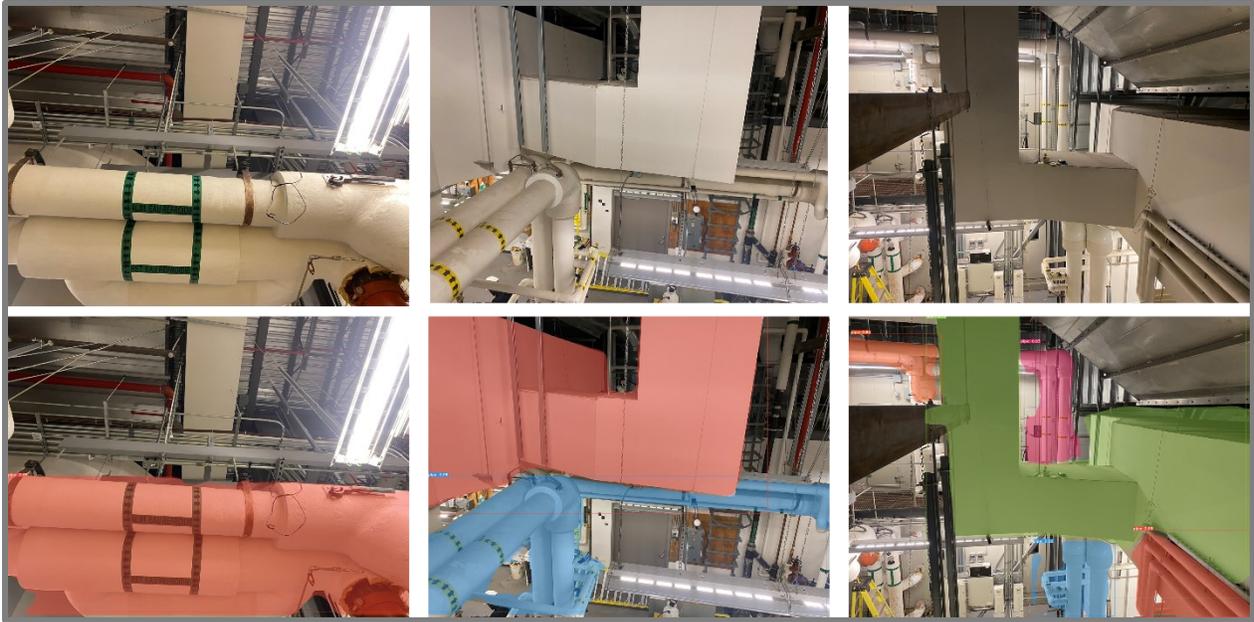


Figure 4-26. Sample images taken from the mechanical room and their corresponding predicted images by YOLACT++.

4.3.2 The Developed Application for Data Integration

For integrating UWB and object recognition data into an organized and comprehensive report, an application was developed. The application was designed in the Visual Studio environment using the C-sharp (C#) programming language that once deployed, can be easily accessed and utilized by the users regardless of their programming expertise. The computer used for testing the integrated model was equipped with Corei7-1065G7 processor, 12GB DDR4 RAM, and Intel Iris Plus Graphics G7.

In the initial step, users are requested to specify the location on the computer system where folders with predicted images exist. Each folder includes the predicted images assigned to a

specific class. The application automatically reads all the folders at the specified location and lists the folder names, which in this case study are HVAC ducts and pipes. In addition, a “Default Tag” name will be added to the list within the application which corresponds to the tag attached to the camera, if available. To obtain the list of classified tags, a text box next to each class allows the user to specify the list of tag IDs assigned to that class.

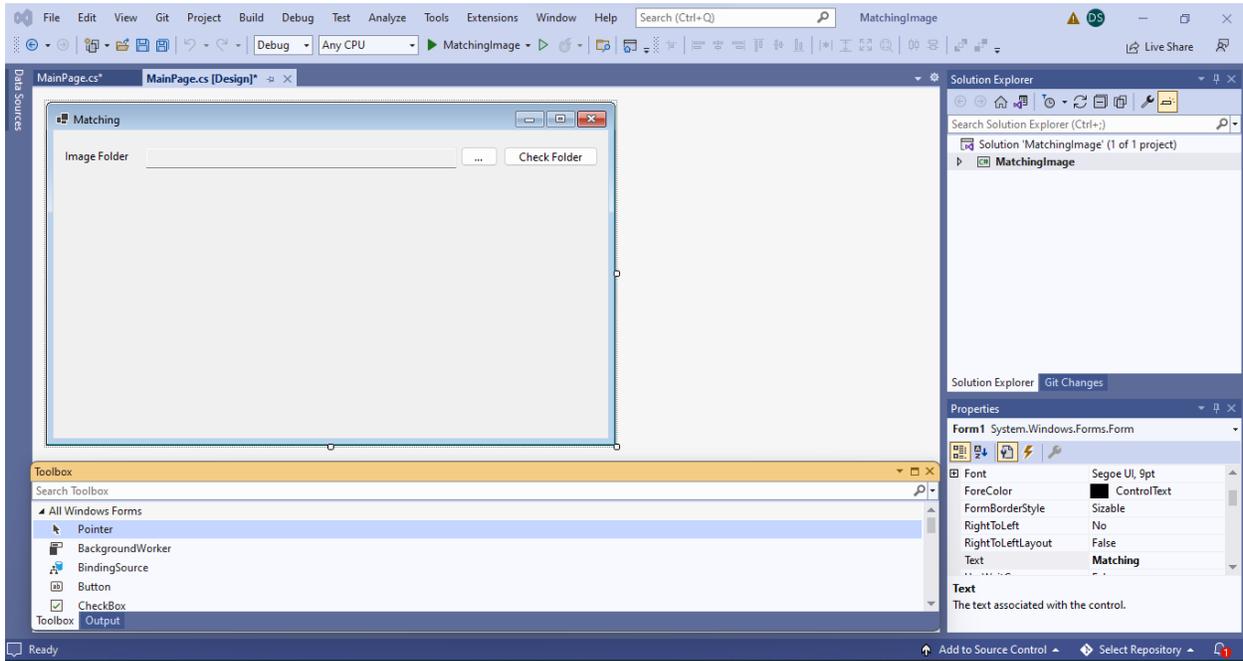


Figure 4-27. Visual studio environment for designing the application user interface.

In the next step, to identify the actual captured time of each image, the user is asked to specify the location path where the original images, before processing by YOLACT++, are stored. This step is necessary since the “date taken” metadata of each image is changed after processing by the deep learning models to the time when the image is predicted by the models. Therefore, this metadata no longer provides an accurate “date taken” and it is required to take the correct one from the original image. The application achieves this by matching the image names of the predicted images with those in the folder that contains the original images.

The user is requested to import the UWB output to begin the data integration process, as described in Section 3.4.1. The corresponding code section is provided in the Appendix. Following the processing of all data, the application automatically generates a structured report in an Excel (.xls) format. This report contains essential information for each tracked component necessary for

efficient monitoring such as visual information of the element from images, object class, precise 3D location information, unique ID, and captured date and time. Table 4-15 presents samples of the generated report records. This organized report simplifies and facilitates the process for project managers to obtain needed information about tracked components by enhancing data management from multiple sources.

The application user interface is depicted in Figure 4-28. It requires data entry from users to automatically generate the report. The model was tested with two UWB data files each containing around 6,000 datapoints and 32 images from the mechanical room. The processing time for generating reports took around 11 seconds.

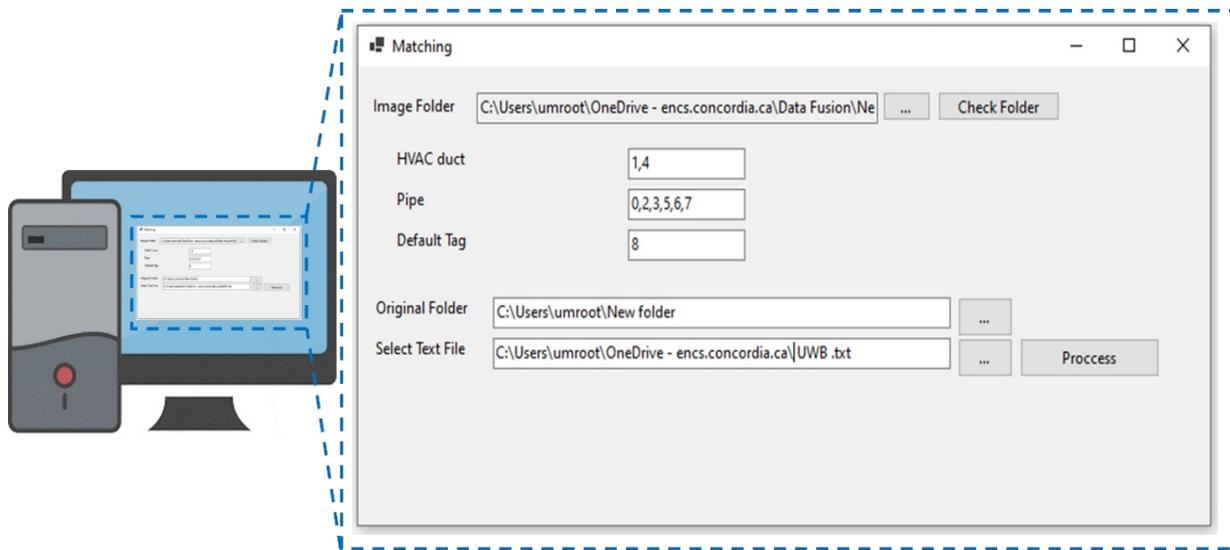


Figure 4-28. The application user interface.

Table 4-15. Record samples from the generated reports.

Record No.	Image Features				UWB Features				
	Image full path	Date taken	Image time (HH:MM:SS)	Object class	Tag ID	UWB timestamp (HH:MM:SS:MS)	Location x (m)	Location y (m)	Location z (m)
1	C:\...\IMG_1691.JPEG	2023-06-08	10:37:24	Pipe	0	10:35:59:27	-0.35	2.88	2.55
2	C:\...\IMG_1674.JPEG	2023-06-08	10:31:46	Duct	1	10:31:06:43	-1.12	3.65	4.52

3	C:\...\IMG_2527.JPEG	2023-06-09	14:23:46	Duct	4	14:23:46:26	4.27	6.94	5.30
4	C:\...\IMG_2543.JPEG	2023-06-09	14:24:21	Duct	4	14:24:21:25	4.26	7.05	5.43
5	C:\...\IMG_2514.JPEG	2023-06-09	14:23:02	Pipe	7	14:23:02:18	2.89	2.62	3.09

In this context, this framework can be considered as a beneficial method for providing both visual information and location data about tracked elements simultaneously through an organized report. This report provides essential material for progress monitoring and reporting as it includes information about the status of installed elements. It allows for the automatic extraction of the actual number of installed elements, their unique IDs, locations, and installation dates, which can contribute to automated construction progress monitoring. The report can be used in progress completion percentage by comparing the number of actual installed components with the total planned. It also provides visual information for each component, ensuring correct installation and minimizing uncertainties. The capabilities of the developed integrated method in providing comprehensive data can be extended to different monitoring domains such as material and inventory management and safety monitoring.

4.3.3 Summary and Conclusions

The shift towards automated progress monitoring and reporting using new technologies for the efficient delivery of construction projects has received significant attention in recent years. The application of vision-based techniques and deep learning algorithms for automated object recognition and RTLS for object localization has been explored in many studies. However, an individual technology cannot provide all the information necessary to determine the status of tracked elements on a job site. Therefore, technologies such as vision-based systems and RTLS can be integrated to alleviate their individual shortcomings and ensure that sufficient information about tracked resources is obtainable (Ekanayake et al., 2021; Noruwa et al., 2020; Rafiee et al., 2013).

This section developed a user-friendly application that integrates the data from a deep learning model (YOLACT++) and the UWB system for automated progress monitoring and reporting in construction projects. YOLACT++ was implemented to automatically recognize and classify tracked components within images in a timely manner, while the UWB system provided

a unique ID and location of corresponding components in real-time. The output of this integration process is a report in an Excel (.xls) format that contains essential information from both sources for an accurate and reliable assessment of the installation status of each tracked component. This report delivers each element's unique ID, location in 3D coordinates, visual information, and the capture time.

The developed method is validated in a highly occluded mechanical room filled with metallic objects in a confined space, a challenging environment for the UWB system and object recognition models due to signal interferences and occlusions. Similar to Sections 4.1.4 and 4.2.3, the developed application was implemented to monitor HVAC ducts and pipes.

An advantage of this system is its user-friendly application, which does not require human expertise to operate. For all users, the application has been simplified and only requires data entry which are images and UWB files. Moreover, this application is not limited to only HVAC ducts and pipes, and it can be generalized to different building elements in the job sites. The integrated model can deliver the necessary information to project managers in a timely manner, which is essential for efficient progress monitoring systems. The operation and computational processing of the developed method are fast, enabling reports to be generated timely on-site. The UWB generates reports in real-time and YOLACT++ offers real-time prediction capability. However, the prediction speed of the object recognition algorithm and the integrated model may be affected by computer specifications. Depending on project requirements and stakeholder needs, this report can be generated daily, weekly or monthly.

4.4 Limitations

Despite the contributions of this study, there are some limitations that need to be discussed.

- The manual generation of synthetic images from BIM models, was a time-consuming process. In the future, automated approaches should be explored for image generation with considering scenarios to reflect the real conditions of indoor environments such as lighting variations, occlusion, object complexity and the scale of objects across multiple viewpoints.
- The quality of synthetic and real image mixes within the datasets needs to be improved by taking into account not only the ratio of synthetic images, but also other factors such as

consistency in image modifications, lighting, complexity, scale, occlusion, clutter degree and viewpoints across all the datasets. This can be helpful in creating more consistent and higher quality datasets. Moreover, increasing the number of real images through advanced augmentation techniques before model training, such as flipping, rotation, cropping, and lighting variations needs to be considered.

- Object recognition models are limited to two building classes and need to be more generalized and encompass more building elements in indoor environments. The details associated with each component, including various types of pipes and added features of HVAC ducts and pipes, as separate classes, are not investigated in this study. The dataset needs to be enhanced to cover different categories of these components and their specific details. However, it is important to note that manual data labeling is labour-intensive, especially when dataset size increases. As a result, other algorithms that need less annotated data, such as semi-supervised or self-supervised learning, can be considered as potential alternatives for future studies.
- The fine-tuning of object recognition algorithms was done manually. While the selected hyperparameters achieved good performance, advanced optimization techniques that could potentially improve model efficiency were not explored.
- The conducted experiments using the UWB system were focused on static objects which are installed at their final installation locations, aligned with the study's scope. Although the UWB system can localize moving objects in jobsites and support monitoring and tracking during the construction process, this capability was not explored in this study. Tracking moving objects introduces additional complexities, as they may be located in different areas with varying obstruction levels, which may cause data loss or higher errors from signal interference or blockages.
- Optimizing the placement of UWB tags and receivers, as well as adjustments to the receiver layout at different construction stages, were not examined. The experiments need to be expanded in more indoor environments with varied settings to obtain a comprehensive performance assessment of the UWB system.
- The number of UWB tags was limited in this study and the UWB system's capability with a larger number of tags needs to be assessed. This consideration could affect the required

number of receivers, the procedure of attaching tags to elements, and the design of the layout plan to maintain the required performance.

- The use of the UWB system as the RTLS for all tracked components is expensive. Therefore, only critical components can be tracked on construction sites. To alleviate this issue, the UWB system could be integrated with other RTLS systems to expand monitoring capabilities and improve cost efficiency. Tracked objects can be categorized based on criteria such as criticality, cost, or required accuracy. This categorization enables the selection of the most suitable RTLS type for each category.
- For the integrated model, the current integration process works based on sequential, one to one matching basis for an element of each class at a time. The method needs to be generalized to enable integrating data from multiple elements of the same class simultaneously. Moreover, the integrated model is not connected to a cloud-based platform that would enable real-time data entry and output storage to be shared with responsible parties.

CHAPTER 5: SUMMARY AND CONCLUDING REMARKS

In this chapter, the conclusions of this research study are discussed. Its contributions are identified, and recommendations for potential future works are provided to improve automated progress monitoring and reporting in construction sites.

5.1 Conclusions

The main purpose of this research was to develop an automated framework for the recognition and localization of tracked objects in indoor construction environments. This can improve progress monitoring and reporting systems and facilitate decision-making during the construction phase. To achieve this, a framework consisting of three modules was developed. In the first module which is object recognition, two novel deep learning-based instance segmentation algorithms, Mask R-CNN and YOLACT++, were selected to automatically recognize tracked objects from images. To train the models, synthetic images were generated and combined with real images from construction sites.

The purpose of synthetic image generation is to overcome the lack of available real images and enhance the model's generalizability. Hence, synthetic images with different viewpoints, lighting conditions, complexity, clutter and occlusion degree were generated. To identify the optimum mix of synthetic and real images for model training, datasets with different mixes of these images were created. Transfer learning and data augmentation techniques were applied to these deep learning models to improve their robustness and overcome the overfitting problem. Their performance was then evaluated in terms of precision, recall, and mAP. This module can alleviate the problems of manual site monitoring, which are time-consuming and error-prone, to better understand the as-built status of building elements.

In addition to object recognition, it is crucial to identify the locations and unique IDs of tracked objects within construction sites in real-time. In the second module which is object localization, the UWB system was selected as RTLS that is capable of precise object localization even in high multipath environments. Different laboratory and field experiments were conducted to evaluate the system's performance and identify factors that may affect it. These factors include the height of the tags, proximity of tags to receivers, the number of active tags and receivers, and LOS and NLOS conditions. The UWB system provides accurate 2D and 3D location information

about each tracked object in real-time, assisting project managers in determining the status of tracked objects.

In the third module, a user-friendly application was developed that automatically integrates the data from both the object recognition model, YOLACT++, and the UWB system based on a matching process. This model provides comprehensive information about each tracked element, including its location information, unique ID, object class, captured time, and visual characteristics, in a detailed report. This application helps decision-makers to timely obtain required information about tracked elements in construction sites and apply corrective actions at the right time.

The uniqueness of our research as compared to other integrated methods is the strategic utilization of vision-based techniques and the UWB system for enhanced monitoring coverage, flexibility, and reliability in complex indoor sites. Employing both systems within the environment allows for capturing comprehensive data from tracked elements at the same time. Other studies have relied on fixed cameras with restricted fields of view and ideal conditions such as minimum obstruction and occlusions within job sites. However, they did not address the challenges in congested environments with high levels of obstruction, which can adversely affect both vision-based systems and RTLS. In contrast, this study fully utilizes each system's capability to capture data from different areas of job sites under various conditions. This flexibility is crucial for progress monitoring in indoor environments where a variety of tracked components are installed throughout the building. In this study, the most appropriate object recognition algorithm among a wide range of alternatives was selected. The selected algorithm, YOLACT++, is an instance segmentation algorithm which can detect and segment objects of interest in the image simultaneously. The segmentation capability precisely determines the boundaries and shapes of objects which enables for further boundary extraction and geometry analysis (Shamsollahi et al., 2022). In addition, YOLACT++ can recognize components under different viewpoints, shapes and scales due to its network architecture, making it suitable for recognizing irregular objects that exist in indoor environments. Moreover, the UWB system was selected among other available RTLS technologies because it is capable of accurate 2D/3D localization of metallic components in high multipath indoor environments.

5.2 Research Contributions

The contributions of this research study for automated progress monitoring and reporting are indicated below:

- 1) Implementing two deep learning-based instance segmentation algorithms, Mask R-CNN and YOLACT++, for automated recognition of tracked building elements from images and comparing their performance. These algorithms are capable of object classification, detection and segmentation which are complementary tasks and contribute to a more robust performance.
- 2) Employing the novel YOLACT++, incorporating DCNv2, to enhance the model capability in dealing objects with different poses, sizes and viewpoints which is an existing challenge for recognition in indoor sites. This feature is helpful in cluttered indoor environments where components can appear in numerous shapes and scales across different viewpoints. The model's performance was evaluated using both detection and segmentation metrics on a large test dataset demonstrating the model's capability in recognition of objects in different conditions.
- 3) Enhancing dataset quality through synthetic image generation using BIM models, which consider actual indoor challenges such as different lighting conditions, object complexity and scale, occlusion, clutter, and viewpoints, addressing the lack of available real images. Such synthetic images reduce the manual efforts required for capturing images through construction site visits and extensive web searches for image collection that would result in saving time and resources.
- 4) Training deep learning models on datasets with different ratios of synthetic to total images to determine the most effective mix of synthetic and real images for training. Meanwhile, to improve model generalization and mitigate overfitting, various data augmentation techniques, including color and geometric transformations were applied. Transfer learning was also applied to enhance the model's robustness. The contributions of this study will enhance the model's ability to achieve more accurate identification of tracked elements in difficult indoor conditions with images that reduce reliance on high-quality real datasets. This results in more precise progress quantification through efficient automated monitoring systems.

- 5) Assessing the UWB system for object localization in 2D and 3D planes by conducting different indoor lab and field experiments. Lab experiments focused on ideal conditions with LOS conditions to identify factors that may impact the UWB system's localization performance. These factors were then considered in subsequent field experiments, which focused on localizing metallic objects within a dense and high multipath indoor environment. The field experiments explored how factors such as LOS and NLOS conditions, tag placement and the number of receivers affect the system's localization performance. The findings demonstrated that the UWB system is capable of accurate localization in both 2D and 3D.
- 6) Selecting the appropriate object recognition algorithm and RTLS suitable for tracking complex objects in high-density work areas, and introducing a method that strategically utilizes a deep learning-based instance segmentation model, with the UWB system as complementary and reliable data acquisition sources. These two systems are able to deliver accurate and timely required visual, identification and location information about the tracked components. Implementing these systems simultaneously can improve construction progress estimation by enhancing data-driven monitoring systems.
- 7) Developing a user-friendly application that integrates data from the object recognition model and the UWB system for each tracked component and automatically creates a structured and comprehensive report. This report delivers each tracked element's unique ID, 3D location, visual data, and the capture time, which cannot be extracted solely by relying on one method. It facilitates the analysis process of different systems' outputs and addresses monitoring systems' challenges in data acquisition, documentation, and management. The developed method is not limited to specific types of object recognition model and RTLS and it can be generalized to different building classes. The application's ease of use allows users to import the input files and obtain the generated outputs without any technical expertise, which facilitates the comprehension of the project's status for decision-makers.

5.3 Future works

The developed method conducted in this research study can be extended within the following areas:

1. Object recognition models using images and deep learning models:
 - The enhancement of images with blurriness and low-lighting conditions using pre-processing techniques.
 - Utilizing automated methods for data collection from construction sites such as UAVs configured for indoor settings to mitigate manual data capturing which is time consuming and labor-intensive.
 - Investigating image labeling at the pixel level. For example, labeling strategies for complex scenarios, such as intersecting objects, are needed to prevent negative impacts on model training due to ambiguous labeled shapes lacking semantic information. Also, pre-processing methods such as harmonizing labeling across all datasets and removing labels for very small-scaled instances can be helpful (Poucin et al., 2021).
2. Defining different classes based on the installation status of objects at the construction site to enable the model to differentiate between objects that are installed and those that are not yet installed. Object localization using the UWB system:
 - Designing a layout plan before implementing the UWB system in the BIM model. Optimizing the location and number of tags and receivers can be achieved within as-planned BIM models which leads to time and cost savings, and improved system accuracy.
 - Developing guidelines for implementing the UWB system in construction projects based on implemented experiments. This includes identifying factors to consider during the system implementation such as UWB product specifications, site layout and environmental configurations.
3. Integrated model based on deep learning-based object recognition model and the UWB system:
 - Improving the current integrated model to support continuous data flow, without the need for manual data entry from the deep learning model and the UWB system.
 - Linking the integrated model with the as-built BIM model. The format and fast generation of the reports make it suitable to be imported into a BIM model for timely updating the status of tracked components. This allows project managers to access reliable and up-to-date information of these components on job sites. In the future, there is an opportunity to integrate the developed model with progress measurement

frameworks, like percentage completion methods, to quantify the progress of MEP works.

APPENDICES

Appendix A. Section of C-sharp (C#) Code:

```
# Initialize result models and get file paths
1: resultModels = new List<ResultModel>();
2: var txtTextFilePath = (TextBox)Controls.Find("txtTextFile", false).First();
# Read UWB text data from the file path
3: var textData = readTextFile(txtTextFilePath.Text);
4: var txtOriginalFolderPath = (TextBox)Controls.Find("txtOriginalImageFolder",
false).First();
5: string originalFolder = txtOriginalFolderPath.Text;
# Process each folder in the directory
6: foreach (var folder in folders)
7: {
# ValidTags: tags specific to each object class.
# DefaultTags: tags attached to the camera.
8: List<string> validTags = getValidTags(folder);
9: List<string> defaultTags = getDefaultTags();
10: var files = Directory.GetFiles(folder);
# Process each image file in the folder
11: foreach (var file in files)
12: {
# Prepare the model for each image
13: var model = new ResultModel
14: {
15: ImageName = Path.GetFileName(file),
16: ImageFullPath = file,
17: OriginalImageFullPath = Path.Combine(originalFolder, Path.GetFileName(file))
18: };
# Check if the original image exists
19: if (File.Exists(model.OriginalImageFullPath))
```

```

20:    {
21:        DateTime photoDate = GetDateTakenFromImage(model.OriginalImageFullPath) ??
throw new InvalidOperationException("Date taken data is missing or unreadable.");
22:            DateTime photoTime = new
DateTime(1970,1,1,photoDate.Hour,photoDate.Minute,photoDate.Second);

# Find the related tag based on the closest timestamp
23:        var relatedRow = textData
24:            .Where(c => validTags.Contains(c.TagId) && (c.Date - photoTime).TotalSeconds
< 90)
25:            .OrderBy(c => Math.Abs((c.Date - photoTime).TotalSeconds)).FirstOrDefault();

# If a tag is matched, update the model
26:        if (relatedRow != null)
27:        {
28:            model.DateTaken = photoDate;
29:            model.TimeTaken = photoTime;
30:            model.TagId = relatedRow.TagId;
31:            model.DateInTextFile = relatedRow.DateNumber;
32:            model.Location = relatedRow.Location;
33:            model.Class = getFolderName(folder);
34:        }
35:        else
36:        {
# Handle cases where no tag is matched
37:            var relatedDefaultRow = textData
38:                .Where(c => defaultTags.Contains(c.TagId) && (c.Date -
photoTime).TotalSeconds < 90)
39:                .OrderBy(c => Math.Abs((c.Date - photoTime).TotalSeconds)).FirstOrDefault();
40:            if (relatedRow != null)
41:            {

```

```
42:         model.DateTaken = photoDate;
43:         model.TimeTaken = photoTime;
44:         model.TagId = relatedRow.TagId;
45:         model.DateInTextFile = relatedRow.DateNumber;
46:         model.Location = relatedRow.Location;
47:         model.Class = "Camera";
48:     }
49:     else
50:     {
51:         # Default comment if no data is found
52:         model.Class = "Not Found in Text File";
53:     }
54: }
55: }
56:     resultModels.Add(model);
57: }
58: }
# Save results to an Excel file
59: saveExcelFile();
```

Appendix B. Synthetic Image Generation Procedure

To generate synthetic images from BIM models the following steps are required:

Step 1: Select a BIM model containing elements of interest (software example: Autodesk Revit).

Step 2: Define properties of the elements in the BIM model such as shape, material, texture, and dimensions.

Step 3: Select a rendering tool for synthetic image generation (software example: Enscape Plugin in Autodesk Revit).

Step 4: Consider different image modifications to reflect real scenarios that may exist in indoor environments using BIM model and rendering tool. Sample of image modifications are:

4.1 Lighting conditions: changing lighting adjustments by adding or removing lighting sources including artificial and natural lighting and changing their intensities within the 3D models.

4.2 Complexity of elements: this included changing the complexity level of tracked elements across the images. For instance, generating images with different network arrangements of MEP elements. Some images can contain only a section of an element while others encompass the entire complex network. Also, changes in elements' shapes in the BIM model contribute to this factor as well.

4.3 Scale: scale of elements in the images can be varied to address the challenge in recognizing elements of different sizes. This diversity included scales from small to large, could all appear in a single image or represented individually across different images. Changing the scale of elements can be controlled by setting different capture points within the BIM model.

4.4 Occlusion and cluttered degree: this involves adding or removing elements within the BIM model, which are often located close to elements of interest. By this consideration, we could mimic conditions where the elements may be partially blocked or surrounded by other objects. These modifications are targeted to improve the model's ability to recognize these elements in different levels of occlusion and clutter.

4.5 Viewpoints: images with different shooting angles of the elements needs to be collected in the 3D BIM model, the viewpoints achievable in indoor locations were considered such as those from

fixed cameras on walls or human perspectives. This approach enables the model to recognize elements from various viewpoints.

Appendix C. Model Implementation Guideline

Module 1: Object recognition

Step 1: Collect real images which can be captured by sites visits, or web searches.

Step 2: Generate synthetic images as described in Appendix B.

Step3: Annotate all images using a labeling software (e.g. VGG Image Annotator) and extract labeling files in JSON or COCO format depending on the algorithm type.

Step 4: Create different mixes of synthetic and real images across various datasets.

Step 5: Store the generated image datasets in Step 4 into Google Drive.

Step 6: Select an appropriate deep learning-based object recognition algorithm (e.g. Mask R-CNN or YOLACT++).

Step 7: Select appropriate hyperparameters, with examples provided in Appendix D, for the deep learning model.

Step 8: Select appropriate data augmentation techniques including geometric and color transformations to mitigate overfitting problem and improve generalization of the model.

Step 9: Apply transfer learning using pretrained weights obtained from large datasets such as ImageNet or COCO dataset as described in Chapter 3.

Step 10: Implement the deep learning model using cloud-based platforms such as Google Colab that can be easily linked with the datasets stored in Google Drive.

Step 11: Train the model and select appropriate evaluation metrics to assess its performance.

Module 2: Object localization using UWB system

Step 1: Define tags ID numbers using Printed Circuit Board (PCB) unit.

Step 2: Select appropriate configurations such as channel type and data rate using PCB unit.

Step 3: The installation location of tags and receivers needs to be determined using several approaches. 1) Existing building drawings, which illustrate the locations of tracked elements can be utilized to identify placements. 2) BIM models can simulate various layouts and placements in

a 3D environment to ensure maximum LOS between receivers and tags. 3) Site visits and consultation with the field managers to refine these placements by considering site conditions.

Step 4: Attach tags to the tracked elements.

Step 5: Consider the following considerations for installing the receivers: 1- Ensure proper signal communication by maintaining a clear line of sight between receivers. 2- Have accessibility to their locations on the site. 3- Maintain a minimum distance of 15 centimeters from any surrounding objects. 4- Locate them higher than tags at the site.

Step 6: Fix receivers into the predefined locations as described in Steps 3 and 5.

Step 7: Connect a laptop to one of the receivers via a USB cable to initiate the UWB application.

Step 8: Start UWB data collection and after obtaining sufficient data, stop collecting data manually using the UWB application. A report will be automatically generated in a text format containing information about each tag.

Module 3: Integrated Object recognition and Localization

Step 1: Collect images and UWB data simultaneously from each element once installed at the construction sites. In addition, the tags from the same class could not be activated at once. Instead, once an element is installed in its location the tag will be activated. Image capturing from the corresponding element and the UWB data collection are initiated concurrently. After the images are captured, the tag would be turned off, removed, renumbered, and attached to other building element that is set to be installed.

Step 2: Images are analyzed by deep learning-based object recognition model. Following the model's prediction results, each predicted image is stored into distinct folders on the local computer based on its identified object class.

Step 3: An application was developed in C# within the Visual Studio environment, with a section of the computer code presented in Appendix A.

Step 4: Once the application is initiated, it requires data entry from users (as explained in next steps) to automatically generate the report.

Step 5: Specify the location on the computer system where folders with predicted images exist. Each folder includes the predicted images assigned to a specific class.

Step 6: The application reads all the folders at the specified location and lists the folder names. In addition, a “Default Tag” name will be added to the list within the application which corresponds to the tag attached to the camera, if available.

Step 7: A text box next to each element class allows the user to specify the list of tag IDs assigned to that class to obtain the list of classified tags.

Step 8: Specify the location path where the original images are stored before being processed by the deep learning model. This step is crucial because the “date taken” metadata of each image is changed after processing by the model, requiring the correct date to be retrieved from the original image.

Step 9: Import the UWB output into the application to begin the data integration process.

Step 10: Click the “process” button within the application, to begin the data integration process. The integration is based on an automated matching process that match each predicted image with the UWB data of the corresponding element. This matching is based on the alignment of the “captured time”, using the timestamp from the UWB system and the 'date taken' metadata from each image.

Step 11: Following the processing of all data, the application automatically generates a structured report in an Excel (.xls) format. This report contains necessary information for each tracked element such as visual information of the element from images, object class, precise 3D location information, unique ID, and captured date and time.

Appendix D. Sample Trial for Selecting Hyperparameters

Model specifications						Results		
No.	Training set	Test set	Epoch	Batch size	Learning rate	Overfitting	Precision	Recall
1	(60% synthetic images + 40% real images) 1143 S + 782 R	(172 synthetic images + 168 real images)	All epochs	32 and 16	0.001 and 0.01	- Colab error	- Colab error	- Colab error
2	(60% synthetic images + 40% real images) 1143 S + 782 R	(172 synthetic images + 168 real images)	Epochs 70, 50 and 30	2 and 4	0.001 and 0.01	-	Low Performance \approx 50-60	
3	(60% synthetic images + 40% real images) 1143 S + 782 R	(172 synthetic images + 168 real images)	90	2	0.01	Yes	63.48	76.68
4	(60% synthetic images + 40% real images) 1143 S + 782 R	(172 synthetic images + 168 real images)	90	2	0.001	No	58.05	78.95
5	(60% synthetic images + 40% real images) 1143 S + 782 R	(172 synthetic images + 168 real images)	110	2	0.001	No	52.39	81.08
6	(60% synthetic images + 40% real images) 1143 S + 782 R	(172 synthetic images + 168 real images)	110	4	0.01	Yes	60.10	76.11
7	(55% Real images + 45% synthetic) 782 R + 956 S	(172 synthetic images + 168 real images)	Epochs 70, 50 and 30	2 and 4	0.001 and 0.01	-	Low Performance \approx 50-60	

8	(55% Real images + 45% synthetic) 782 R + 956 S	(172 synthetic images + 168 real images)	110	2	0.01	Yes	54.95	62.89
9	(55% Real images + 45% synthetic) 782 R + 956 S	(172 synthetic images + 168 real images)	90	2	0.001	No	52.01	76.74

REFERENCES

- Abdulla W. (2017). *Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow*. *GitHub repository*. https://github.com/matterport/Mask_RCNN
- Afsari, K., Eastman, C. M., & Shelden, D. R. (2016). Cloud-based BIM data transmission: Current status and challenges. *ISARC 2016 - 33rd International Symposium on Automation and Robotics in Construction*. <https://doi.org/10.22260/isarc2016/0129>
- Agapaki, E., & Brilakis, I. (2020). CLOI-NET: Class segmentation of industrial facilities' point cloud datasets. *Advanced Engineering Informatics*, 45. <https://doi.org/10.1016/j.aei.2020.101121>
- Ahmed, M., Haas, C. T., & Haas, R. (2012). Using digital photogrammetry for pipe-works progress tracking. *Canadian Journal of Civil Engineering*, 39(9). <https://doi.org/10.1139/L2012-055>
- Akhil, R. P., & Das, B. B. (2019). Cost reduction techniques on MEP projects. In *Lecture Notes in Civil Engineering* (Vol. 25). https://doi.org/10.1007/978-981-13-3317-0_44
- Alaloul, W. S., Qureshi, A. H., Musarat, M. A., & Saad, S. (2021). Evolution of close-range detection and data acquisition technologies towards automation in construction progress monitoring. In *Journal of Building Engineering* (Vol. 43). <https://doi.org/10.1016/j.jobbe.2021.102877>
- Alarifi, A., Al-Salman, A., Alsaleh, M., Alnafessah, A., Al-Hadhrami, S., Al-Ammar, M. A., & Al-Khalifa, H. S. (2016). Ultra wideband indoor positioning technologies: Analysis and recent advances. In *Sensors (Switzerland)* (Vol. 16, Issue 5). MDPI AG. <https://doi.org/10.3390/s16050707>
- Alizadeh Salehi, S., & Yitmen, İ. (2018). Modeling and analysis of the impact of BIM-based field data capturing technologies on automated construction progress monitoring. *International Journal of Civil Engineering*, 16(12). <https://doi.org/10.1007/s40999-018-0320-1>
- Alizadehsalehi, S., & Yitmen, I. (2016). The Impact of Field Data Capturing Technologies on Automated Construction Project Progress Monitoring. *Procedia Engineering*, 161. <https://doi.org/10.1016/j.proeng.2016.08.504>
- Alshibani, A., & Moselhi, O. (2016). Productivity based method for forecasting cost & time of earthmoving operations using sampling GPS data. *Journal of Information Technology in Construction*, 21.
- Athira, M. V., & Khan, D. M. (2020). Recent Trends on Object Detection and Image Classification: A Review. *2020 International Conference on Computational Performance Evaluation, ComPE 2020*. <https://doi.org/10.1109/ComPE49325.2020.9200080>

- Awolusi, I., Marks, E., & Hallowell, M. (2018). Wearable technology for personalized construction safety monitoring and trending: Review of applicable devices. *Automation in Construction*, 85. <https://doi.org/10.1016/j.autcon.2017.10.010>
- Bardareh, & Moselhi. (2022). An integrated RFID–UWB method for indoor localization of materials in construction. *Journal of Information Technology in Construction*. <https://doi.org/10.36680/j.itcon.2022.032>
- Barrera-Animas, A. Y., & Davila Delgado, J. M. (2023). Generating real-world-like labelled synthetic datasets for construction site applications. *Automation in Construction*, 151. <https://doi.org/10.1016/j.autcon.2023.104850>
- BeSpoon. (n.d.). Retrieved August 1, 2023, from www.bespoon.com
- Boje, C., Guerriero, A., Kubicki, S., & Rezgui, Y. (2020). Towards a semantic Construction Digital Twin: Directions for future research. In *Automation in Construction* (Vol. 114). <https://doi.org/10.1016/j.autcon.2020.103179>
- Bolya, D., Zhou, C., Xiao, F., & Lee, Y. J. (2019). YOLACT: Real-time instance segmentation. *Proceedings of the IEEE International Conference on Computer Vision, 2019-October*. <https://doi.org/10.1109/ICCV.2019.00925>
- Bolya, D., Zhou, C., Xiao, F., & Lee, Y. J. (2020). YOLACT++ Better Real-Time Instance Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(2). <https://doi.org/10.1109/TPAMI.2020.3014297>
- Bosché, F., Ahmed, M., Turkan, Y., Haas, C. T., & Haas, R. (2015). The value of integrating Scan-to-BIM and Scan-vs-BIM techniques for construction monitoring using laser scanning and BIM: The case of cylindrical MEP components. *Automation in Construction*. <https://doi.org/10.1016/j.autcon.2014.05.014>
- Braun, A., Tuttas, S., Borrmann, A., & Stilla, U. (2015). A concept for automated construction progress monitoring using BIM-based geometric constraints and photogrammetric point clouds. *Journal of Information Technology in Construction*, 20.
- Cai, J., & Cai, H. (2020). Robust Hybrid Approach of Vision-Based Tracking and Radio-Based Identification and Localization for 3D Tracking of Multiple Construction Workers. *Journal of Computing in Civil Engineering*, 34(4). [https://doi.org/10.1061/\(asce\)cp.1943-5487.0000901](https://doi.org/10.1061/(asce)cp.1943-5487.0000901)
- Campagna-Wilson, J., & Botton, C. (2020). Challenges Related to 4D BIM Simulation in the Construction Industry. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12341 LNCS. https://doi.org/10.1007/978-3-030-60816-3_30
- Carranza-García, M., Torres-Mateo, J., Lara-Benítez, P., & García-Gutiérrez, J. (2021). On the performance of one-stage and two-stage object detectors in autonomous vehicles using camera data. *Remote Sensing*, 13(1). <https://doi.org/10.3390/rs13010089>

- Chen, C., Zhu, Z., & Hammad, A. (2020). Automated excavators activity recognition and productivity analysis from construction site surveillance videos. *Automation in Construction, 110*. <https://doi.org/10.1016/j.autcon.2019.103045>
- Chen, J., Bulbul, T., Taylor, J. E., & Olgun, G. (2014). A case study of embedding real-time infrastructure sensor data to BIM. *Construction Research Congress 2014: Construction in a Global Network - Proceedings of the 2014 Construction Research Congress*. <https://doi.org/10.1061/9780784413517.0028>
- Cheng, T., Venugopal, M., Teizer, J., & Vela, P. A. (2011). Performance evaluation of ultra wideband technology for construction resource location tracking in harsh environments. *Automation in Construction, 20*(8). <https://doi.org/10.1016/j.autcon.2011.05.001>
- Chin, S., Yoon, S., Kim, Y. S., Ryu, J., Choi, C., & Cho, C. Y. (2005). Realtime 4D CAD + RFID for project progress management. *Construction Research Congress 2005: Broadening Perspectives - Proceedings of the Congress*. [https://doi.org/10.1061/40754\(183\)33](https://doi.org/10.1061/40754(183)33)
- Cho, Y. K., Youn, J. H., & Martinez, D. (2010). Error modeling for an untethered ultra-wideband system for construction indoor asset tracking. *Automation in Construction, 19*(1). <https://doi.org/10.1016/j.autcon.2009.08.001>
- Chong, Y., Xu, X., Guo, N., Shu, L., & Zhang, Q. (2023). Adaptive Decentralized Cooperative Localization for Firefighters Based on UWB and Autonomous Navigation. *Applied Sciences (Switzerland), 13*(8). <https://doi.org/10.3390/app13085177>
- Czerniawski, T., & Leite, F. (2018). 3DFacilities: Annotated 3D reconstructions of building facilities. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 10863 LNCS*. https://doi.org/10.1007/978-3-319-91635-4_10
- Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H., & Wei, Y. (2017). Deformable Convolutional Networks. *Proceedings of the IEEE International Conference on Computer Vision, 2017-October*. <https://doi.org/10.1109/ICCV.2017.89>
- Dardari, D., Conti, A., Ferner, U., Giorgetti, A., & Win, M. Z. (2009). Ranging with ultrawide bandwidth signals in multipath environments. *Proceedings of the IEEE, 97*(2). <https://doi.org/10.1109/JPROC.2008.2008846>
- Decawave. (n.d.). Retrieved August 1, 2023, from www.decawave.com
- Deng, H., Hong, H., Luo, D., Deng, Y., & Su, C. (2020). Automatic Indoor Construction Process Monitoring for Tiles Based on BIM and Computer Vision. *Journal of Construction Engineering and Management, 146*(1). [https://doi.org/10.1061/\(asce\)co.1943-7862.0001744](https://doi.org/10.1061/(asce)co.1943-7862.0001744)
- Deng, J., Dong, W., Socher, R., Li, L.-J., Kai Li, & Li Fei-Fei. (2010). *ImageNet: A large-scale hierarchical image database*. <https://doi.org/10.1109/cvpr.2009.5206848>

- Duan, R., Deng, H., Tian, M., Deng, Y., & Lin, J. (2022). SODA: A large-scale open site object detection dataset for deep learning in construction. *Automation in Construction*, 142. <https://doi.org/10.1016/j.autcon.2022.104499>
- Eder, M., Moulon, P., & Guan, L. (2019). Pano Poppers: Indoor 3D Reconstruction with a Plane-Aware Network. *Proceedings - 2019 International Conference on 3D Vision, 3DV 2019*. <https://doi.org/10.1109/3DV.2019.00018>
- Ekanayake, B., Ahmadian Fard Fini, A., Wong, J. K. W., & Smith, P. (2022). A deep learning-based approach to facilitate the as-built state recognition of indoor construction works. *Construction Innovation*. <https://doi.org/10.1108/CI-05-2022-0121>
- Ekanayake, B., Wong, J. K. W., Fini, A. A. F., & Smith, P. (2021). Computer vision-based interior construction progress monitoring: A literature review and future research directions. In *Automation in Construction* (Vol. 127). <https://doi.org/10.1016/j.autcon.2021.103705>
- Everingham, M., van Gool, L., Williams, C. K. I., Winn, J., & Zisserman, A. (2007). The PASCAL visual object classes challenge 2007 (VOC2007) results. In *International Journal*.
- Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., & Zisserman, A. (2012). The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Development Kit. *Pattern Analysis, Statistical Modelling and Computational Learning, Tech. Rep.* <http://host.robots.ox.ac.uk/pascal/VOC/voc2012/html/doc/index.html>
- Fang, Y., Cho, Y. K., Zhang, S., & Perez, E. (2016). Case Study of BIM and Cloud-Enabled Real-Time RFID Indoor Localization for Construction Management Applications. *Journal of Construction Engineering and Management*, 142(7). [https://doi.org/10.1061/\(asce\)co.1943-7862.0001125](https://doi.org/10.1061/(asce)co.1943-7862.0001125)
- Fathi, H., Brilakis, I., & Vela, P. (2011). Automated 3D structure inference of civil infrastructure using a stereo camera set. *Congress on Computing in Civil Engineering, Proceedings*. [https://doi.org/10.1061/41182\(416\)15](https://doi.org/10.1061/41182(416)15)
- Feng, X., Jiang, Y., Yang, X., Du, M., & Li, X. (2019). Computer vision algorithms and hardware implementations: A survey. In *Integration* (Vol. 69). <https://doi.org/10.1016/j.vlsi.2019.07.005>
- Georgy, M., & Basily, S. Y. (2008). Using genetic algorithms in optimizing construction material delivery schedules. *Construction Innovation*. <https://doi.org/10.1108/14714170810846503>
- Ghanem, A. G., & Abdelrazig, Y. A. (2006). A framework for real-time construction project progress tracking. *Earth and Space 2006 - Proceedings of the 10th Biennial International Conference on Engineering, Construction, and Operations in Challenging Environments*. [https://doi.org/10.1061/40830\(188\)112](https://doi.org/10.1061/40830(188)112)

- Gharib, S., & Moselhi, O. (2023). A review of Computer Vision-Based Techniques for Construction Progress Monitoring. *Proceedings of the 40th International Symposium on Automation and Robotics in Construction*. <https://doi.org/10.22260/isarc2023/0071>
- Girshick, R. (2015). Fast R-CNN. *Proceedings of the IEEE International Conference on Computer Vision*. <https://doi.org/10.1109/ICCV.2015.169>
- Golkhoo, F. (2020). *Material Management Framework based on Near Real-Time Monitoring of Construction Operations* [Concordia University]. https://spectrum.library.concordia.ca/id/eprint/987386/1/Golkhoo_PhD_F2020.pdf
- Golparvar-Fard, M., Bohn, J., Teizer, J., Savarese, S., & Peña-Mora, F. (2011). Evaluation of image-based modeling and laser scanning accuracy for emerging automated performance monitoring techniques. *Automation in Construction*, 20(8). <https://doi.org/10.1016/j.autcon.2011.04.016>
- Golparvar-Fard, M., Peña-Mora, F., & Savarese, S. (2015). Automated Progress Monitoring Using Unordered Daily Construction Photographs and IFC-Based Building Information Models. *Journal of Computing in Civil Engineering*, 29(1). [https://doi.org/10.1061/\(asce\)cp.1943-5487.0000205](https://doi.org/10.1061/(asce)cp.1943-5487.0000205)
- Grau, D., Caldas, C. H., Haas, C. T., Goodrum, P. M., & Gong, J. (2009). Assessing the impact of materials tracking technologies on construction craft productivity. *Automation in Construction*, 18(7). <https://doi.org/10.1016/j.autcon.2009.04.001>
- Gu, Y., Lo, A., & Niemegeers, I. (2009). A survey of indoor positioning systems for wireless personal networks. *IEEE Communications Surveys and Tutorials*, 11(1). <https://doi.org/10.1109/SURV.2009.090103>
- Ham, H., Wesley, J., & Hendra. (2019). Computer vision based 3D reconstruction : A review. *International Journal of Electrical and Computer Engineering*, 9(4). <https://doi.org/10.11591/ijece.v9i4.pp2394-2402>
- Hamledari, H., McCabe, B., & Davari, S. (2017). Automated computer vision-based detection of components of under-construction indoor partitions. *Automation in Construction*. <https://doi.org/10.1016/j.autcon.2016.11.009>
- Han, K., Degol, J., & Golparvar-Fard, M. (2018). Geometry- and Appearance-Based Reasoning of Construction Progress Monitoring. *Journal of Construction Engineering and Management*, 144(2). [https://doi.org/10.1061/\(asce\)co.1943-7862.0001428](https://doi.org/10.1061/(asce)co.1943-7862.0001428)
- Han, K. K., & Golparvar-Fard, M. (2015). Appearance-based material classification for monitoring of operation-level construction progress using 4D BIM and site photologs. *Automation in Construction*, 53. <https://doi.org/10.1016/j.autcon.2015.02.007>
- Han, X. F., Laga, H., & Bennamoun, M. (2021). Image-based 3d object reconstruction: State-of-the-art and trends in the deep learning era. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(5). <https://doi.org/10.1109/TPAMI.2019.2954885>

- Hao, S., Zhou, Y., & Guo, Y. (2020). A Brief Survey on Semantic Segmentation with Deep Learning. *Neurocomputing*, 406. <https://doi.org/10.1016/j.neucom.2019.11.118>
- He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. *Proceedings of the IEEE International Conference on Computer Vision*. <https://doi.org/10.1109/ICCV.2017.322>
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016-December*. <https://doi.org/10.1109/CVPR.2016.90>
- Hou, X., Zeng, Y., & Xue, J. (2020). Detecting Structural Components of Building Engineering Based on Deep-Learning Method. *Journal of Construction Engineering and Management*. [https://doi.org/10.1061/\(asce\)co.1943-7862.0001751](https://doi.org/10.1061/(asce)co.1943-7862.0001751)
- Huang, Y., Hammad, A., Torabi, G., Ghelmani, A., & Guevremont, M. (2021). Towards Near Real-time Digital Twins of Construction Sites: Developing High LOD 4D Simulation Based on Computer Vision and RTLS. *Proceedings of the International Symposium on Automation and Robotics in Construction, 2021-November*. <https://doi.org/10.22260/isarc2021/0036>
- Huang, Y., Hammad, A., & Zhu, Z. (2021). Providing proximity alerts to workers on construction sites using Bluetooth Low Energy RTLS. *Automation in Construction*, 132. <https://doi.org/10.1016/j.autcon.2021.103928>
- Hui, L., Park, M.-W., & Brilakis, I. (2015). Automated Brick Counting for Façade Construction Progress Estimation. *Journal of Computing in Civil Engineering*, 29(6). [https://doi.org/10.1061/\(asce\)cp.1943-5487.0000423](https://doi.org/10.1061/(asce)cp.1943-5487.0000423)
- Hwang, J., Kim, J., & Chi, S. (2023). Site-optimized training image database development using web-crawled and synthetic images. *Automation in Construction*, 151. <https://doi.org/10.1016/j.autcon.2023.104886>
- Ibem, E. O., & Laryea, S. (2014). Survey of digital technologies in procurement of construction projects. In *Automation in Construction* (Vol. 46). <https://doi.org/10.1016/j.autcon.2014.07.003>
- Jiang, S., Jiang, C., & Jiang, W. (2020). Efficient structure from motion for large-scale UAV images: A review and a comparison of SfM tools. In *ISPRS Journal of Photogrammetry and Remote Sensing* (Vol. 167). <https://doi.org/10.1016/j.isprsjprs.2020.04.016>
- Jiménez, A. R., & Seco, F. (2016). Comparing Decawave and Bespoon UWB location systems: Indoor/outdoor performance analysis. *2016 International Conference on Indoor Positioning and Indoor Navigation, IPIN 2016*. <https://doi.org/10.1109/IPIN.2016.7743686>
- Jin, T., Sadeghpour, F., & Jergeas, G. (2019). Feasibility investigation and accuracy assessment for a new generation UWB tracking system. *Proceedings, Annual Conference - Canadian Society for Civil Engineering, 2019-June*.

https://legacy.csce.ca/elf/apps/CONFERENCEVIEWER/conferences/2019/pdfs/PaperPDFversion_162_0424020327.pdf

- Jung, A. B. , Wada, K., Crall, J., Tanaka, S., Graving, J., Reinders, C., & Yadav, S. (2020). *imgaug*.
- Kang, K. S., Cho, Y. W., Jin, K. H., Kim, Y. Bin, & Ryu, H. G. (2022). Application of one-stage instance segmentation with weather conditions in surveillance cameras at construction sites. *Automation in Construction*, 133. <https://doi.org/10.1016/j.autcon.2021.104034>
- Kang, Z., Yang, J., Yang, Z., & Cheng, S. (2020). A review of techniques for 3D reconstruction of indoor environments. In *ISPRS International Journal of Geo-Information* (Vol. 9, Issue 5). <https://doi.org/10.3390/ijgi9050330>
- Kim, H., Kim, H., Hong, Y. W., & Byun, H. (2018). Detecting Construction Equipment Using a Region-Based Fully Convolutional Network and Transfer Learning. *Journal of Computing in Civil Engineering*, 32(2). [https://doi.org/10.1061/\(asce\)cp.1943-5487.0000731](https://doi.org/10.1061/(asce)cp.1943-5487.0000731)
- Kim, J., Hwang, J., Chi, S., & Seo, J. O. (2020). Towards database-free vision-based monitoring on construction sites: A deep active learning approach. *Automation in Construction*. <https://doi.org/10.1016/j.autcon.2020.103376>
- Kim, J., Kim, D., Lee, S. H., & Chi, S. (2023). Hybrid DNN training using both synthetic and real construction images to overcome training data shortage. *Automation in Construction*, 149. <https://doi.org/10.1016/j.autcon.2023.104771>
- Kini, D. U. (1999). Materials Management: The Key to Successful Project Management. *Journal of Management in Engineering*. [https://doi.org/10.1061/\(asce\)0742-597x\(1999\)15:1\(30\)](https://doi.org/10.1061/(asce)0742-597x(1999)15:1(30))
- Koo, B., & Fischer, M. (2000). Feasibility Study of 4D CAD in Commercial Construction. *Journal of Construction Engineering and Management*, 126(4). [https://doi.org/10.1061/\(asce\)0733-9364\(2000\)126:4\(251\)](https://doi.org/10.1061/(asce)0733-9364(2000)126:4(251))
- Kopsida, M., & Brilakis, I. (2020). Real-Time Volume-to-Plane Comparison for Mixed Reality-Based Progress Monitoring. *Journal of Computing in Civil Engineering*, 34(4). [https://doi.org/10.1061/\(asce\)cp.1943-5487.0000896](https://doi.org/10.1061/(asce)cp.1943-5487.0000896)
- Kopsida, M., Brilakis, I., & Vela, P. (2015). A Review of Automated Construction Progress and Inspection Methods. *Proceedings of the 32nd CIB W78 Conference on Construction IT*. <https://api.semanticscholar.org/CorpusID:55475335>
- Korman, T. M., Fischer, M. A., & Tatum, C. B. (2003). Knowledge and Reasoning for MEP Coordination. *Journal of Construction Engineering and Management*, 129(6). [https://doi.org/10.1061/\(asce\)0733-9364\(2003\)129:6\(627\)](https://doi.org/10.1061/(asce)0733-9364(2003)129:6(627))
- Kristensen, J. B., Massanet Ginard, M., Jensen, O. K., & Shen, M. (2019). Non-Line-of-Sight Identification for UWB Indoor Positioning Systems using Support Vector Machines. *2019 IEEE MTT-S International Wireless Symposium, IWS 2019 - Proceedings*. <https://doi.org/10.1109/IWS.2019.8804072>

- Kropp, C., Koch, C., & König, M. (2018). Interior construction state recognition with 4D BIM registered image sequences. *Automation in Construction*, 86. <https://doi.org/10.1016/j.autcon.2017.10.027>
- Kropp, C., Koch, C., König, M., & Brilakis, I. (2012). A framework for automated delay prediction of finishing works using video data and BIM-based construction simulation. *Proceedings of the 14th International Conference on Computing in Civil and Building Engineering, January*.
- Kropp, C., König, M., & Koch, C. (2013a). Object recognition in BIM registered videos for indoor progress monitoring. *European Group for Intelligent Computing in Engineering, EG-ICE 2013 - 20th International Workshop: Intelligent Computing in Engineering*.
- Kshetrimayum, R. S. (2009). An introduction to UWB communication systems. *IEEE Potentials*, 28(2). <https://doi.org/10.1109/MPOT.2009.931847>
- Kumar, S., & Moore, K. B. (2002). The Evolution of Global Positioning System (GPS) Technology. *Journal of Science Education and Technology*. <https://doi.org/10.1023/A:1013999415003>
- Lee, A. (2015). Comparing Deep Neural Networks and Traditional Vision Algorithms in Mobile Robotics. *Pdfs.Semanticscholar.Org*.
- Li, H., Chan, G., & Skitmore, M. (2013). Integrating real time positioning systems to improve blind lifting and loading crane operations. *Construction Management and Economics*. <https://doi.org/10.1080/01446193.2012.756144>
- Li, H., Chan, G., Wong, J. K. W., & Skitmore, M. (2016). Real-time locating systems applications in construction. In *Automation in Construction* (Vol. 63). <https://doi.org/10.1016/j.autcon.2015.12.001>
- Li, H., Chen, Z., Yong, L., & Kong, S. C. W. (2005). Application of integrated GPS and GIS technology for reducing construction waste and improving construction efficiency. *Automation in Construction*. <https://doi.org/10.1016/j.autcon.2004.08.007>
- Li, Y., & Chen, J. (2022). Computer Vision–Based Counting Model for Dense Steel Pipe on Construction Sites. *Journal of Construction Engineering and Management*, 148(1). [https://doi.org/10.1061/\(asce\)co.1943-7862.0002217](https://doi.org/10.1061/(asce)co.1943-7862.0002217)
- Lin, J. J., & Golparvar-Fard, M. (2020). Construction Progress Monitoring Using Cyber-Physical Systems. In *Cyber-Physical Systems in the Built Environment*. https://doi.org/10.1007/978-3-030-41560-0_5
- Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017-January*. <https://doi.org/10.1109/CVPR.2017.106>

- Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8693 LNCS(PART 5). https://doi.org/10.1007/978-3-319-10602-1_48
- Liu, A., Lin, S., Wang, J., & Kong, X. (2022). A Succinct Method for Non-Line-of-Sight Mitigation for Ultra-Wideband Indoor Positioning System. *Sensors*, 22(21). <https://doi.org/10.3390/s22218247>
- Liu, C., Kadja, T., & Chodavarapu, V. P. (2022). Experimental Evaluation of Sensor Fusion of Low-Cost UWB and IMU for Localization under Indoor Dynamic Testing Conditions. *Sensors*, 22(21). <https://doi.org/10.3390/s22218156>
- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). SSD: Single shot multibox detector. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9905 LNCS. https://doi.org/10.1007/978-3-319-46448-0_2
- Liu, X., Deng, Z., & Yang, Y. (2019). Recent progress in semantic image segmentation. *Artificial Intelligence Review*, 52(2). <https://doi.org/10.1007/s10462-018-9641-3>
- Lu, Q., & Lee, S. (2017). Image-Based Technologies for Constructing As-Is Building Information Models for Existing Buildings. *Journal of Computing in Civil Engineering*, 31(4). [https://doi.org/10.1061/\(asce\)cp.1943-5487.0000652](https://doi.org/10.1061/(asce)cp.1943-5487.0000652)
- Lu, W., Huang, G. Q., & Li, H. (2011). Scenarios for applying RFID technology in construction project management. *Automation in Construction*. <https://doi.org/10.1016/j.autcon.2010.09.007>
- Ma, J. W., 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. <https://doi.org/10.1016/j.autcon.2020.103144>
- Ma, Z., & Liu, S. (2018). A review of 3D reconstruction techniques in civil engineering and their applications. In *Advanced Engineering Informatics* (Vol. 37). <https://doi.org/10.1016/j.aei.2018.05.005>
- Maalek, R., & Sadeghpour, F. (2013). Accuracy assessment of Ultra-Wide Band technology in tracking static resources in indoor construction scenarios. *Automation in Construction*, 30. <https://doi.org/10.1016/j.autcon.2012.10.005>
- Man, K., & Chahl, J. (2022). A Review of Synthetic Image Data and Its Use in Computer Vision. In *Journal of Imaging* (Vol. 8, Issue 11). <https://doi.org/10.3390/jimaging8110310>
- Martinez, P., Barkokebas, B., Hamzeh, F., Al-Hussein, M., & Ahmad, R. (2021). A vision-based approach for automatic progress tracking of floor paneling in offsite construction facilities. *Automation in Construction*, 125. <https://doi.org/10.1016/j.autcon.2021.103620>

- Matthews, J., Love, P. E. D., Heinemann, S., Chandler, R., Rumsey, C., & Olatunj, O. (2015). Real time progress management: Re-engineering processes for cloud-based BIM in construction. *Automation in Construction*, 58. <https://doi.org/10.1016/j.autcon.2015.07.004>
- Mautz, R. (2012). *Indoor positioning technologies*. <https://doi.org/10.3929/ethz-a-007313554>
- Mirzaei, K., Arashpour, M., Asadi, E., Masoumi, H., Bai, Y., & Behnood, A. (2022). 3D point cloud data processing with machine learning for construction and infrastructure applications: A comprehensive review. In *Advanced Engineering Informatics* (Vol. 51). <https://doi.org/10.1016/j.aei.2021.101501>
- Mirzaei, K., Arashpour, M., Asadi, E., Masoumi, H., Mahdiyar, A., & Gonzalez, V. (2023). End-to-end point cloud-based segmentation of building members for automating dimensional quality control. *Advanced Engineering Informatics*, 55. <https://doi.org/10.1016/j.aei.2023.101878>
- Mohammadali, K., Mazdak, N. B., & Osama, M. (2023). Proximity Detection on Construction Sites, Using Bluetooth Low Energy Beacons. *Lecture Notes in Civil Engineering*, 240. https://doi.org/10.1007/978-981-19-0507-0_20
- Mohanty, L., Chae, S., & Yang, Y. (2020). Identifying productive working patterns at construction sites using BLE sensor networks. *Developments in the Built Environment*, 4. <https://doi.org/10.1016/j.dibe.2020.100025>
- Montaser, A., Bakry, I., Alshibani, A., & Moselhi, O. (2012). Estimating productivity of earthmoving operations using spatial technologies. *Canadian Journal of Civil Engineering*, 39(9). <https://doi.org/10.1139/L2012-059>
- Montaser, A., & Moselhi, O. (2012a). RFID+ for tracking earthmoving operations. *Construction Research Congress 2012: Construction Challenges in a Flat World, Proceedings of the 2012 Construction Research Congress*. <https://doi.org/10.1061/9780784412329.102>
- Montaser, A., & Moselhi, O. (2012c). RFID and BIM for automated progress reporting. In *AACE International Transactions, 56th Annual Meeting, San Antonio, Texas, United States, Pp. 08-11 July. 2012., 1*, 8–11. <https://library.aacei.org/v15/cgi-bin/litsearch.pl>
- Montaser, A., & Moselhi, O. (2014). RFID indoor location identification for construction projects. *Automation in Construction*, 39. <https://doi.org/10.1016/j.autcon.2013.06.012>
- Moselhi, O., Bardareh, H., & Zhu, Z. (2020). Automated data acquisition in construction with remote sensing technologies. *Applied Sciences (Switzerland)*. <https://doi.org/10.3390/APP10082846>
- Motamedi, A., Soltani, M. M., & Hammad, A. (2013). Localization of RFID-equipped assets during the operation phase of facilities. *Advanced Engineering Informatics*, 27(4). <https://doi.org/10.1016/j.aei.2013.07.001>
- Murphy, K., Torralba, A., Eaton, D., & Freeman, W. (2006). *Object Detection and Localization Using Local and Global Features*. https://doi.org/10.1007/11957959_20

- Natephra, W., & Motamedi, A. (2019). BIM-based live sensor data visualization using virtual reality for monitoring indoor conditions. *Intelligent and Informed - Proceedings of the 24th International Conference on Computer-Aided Architectural Design Research in Asia, CAADRIA 2019*, 2. <https://doi.org/10.52842/conf.caadria.2019.2.191>
- Nath, N. D., & Behzadan, A. H. (2019). Deep Learning Models for Content-Based Retrieval of Construction Visual Data. *Computing in Civil Engineering 2019: Data, Sensing, and Analytics - Selected Papers from the ASCE International Conference on Computing in Civil Engineering 2019*. <https://doi.org/10.1061/9780784482438.009>
- Nath, N. D., & Behzadan, A. H. (2020). Deep Convolutional Networks for Construction Object Detection Under Different Visual Conditions. *Frontiers in Built Environment*, 6. <https://doi.org/10.3389/fbuil.2020.00097>
- National Institute of Building Sciences (NIBS), B. alliance-common, & Tools, building information model files and. (2017). *No Title*. https://www.nibs.org/?page=bsa_commonbimfiles
- Navon, R., & Shpatnitsky, Y. (2005). Field Experiments in Automated Monitoring of Road Construction. *Journal of Construction Engineering and Management*, 131(4). [https://doi.org/10.1061/\(asce\)0733-9364\(2005\)131:4\(487\)](https://doi.org/10.1061/(asce)0733-9364(2005)131:4(487))
- Nieto, A., Lpez, D., & Brea, V. (2012). Towards the Optimal Hardware Architecture for Computer Vision. In *Machine Vision - Applications and Systems*. <https://doi.org/10.5772/34023>
- Niu, L., Cong, W., Liu, L., Hong, Y., Zhang, B., Liang, J., & Zhang, L. (2021). *Making Images Real Again: A Comprehensive Survey on Deep Image Composition*. <http://arxiv.org/abs/2106.14490>
- Noruwa, B. I., Arewa, A. O., & Merschbrock, C. (2020). Effects of emerging technologies in minimising variations in construction projects in the UK. *International Journal of Construction Management*. <https://doi.org/10.1080/15623599.2020.1772530>
- Oh, M., Lee, J., Hong, S. W., & Jeong, Y. (2015). Integrated system for BIM-based collaborative design. *Automation in Construction*, 58. <https://doi.org/10.1016/j.autcon.2015.07.015>
- O'Mahony, N., Campbell, S., Carvalho, A., Harapanahalli, S., Hernandez, G. V., Krpalkova, L., Riordan, D., & Walsh, J. (2019). Deep Learning vs . Traditional Computer Vision. *Computer Vision Conference (CVC)*, Cv, 128–144.
- Omar, H., Mahdjoubi, L., & Kheder, G. (2018). Towards an automated photogrammetry-based approach for monitoring and controlling construction site activities. *Computers in Industry*, 98. <https://doi.org/10.1016/j.compind.2018.03.012>
- Omar, T., & Nehdi, M. L. (2016). Data acquisition technologies for construction progress tracking. In *Automation in Construction*. <https://doi.org/10.1016/j.autcon.2016.06.016>

- Oquab, M., Bottou, L., Laptev, I., & Sivic, J. (2014). Learning and transferring mid-level image representations using convolutional neural networks. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.
<https://doi.org/10.1109/CVPR.2014.222>
- Padilla, R., Netto, S. L., & Da Silva, E. A. B. (2020). A Survey on Performance Metrics for Object-Detection Algorithms. *International Conference on Systems, Signals, and Image Processing*. <https://doi.org/10.1109/IWSSIP48289.2020.9145130>
- Pal, A., & Hsieh, S. H. (2021). Deep-learning-based visual data analytics for smart construction management. In *Automation in Construction* (Vol. 131).
<https://doi.org/10.1016/j.autcon.2021.103892>
- Pal, A., Lin, J. J., Hsieh, S. H., & Golparvar-Fard, M. (2023). Automated vision-based construction progress monitoring in built environment through digital twin. In *Developments in the Built Environment* (Vol. 16). Elsevier Ltd.
<https://doi.org/10.1016/j.dibe.2023.100247>
- Pal, A., Lin, J. J., Hsieh, S. H., & Golparvar-Fard, M. (2024). Activity-level construction progress monitoring through semantic segmentation of 3D-informed orthographic images. *Automation in Construction*, 157. <https://doi.org/10.1016/j.autcon.2023.105157>
- Paneru, S., & Jeelani, I. (2021). Computer vision applications in construction: Current state, opportunities & challenges. In *Automation in Construction* (Vol. 132).
<https://doi.org/10.1016/j.autcon.2021.103940>
- Park, J., Kim, K., & Cho, Y. K. (2017). Framework of Automated Construction-Safety Monitoring Using Cloud-Enabled BIM and BLE Mobile Tracking Sensors. *Journal of Construction Engineering and Management*, 143(2). [https://doi.org/10.1061/\(asce\)co.1943-7862.0001223](https://doi.org/10.1061/(asce)co.1943-7862.0001223)
- Park, J. W., Yang, X., Cho, Y. K., & Seo, J. (2017). Improving dynamic proximity sensing and processing for smart work-zone safety. *Automation in Construction*, 84.
<https://doi.org/10.1016/j.autcon.2017.08.025>
- Park, M. W., & Brilakis, I. (2016). Continuous localization of construction workers via integration of detection and tracking. *Automation in Construction*, 72.
<https://doi.org/10.1016/j.autcon.2016.08.039>
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., ... Chintala, S. (2019). PyTorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems*, 32.
- Patel, T., Guo, B. H. W., & Zou, Y. (2021). A scientometric review of construction progress monitoring studies. In *Engineering, Construction and Architectural Management*.
<https://doi.org/10.1108/ECAM-10-2020-0799>

- Pour Rahimian, F., Seyedzadeh, S., Oliver, S., Rodriguez, S., & Dawood, N. (2020). On-demand monitoring of construction projects through a game-like hybrid application of BIM and machine learning. *Automation in Construction*.
<https://doi.org/10.1016/j.autcon.2019.103012>
- Pradhananga, N., & Teizer, J. (2013). Automatic spatio-temporal analysis of construction site equipment operations using GPS data. *Automation in Construction*, 29.
<https://doi.org/10.1016/j.autcon.2012.09.004>
- Rafiee, M., Siddiqui, H., & Hammad, A. (2013). Improving indoor security surveillance by fusing data from BIM, UWB and video. *ISARC 2013 - 30th International Symposium on Automation and Robotics in Construction and Mining, Held in Conjunction with the 23rd World Mining Congress*, 742–752. <https://doi.org/10.22260/isarc2013/0081>
- Rao, A. S., Radanovic, M., Liu, Y., Hu, S., Fang, Y., Khoshelham, K., Palaniswami, M., & Ngo, T. (2022). Real-time monitoring of construction sites: Sensors, methods, and applications. *Automation in Construction*, 136, 104099. <https://doi.org/10.1016/j.autcon.2021.104099>
- Razavi, S. N., & Moselhi, O. (2012). GPS-less indoor construction location sensing. *Automation in Construction*, 28. <https://doi.org/10.1016/j.autcon.2012.05.015>
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016-December*. <https://doi.org/10.1109/CVPR.2016.91>
- Reja, V. K., Varghese, K., & Ha, Q. P. (2022). Computer vision-based construction progress monitoring. In *Automation in Construction* (Vol. 138).
<https://doi.org/10.1016/j.autcon.2022.104245>
- Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems, 2015-January*.
- Ruiz, A. R. J., & Granja, F. S. (2017). Comparing Ubisense, BeSpoon, and DecaWave UWB Location Systems: Indoor Performance Analysis. *IEEE Transactions on Instrumentation and Measurement*, 66(8). <https://doi.org/10.1109/TIM.2017.2681398>
- Sabri, N., Aljunid, S. A., & Yahya, A. (2012). *Performance evaluation of wireless sensor network channel in agricultural application*.
<https://www.researchgate.net/publication/306204263>
- Sami Ur Rehman, M., Shafiq, M. T., & Ullah, F. (2022). Automated Computer Vision-Based Construction Progress Monitoring: A Systematic Review. In *Buildings* (Vol. 12, Issue 7).
<https://doi.org/10.3390/buildings12071037>
- Sang, C. L., Steinhagen, B., Homburg, J. D., Adams, M., Hesse, M., & Rückert, U. (2020). Identification of NLOS and multi-path conditions in UWB localization using machine

- learning methods. *Applied Sciences (Switzerland)*, 10(11).
<https://doi.org/10.3390/app10113980>
- Sardroud, J. M. (2014). Applications of RFID technology in the construction industry: An academic literature review. In *Radio-Frequency Identification (RFID): Emerging Technologies, Applications and Improvement Strategies*.
- Schjørring, A., Cretu-Sircu, A. L., Rodriguez, I., Cederholm, P., Berardinelli, G., & Mogensen, P. (2022). Performance Evaluation of a UWB Positioning System Applied to Static and Mobile Use Cases in Industrial Scenarios. *Electronics (Switzerland)*, 11(20).
<https://doi.org/10.3390/electronics11203294>
- Seo, J., Han, S., Lee, S., & Kim, H. (2015). Computer vision techniques for construction safety and health monitoring. *Advanced Engineering Informatics*, 29(2).
<https://doi.org/10.1016/j.aei.2015.02.001>
- Sergi, I., Montanaro, T., Shumba, A. T., Gammariello, M. C., Imperiale, E., & Patrono, L. (2022). A Literature Review on Outdoor Localization Systems based on the Bluetooth Technology. *2022 7th International Conference on Smart and Sustainable Technologies, SpliTech 2022*. <https://doi.org/10.23919/SpliTech55088.2022.9854361>
- Shahi, A., Aryan, A., West, J. S., Haas, C. T., & Haas, R. C. G. (2012). Deterioration of UWB positioning during construction. *Automation in Construction*, 24.
<https://doi.org/10.1016/j.autcon.2012.02.009>
- Shahi, A., Safa, M., Haas, C. T., & West, J. S. (2015). Data Fusion Process Management for Automated Construction Progress Estimation. *Journal of Computing in Civil Engineering*.
[https://doi.org/10.1061/\(asce\)cp.1943-5487.0000436](https://doi.org/10.1061/(asce)cp.1943-5487.0000436)
- Shahi, A., West, J. S., & Haas, C. T. (2013). Onsite 3D marking for construction activity tracking. *Automation in Construction*, 30, 136–143.
<https://doi.org/10.1016/j.autcon.2012.11.027>
- Shamsollahi, D., Moselhi, O., & Khorasani, k. (2023). Integrated Framework Using Computer Vision and Ultra-Wide Band Techniques for Progress Reporting in Construction Projects. *Annual Conference - Canadian Society for Civil Engineering*.
- Shamsollahi, D., Moselhi, O., & Khorasani, K. (2021). A Timely Object Recognition Method for Construction using the Mask R-CNN Architecture. *Proceedings of the 38th International Symposium on Automation and Robotics in Construction (ISARC)*.
<https://doi.org/10.22260/isarc2021/0052>
- Shamsollahi, D., Moselhi, O., & Khorasani, K. (2022). Construction Progress Monitoring and Reporting using Computer Vision Techniques - A Review. *Proceedings of the International Symposium on Automation and Robotics in Construction, 2022-July*.
<https://doi.org/https://doi.org/10.22260/ISARC2022/0064>

- Shamsollahi D., Moselhi O., & Khorasani K. (2024). Automated Detection and Segmentation of Mechanical, Electrical, and Plumbing (MEP) Components in Indoor Environments by Using the YOLACT++ Architecture. *Journal of Construction Engineering and Management*, 150(8). <https://doi.org/10.1061/JCEMD4.COENG-15115>
- Shamsollahi, D., Moselhi, O., & Khorasani, K. (2024). A Data integration method using a deep learning algorithm and real-time locating system (RTLS) for automated construction progress monitoring and reporting. *Automation in Construction*.
- Shekhar, S., Shukla, P., & Das, B. B. (2021). Developing a Standard Template for Activity Linkage and Resource Estimation of MEP Works. *Lecture Notes in Civil Engineering*, 105. https://doi.org/10.1007/978-981-15-8293-6_40
- Siddiqui, H., Vahdatikhaki, F., & Hammad, A. (2019). Case study on application of wireless ultra-wideband technology for tracking equipment on a congested site. *Journal of Information Technology in Construction*, 24. <http://www.itcon.org/2019/10>
- Soltani, M. M., Zhu, Z., & Hammad, A. (2016). Automated annotation for visual recognition of construction resources using synthetic images. *Automation in Construction*, 62. <https://doi.org/10.1016/j.autcon.2015.10.002>
- Soltani, M. M., Zhu, Z., & Hammad, A. (2018). Framework for Location Data Fusion and Pose Estimation of Excavators Using Stereo Vision. *Journal of Computing in Civil Engineering*, 32(6). [https://doi.org/10.1061/\(asce\)cp.1943-5487.0000783](https://doi.org/10.1061/(asce)cp.1943-5487.0000783)
- Song, J., Haas, C. T., Caldas, C., Ergen, E., & Akinci, B. (2006). Automating the task of tracking the delivery and receipt of fabricated pipe spools in industrial projects. *Automation in Construction*, 15(2). <https://doi.org/10.1016/j.autcon.2005.03.001>
- Song, L., Mohammed, T., Stayshich, D., & Eldin, N. (2015). A Cost Effective Material Tracking and Locating Solution for Material Laydown Yard. *Procedia Engineering*, 123. <https://doi.org/10.1016/j.proeng.2015.10.106>
- Sun, C., Jiang, F., & Jiang, S. (2013). Research on RFID applications in construction industry. *Journal of Networks*, 8(5). <https://doi.org/10.4304/jnw.8.5.1221-1228>
- Techasartikul, N., & Mashita, T. (2022). Instance Segmentation of Images Above the Ceiling Using Mask R-CNN. <https://doi.org/10.1109/icecet52533.2021.9698438>
- Teizer, J. (2015). Status quo and open challenges in vision-based sensing and tracking of temporary resources on infrastructure construction sites. *Advanced Engineering Informatics*, 29(2). <https://doi.org/10.1016/j.aei.2015.03.006>
- Teizer, J., Golovina, O., Wang, D., & Pradhananga, N. (2015). Automated collection, identification, localization, and analysis of worker-related proximity hazard events in heavy construction equipment operation. *32nd International Symposium on Automation and Robotics in Construction and Mining: Connected to the Future, Proceedings*. <https://doi.org/10.22260/isarc2015/0005>

- Teizer, J., Neve, H., Li, H., Wandahl, S., König, J., Ochner, B., König, M., & Lerche, J. (2020). Construction resource efficiency improvement by Long Range Wide Area Network tracking and monitoring. *Automation in Construction*, 116. <https://doi.org/10.1016/j.autcon.2020.103245>
- Teizer, J., Venugopal, M., & Walia, A. (2008a). Ultrawideband for automated real-time three-dimensional location sensing for workforce, equipment, and material positioning and tracking. *Transportation Research Record*, 2081. <https://doi.org/10.3141/2081-06>
- Teizer, J., Venugopal, M., & Walia, A. (2008b). Ultrawideband for automated real-time three-dimensional location sensing for workforce, equipment, and material positioning and tracking. *Transportation Research Record*, 2081, 56–64. <https://doi.org/10.3141/2081-06>
- Teizer, J., Wolf, M., Golovina, O., Perschewski, M., Propach, M., Neges, M., & König, M. (2017). Internet of Things (IoT) for integrating environmental and localization data in Building Information Modeling (BIM). *ISARC 2017 - Proceedings of the 34th International Symposium on Automation and Robotics in Construction*. <https://doi.org/10.22260/isarc2017/0084>
- Teo, Y. H., Yap, J. H., An, H., Yu, S. C. M., Zhang, L., Chang, J., & Cheong, K. H. (2022). Enhancing the MEP Coordination Process with BIM Technology and Management Strategies. In *Sensors* (Vol. 22, Issue 13). <https://doi.org/10.3390/s22134936>
- Tomasi, R., Sottile, F., Pastrone, C., Mozumdar, M. M. R., Osello, A., & Lavagno, L. (2015). Leveraging BIM Interoperability for UWB-Based WSN Planning. *IEEE Sensors Journal*, 15(10). <https://doi.org/10.1109/JSEN.2015.2451736>
- Topak, F., Pekerli, M. K., & Tanyer, A. M. (2018). Technological Viability Assessment of Bluetooth Low Energy Technology for Indoor Localization. *Journal of Computing in Civil Engineering*, 32(5). [https://doi.org/10.1061/\(asce\)cp.1943-5487.0000778](https://doi.org/10.1061/(asce)cp.1943-5487.0000778)
- Tuttas, S., Braun, A., Borrmann, A., & Stilla, U. (2017). Acquisition and Consecutive Registration of Photogrammetric Point Clouds for Construction Progress Monitoring Using a 4D BIM. *Photogrammetrie, Fernerkundung, Geoinformation*, 85(1). <https://doi.org/10.1007/s41064-016-0002-z>
- Ubisense*. Retrieved August 1, 2023, from www.ubisense.net
- Umer, W., & Siddiqui, M. K. (2020). Use of ultra wide band real-time location system on construction jobsites: Feasibility study and deployment alternatives. *International Journal of Environmental Research and Public Health*, 17(7). <https://doi.org/10.3390/ijerph17072219>
- Valero, E., Adán, A., & Cerrada, C. (2015). Evolution of RFID applications in construction: A literature review. *Sensors (Switzerland)*, 15(7). <https://doi.org/10.3390/s150715988>

- Wang, J., Ma, Y., Zhang, L., Gao, R. X., & Wu, D. (2018). Deep learning for smart manufacturing: Methods and applications. *Journal of Manufacturing Systems*. <https://doi.org/10.1016/j.jmsy.2018.01.003>
- Wang, J., Raja, A. K., & Pang, Z. (2015). Prototyping and Experimental Comparison of IR-UWB Based High Precision Localization Technologies. *2015 IEEE 12th Intl Conf on Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom)*, 1187–1192. <https://doi.org/10.1109/UIC-ATC-ScalCom-CBDCCom-IoP.2015.216>
- Wang, Q., & Kim, M. K. (2019). Applications of 3D point cloud data in the construction industry: A fifteen-year review from 2004 to 2018. In *Advanced Engineering Informatics* (Vol. 39). <https://doi.org/10.1016/j.aei.2019.02.007>
- Wang, Q., Tan, Y., & Mei, Z. (2020). Computational Methods of Acquisition and Processing of 3D Point Cloud Data for Construction Applications. *Archives of Computational Methods in Engineering*, 27(2). <https://doi.org/10.1007/s11831-019-09320-4>
- Wang, X., & Love, P. E. D. (2012). BIM + AR: Onsite information sharing and communication via advanced visualization. *Proceedings of the 2012 IEEE 16th International Conference on Computer Supported Cooperative Work in Design, CSCWD 2012*. <https://doi.org/10.1109/CSCWD.2012.6221920>
- Wang, Z., Zhang, Q., Yang, B., Wu, T., Lei, K., Zhang, B., & Fang, T. (2021). Vision-Based Framework for Automatic Progress Monitoring of Precast Walls by Using Surveillance Videos during the Construction Phase. *Journal of Computing in Civil Engineering*. [https://doi.org/10.1061/\(asce\)cp.1943-5487.0000933](https://doi.org/10.1061/(asce)cp.1943-5487.0000933)
- Wei, S., Gao, G., Ke, Z., Fan, G., Liu, Y., & Gu, M. (2022). *Semantic Segmentation of building point clouds based on Point Transformer and IFC*. <https://doi.org/10.7146/aul.455.c197>
- 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. <https://doi.org/10.1016/j.autcon.2023.105048>
- Wei Wei, Yujie Lu, Tao Zhong, Peixian Li, & Bo Liu. (2022). Integrated vision-based automated progress monitoring of indoor construction using mask region-based convolutional neural networks and BIM. *Automation in Construction*. <https://doi.org/https://doi.org/10.1016/j.autcon.2022.104327>
- Wing, R. (2006). RFID applications in construction and facilities management. *Electronic Journal of Information Technology in Construction*, 11.
- Wong, J., Wang, X., Li, H., Chan, G., & Li, H. (2014). A review of cloud-based bim technology in the construction sector. In *Journal of Information Technology in Construction* (Vol. 19).

- Wu, H., Marshall, A., & Yu, W. (2007). Path planning and following algorithms in an indoor navigation model for visually impaired. *Second International Conference on Internet Monitoring and Protection, ICIMP 2007*. <https://doi.org/10.1109/ICIMP.2007.31>
- Wu, X., Sahoo, D., & Hoi, S. C. H. (2020). Recent advances in deep learning for object detection. *Neurocomputing*, 396. <https://doi.org/10.1016/j.neucom.2020.01.085>
- Xia, L., Retscher, G., Tian, H., & Tian, H. (2010). A case study on the feasibility and performance of an UWB-AoA real time location system for resources management of civil construction projects. *Journal of Applied Geodesy*, 4(1), 23–32. <https://doi.org/10.1515/jag.2010.003>
- 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)
- Xiao, B., Zhang, Y., Chen, Y., & Yin, X. (2021). A semi-supervised learning detection method for vision-based monitoring of construction sites by integrating teacher-student networks and data augmentation. *Advanced Engineering Informatics*, 50. <https://doi.org/10.1016/j.aei.2021.101372>
- Xie, Y., Li, S., Liu, T., & Cai, Y. (2023). As-built BIM reconstruction of piping systems using PipeNet. *Automation in Construction*, 147. <https://doi.org/10.1016/j.autcon.2022.104735>
- Xu, H., Wu, M., Li, P., Zhu, F., & Wang, R. (2018). An RFID indoor positioning algorithm based on support vector regression. *Sensors (Switzerland)*, 18(5). <https://doi.org/10.3390/s18051504>
- Xu, S., Wang, J., Shou, W., Ngo, T., Sadick, A. M., & Wang, X. (2020). Computer Vision Techniques in Construction: A Critical Review. *Archives of Computational Methods in Engineering*. <https://doi.org/10.1007/s11831-020-09504-3>
- Xu, Y., Tong, X., & Stilla, U. (2021). Voxel-based representation of 3D point clouds: Methods, applications, and its potential use in the construction industry. In *Automation in Construction* (Vol. 126). <https://doi.org/10.1016/j.autcon.2021.103675>
- Xue, J., Hou, X., & Zeng, Y. (2021). Review of image-based 3d reconstruction of building for automated construction progress monitoring. *Applied Sciences (Switzerland)*, 11(17). <https://doi.org/10.3390/app11177840>
- Xuehui, A., Li, Z., Zuguang, L., Chengzhi, W., Pengfei, L., & Zhiwei, L. (2021). Dataset and benchmark for detecting moving objects in construction sites. *Automation in Construction*, 122. <https://doi.org/10.1016/j.autcon.2020.103482>
- Yang, J., Shi, Z., & Wu, Z. (2016). Vision-based action recognition of construction workers using dense trajectories. *Advanced Engineering Informatics*, 30(3). <https://doi.org/10.1016/j.aei.2016.04.009>

- Yang, Q., Dong, E., & Zhu, L. (2020). An Instance Segmentation Algorithm Based on Improved Mask R-CNN. *Proceedings - 2020 Chinese Automation Congress, CAC 2020*. <https://doi.org/10.1109/CAC51589.2020.9326740>
- Yao, L., Yao, L., & Wu, Y. W. (2021). Analysis and improvement of indoor positioning accuracy for uwb sensors. *Sensors, 21*(17). <https://doi.org/10.3390/s21175731>
- Yarmohammadi, S., & Ashuri, B. (2015). Exploring the approaches in the implementation of BIM-based MEP coordination in the USA. *Journal of Information Technology in Construction, 20*. <http://www.itcon.org/2015/22>
- Yates, J. K., & Epstein, A. (2006). Avoiding and minimizing construction delay claim disputes in relational contracting. In *Journal of Professional Issues in Engineering Education and Practice* (Vol. 132, Issue 2). [https://doi.org/10.1061/\(ASCE\)1052-3928\(2006\)132:2\(168\)](https://doi.org/10.1061/(ASCE)1052-3928(2006)132:2(168))
- Ying, H. Q., & Lee, S. (2019). A mask R-CNN based approach to automatically construct As-is IFC BIM objects from digital images. *Proceedings of the 36th International Symposium on Automation and Robotics in Construction, ISARC 2019*. <https://doi.org/10.22260/isarc2019/0103>
- Zhang, C., Hammad, A., & Rodriguez, S. (2012). Crane Pose Estimation Using UWB Real-Time Location System. *Journal of Computing in Civil Engineering, 26*(5). [https://doi.org/10.1061/\(asce\)cp.1943-5487.0000172](https://doi.org/10.1061/(asce)cp.1943-5487.0000172)
- Zhang, C., Shen, W., & Ye, Z. (2020). Technical feasibility analysis on applying ultra-wide band technology in construction progress monitoring. *International Journal of Construction Management, 20*(2). <https://doi.org/10.1080/15623599.2020.1834928>
- Zhang, F., Sun, Y., & Wan, Q. (2020). Calibrating the error from sensor position uncertainty in TDOA-AOA localization. *Signal Processing, 166*. <https://doi.org/10.1016/j.sigpro.2019.07.006>
- Zhang, H., Zhang, Z., Gao, N., Xiao, Y., Meng, Z., & Li, Z. (2020). Cost-effective wearable indoor localization and motion analysis via the integration of UWB and IMU. *Sensors (Switzerland), 20*(2). <https://doi.org/10.3390/s20020344>
- Zhang, S., & Zhang, L. (2021). Vision-based Excavator Activity Analysis and Safety Monitoring System. *Proceedings of the International Symposium on Automation and Robotics in Construction, 2021-November*. <https://doi.org/10.22260/isarc2021/0009>
- Zhao, J., Seppänen, O., Peltokorpi, A., Badihi, B., & Olivieri, H. (2019). Real-time resource tracking for analyzing value-adding time in construction. *Automation in Construction, 104*. <https://doi.org/10.1016/j.autcon.2019.04.003>
- Zheng, Z., Zhang, Z., & Pan, W. (2020). Virtual prototyping- and transfer learning-enabled module detection for modular integrated construction. *Automation in Construction, 120*. <https://doi.org/10.1016/j.autcon.2020.103387>

- Zhou, Y., Guo, H., Ma, L., Zhang, Z., & Skitmore, M. (2021). Image-based onsite object recognition for automatic crane lifting tasks. *Automation in Construction*, 123. <https://doi.org/10.1016/j.autcon.2020.103527>
- Zhu, X., Hu, H., Lin, S., & Dai, J. (2019). Deformable convnets V2: More deformable, better results. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2019-June*. <https://doi.org/10.1109/CVPR.2019.00953>
- Zou, J., & Kim, H. (2007). Using Hue, Saturation, and Value Color Space for Hydraulic Excavator Idle Time Analysis. *Journal of Computing in Civil Engineering*. [https://doi.org/10.1061/\(asce\)0887-3801\(2007\)21:4\(238\)](https://doi.org/10.1061/(asce)0887-3801(2007)21:4(238))