# Integration of Construction Progress Monitoring Results using AI Image Recognition from Multiple Cameras onto A BIM

**Chang-Cheng Hsieh[1], Hung-Ming Chen[2], Wan-Yu Chen[3] and Ting-Yu Wu[3]**

[1]Ph.D. Candidate, Department of Civil & Construction Engineering, National Taiwan University of Science and Technology
[2]Professor, Department of Civil & Construction Engineering, National Taiwan University of Science and Technology
[3]Graduate Research Assistant, Department of Civil & Construction Engineering, National Taiwan University of Science and Technology
E-mail: D10605009@mail.ntust.edu.tw, hungming@mail.ntust.edu.tw, M10905508@mail.ntust.edu.tw, M11005501@mail.ntust.edu.tw

**Abstract**

**This study merges artificial intelligence (AI) image recognition technology with Building Information Modeling (BIM), to develop a prototype system for the automation and visualization of construction site progress control. Our focus is the strategic deployment of multiple construction site surveillance cameras using a BIM model to encompass the entirety of the construction site. Following the capture of camera images, the application of object detection techniques within AI image recognition locates all actively constructed objects in the images, subsequently identifying the construction phases to which these objects belong. By integrating multiple camera perspectives from the site into a BIM model, the results of AI detection are automatically inputted into the corresponding components of the model. Finally, real-time on-site progress information obtained from the BIM model is compared with the progress schedule, and the comparative results are visually presented on the BIM model components in distinct colors. Through this visual approach, managerial personnel can intuitively and instantly control the construction progress.**

**Keywords –**

**Deep Learning, Image Recognition, Construction Progress Management, Build Information Modeling, Automation, Visualization**

## 1 Introduction

In the project management of the traditional construction industry, monitoring the progress of the project has always been an important task [1]. In addition to having a profound awareness of the construction environment, the on-site engineer must also understand drawings. Progress data must be collected on-site, and real-time progress information must be presented in the form of text and data for project managers to refer to [2]. For less experienced site engineers, there may be cognitive standards for different construction phases. Moreover, it is not easy for people who are not familiar with engineering to convert two-dimensional drawings into three-dimensional scenes, which will cause differences in information transmission.

In order to fully support the life cycle of construction projects and interpret engineering information models through computer programs, BIM came into being. The application of BIM covers all stages of the building life cycle, including planning, design, procurement, construction, operation and maintenance, etc. [3]. In the construction stage, BIM is commonly integrated with project timelines to create a 4D model for construction simulation. Although the dynamic model exists, on-site engineers still need to update the schedule data of the model components by comparing animations with actual on-site construction conditions to achieve progress control.

Using a huge management manpower to collect and organize complex data, this traditional management method no longer seems efficient enough. Many studies have surveyed how to improve complex data processing procedures that rely on manpower [4]. In order to allow managers to perform progress management tasks more quickly and ensure that progress evaluation standards are unified, thereby making the overall management process smoother.

As technology advances, more research in artificial intelligence (AI) has made significant strides in recent years. Machine learning is a method of learning from past data and experiences to identify operational rules. Deep learning, a subset of machine learning that applies multi-layered neural networks to simulate human neuron functions, has achieved notable breakthroughs in the fields of images, videos, and speech [5-6].

The rapid development of AI image recognition technology has led to its expanding applications in the field of engineering, particularly in construction industries for the management of construction machinery, personnel, and materials. This project aims to apply AI image recognition to construction progress control, focusing on the use of object detection technology in AI image recognition to achieve automated recognition of various work progress in construction sites.

However, constrained by factors such as the site's scope, layout, component obstructions, and camera wide angle, AI image detection can only address detection within a single image, making it challenging to cover the overall area. Even with multiple cameras set up to encompass the overall area, integrating the detected results from these cameras and automating the comparison with the construction planning schedule still requires the development of effective solutions.

Therefore, this project further integrates AI image detection with BIM technology and develops a prototype system. This system, utilizing multiple cameras, applies an AI image recognition model to recognize the construction status of work items within the images. Subsequently, through the BIM model, it achieves image alignment and identifies the corresponding components, inputting the construction progress of the respective components into the BIM model. Finally, in the application program, a visual representation using a color concept is employed to present different progress states, enabling project managers to control the construction progress in real-time.

This study will establish an image-based construction progress detection model applicable to the construction phase, providing project managers with a method to assess construction progress states through image object detection. Adopt the BIM model for the integration of component construction states, automatically inputs the corresponding component progress states into the BIM model. This not only avoids variations in construction progress judgment among different personnel but also enhances the utilization of human resources, thereby improving the efficiency and convenience of on-site progress management.

Based on the background, this study plans to propose a progress object detection module used in the construction phase. By collecting images of the progress of each stage of construction, it can be used as training data. By combining common model architecture, an object detection model suitable for identifying construction progress is selected. Apply the transfer learning method and adjust model parameters to improve the recognition rate, and then find the most suitable model for construction progress detection. Finally, integrating with the BIM model at the application end will achieve component positioning, integrate multiple images of the same component for detection input, and address differences in progress judgment due to manual input and optimize human resource utilization.

The study will be divided into four phases. First is the collection of construction progress image data, followed by the training and testing of the object detection model, optimization and validation of the object detection model, and finally, the integration of BIM for automated and visualized construction progress monitoring.

In terms of data collection, cameras will be installed at the construction site to collect image data, supplemented by collecting relevant construction progress photos from online sources.

The establishment of the object detection model adopted transfer learning. This pre-trained model will undergo training and testing to seek an optimal object detection model. The parameters of the model will be systematically optimized and adjusted to enhance the accuracy of the detection model.

During the model validation and testing phase, images or videos directly obtained from construction sites will be utilized. Image recognition will be applied to identify the construction activities in the data source.

Adopting the BIM model as the data integration hub, simulating camera deployment conditions, utilizing image recognition technology to detect the construction status of components within the coverage area, and inputting it into the model components. The integrated construction progress of components can be automatically input into the scheduling system.

The prototype system will compare with the original schedule, and using different progress visualization methods, categorize the results by color. Through visualization, management personnel can grasp the progress concretely and intuitively in real time.

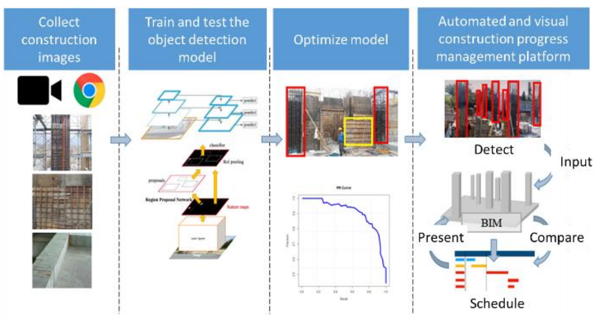


Figure 1.The process diagram of image-based construction progress detection model

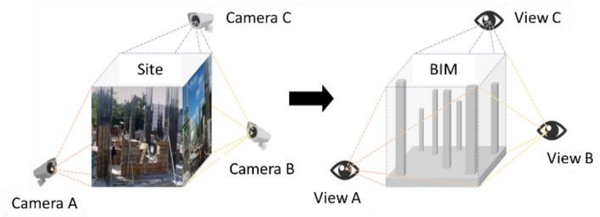


Figure 2. Demo diagram of integrates AI image detection with BIM technology

## 2 Related Research

The research topics related to this study include real-time progress monitoring and AI image recognition, especially for construction progress monitoring. The literature review for these topics are as follows.

### 2.1 Real-time Progress Monitoring

Past research has utilized various Field Data Capturing Technologies (FDCT) [7] to collect progress data from construction sites, integrating them with Building Information Modeling (BIM) to enhance construction progress monitoring. These include image recognition of construction elements, alignment with 4D models for comparison, RFID for tracking personnel and materials, UWB positioning systems for progress tracking, and laser scanners for construction environment scanning to create 4D As-Built BIM models, compared with 4D Designed BIM models [8-11].

These technologies still have limitations for improvement in progress management applications, such as the need for confirmation of work status for UWB and RFID, like verifying completion and installation. Point cloud analysis requires more time for processing to the object level for project progress.

### 2.2 AI Image Recognition for Construction Progress Monitoring

In recent years, the application of image recognition in construction through deep learning has been steadily increasing, encompassing the identification of elements like construction workers, materials, and machinery [12-13]. The studies of applying AI image recognition, for construction progress monitoring are relatively few.

Utilizing image recognition for automated progress monitoring in construction projects involves extracting features through deep learning to detect and identify construction status. Zheng et al. [14] utilized R-CNN model to automatically detect modules and identify their status, such as Hooking, Lifting, and Final Positioning.

With the affordability of cameras, gathering rich information from construction sites to achieve automated visual monitoring of construction sites becomes feasible [15]. Martinez et al. [16] utilized low-resolution CCTV images and combined deep learning methods (R-CNN) with Finite State Machines (VFSM) to identify labor and key equipment in floor manufacturing. The study presents the calculation of task duration and working hours, providing managers with clear and real-time insights into workstation progress.

The study adopts deep learning-based recognition technology to propose a framework suitable for on-site construction progress management. Data collection is conducted using fixed on-site cameras to improve the costly investment in equipment and manpower. The use of YOLO and transfer learning enhances efficient detection, particularly in scenarios with limited construction photos. Furthermore, this study aims to define major operational steps in structural engineering, such as rebar binding, formwork assembly, and concrete pouring, for effective control and management of construction progress. Additionally, suggestions for integrating progress management systems will be provided, improving existing research.

## 3 Methodology

To achieve these objectives, we designed a prototype system, which includes the required model architecture, on-site install flow, and functional displays. Building the progress object detection module in the construction phase, configuring site cameras, data integration and presentation will be detailed in the following sections.

### 3.1 Building Progress Object Detection Module in the Construction Phase

The study proposes the utilization of a progress object detection module for the construction phase. By collecting images of construction progress at various phases and applying transfer learning, the model parameters will be adjusted to enhance recognition. A comparison of various models will be conducted to identify the most suitable one for construction progress detection.

The implementation is divided into three stages: image classification and data collection, selection of the object detection model, and training/testing of the object detection model.

#### 3.1.1 Image Classification and Data Collection

To collect datasets for training the model, this study focuses on collecting relevant images from architectural projects, specifically targeting the structural construction phase. To mitigate redundancy in the training data that may result in high feature similarity and potentially impact the model's training outcomes, photos are collected through three distinct approaches: daily progress photos, time-lapse cameras on site, and online sources.

In this study, focusing on architectural structures, the collected progress images are mainly categorized into rebar tying, formwork assembly, and concrete pouring. Recognizing variations in the construction sequence between columns and walls, which proved to be confusing in initial tests, the construction phases are dissected into distinct phases: rebar tying of columns, pre-rebar tying of walls, completion of rebar tying of walls, formwork assembly of walls, formwork assembly of columns and concrete pouring.

### 3.1.2 Selection of Object Detection Model

The object detection model is composed of three parts: input, convolutional neural network layers, and detection layers. The convolutional neural network layers serve as the main network backbone, responsible for extracting image features and producing a feature map by merging the extracted features through pooling layers. The detection layers are responsible for the final prediction of object categories and generating candidate boxes.

To explore and seek a suitable combination of detection models, this study chose one-stage and two-stage object detection models, and selected models that have performed well in most related studies, YOLOv5 represents the one-stage model, while Faster R-CNN represents the two-stage model. Then, five groups of convolutional neural network layers and detection layers are constructed.

Table 1 Comparison of object detection model combinations

| | Convolutional Neural Network Layers | Detection Layers | mAP(%) | FPS |
|---|---|---|---|---|
| 1 | ResNet50 | YOLOv5 | 27.6 | 11.7 |
| 2 | ResNet50 | Faster R-CNN | 31.8 | 4.3 |
| 3 | DenseNet121 | YOLOv5 | 31.1 | 10.3 |
| 4 | DenseNet121 | Faster R-CNN | 38.3 | 3.9 |
| 5 | CSPDarknet53 | YOLOv5 | 51.1 | 17.6 |

The training conditions are set to 300 epochs, with 16 samples per batch, and image pixels of 640x640. The pixel size of the images utilizes the maximum value allowed by the computer hardware to avoid suboptimal training results. Regarding optimization parameters, the original default values of the model will be used, and to examine the fundamental performance of the model, no transfer learning pre-trained weights will be adopted. The model will be trained from scratch to obtain the detection model's training results.

Among the selected detection model combinations, the model with the highest individual accuracy will be compared. Considering conditions such as detection box overlap and detection speed, YOLOv5 - CSPDarknet53 will be adopted as the detection model for this study.

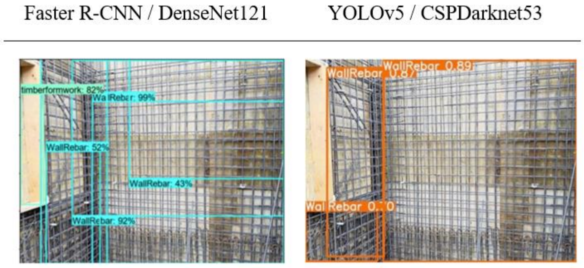


Figure 3. Compare prediction results and candidate boxes of detection models

### 3.1.3 Training and Testing of the Object Detection Model

The quantity of the dataset can impact the accuracy. During the initial stages, obtaining a clear and adequate number of construction progress photos posed challenges. This is mainly due to construction activities developing in different phases, and the collection of construction photos progresses gradually with the advancement of the project, making it difficult to rapidly and substantially increase the data volume.

To address this, the study utilizes the mosaic feature proposed in YOLOv4 [17] as a form of data augmentation to generate additional photo data, aiming to increase the number of the dataset. Additionally, the study adjusts hyperparameters during the training process to enhance the detection accuracy of the model. In this phase, transfer learning is applied, utilizing the dataset named MS COCO (Microsoft Common Objects in Context) [18-19] to train the YOLOv5 model. This source was designed to detect and segment common objects like humans, cars, and buses in daily life. The dataset consists of 328K images and 80 object categories.

Through transfer learning and hyperparameters optimization, the optimized model improves 20.4% accuracy over the original model, with no significant degradation in FPS (frames per second, FPS) performance. On average, the optimized model shows better loss values and performance compared to the initial architecture.

Table 2 Model optimization performance

| YOLOv5 | Accuracy (%) | Recall rate (%) | mAP (%) | FPS |
|---|---|---|---|---|
| Initial model | 63.1 | 74.67 | 51.1 | 17.6 |
| Optimization model | 83.5 | 71.74 | 57.9 | 17.4 |

## 3.2 Site Cameras Deployment and Data Integration and Presentation

Upon the completion of the detection model, many surveillance cameras are deployed on the construction site to record real-time video during the ongoing construction. This process is aimed at further recognizing the construction progress of each component. The positioning and alignment of cameras are constrained by some factors such as the site location and installation conditions. Therefore, systematic positioning and alignment are adopted to facilitate recognition and utilization of image data. In addition, image capture is performed from multiple angles, many components will be recorded repeatedly. This study provides a procedure to integrate images and presents a visual approach that enables management personnel to intuitively and promptly comprehend the construction progress.

#### 3.2.1 Site Cameras Deployment

The input source for this study is captured from the surveillance camera at the site. To ensure alignment between the input and the system's camera, two sets of camera initialization, positioning, and alignment configurations must be executed. The methods vary depending on whether the on-site surveillance cameras are already installed or are to be set up based on parameters. In this study, the coordinate system in Unity is utilized as a reference for calculating the relative spatial position after importing the model. Different processes and methods for installation are planned for two scenarios, as illustrated in the figures below.

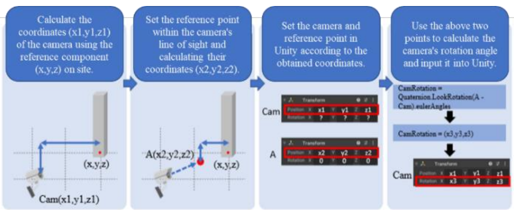


Figure 4. Case of on-site surveillance cameras are already installed

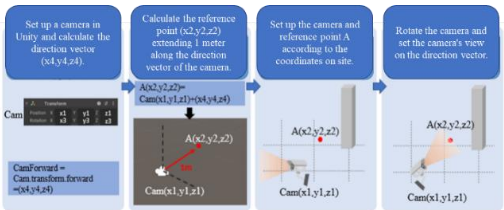


Figure 5. Case of on-site surveillance cameras are set up based on system parameters

#### 3.2.2 Mechanism of Construction Image Detection and Input

The return of images captured by the surveillance cameras at the site to the system's main server, the image detection model is utilized for recognition. The recognition results are outputted as a txt format, including information such as the coordinates of the detection box's center point, length, height, recognized progress phase, and mAP value.

The consistency perspective has been established between Unity and site cameras, the photos captured by the cameras can be considered aligned with the perspective of the model. Therefore, use the Unity Physic. Raycast function for component selection to find the component corresponding to the recognition result.

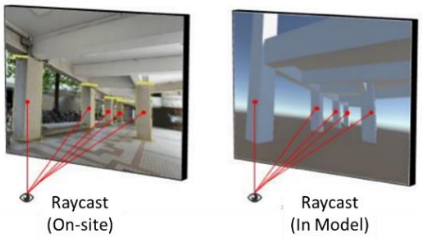


Figure 6. Demo of Physic. Raycast function for component selection

Because the system integrates multiple cameras, it may recognize different phases for the same component. This study proposes three solutions, allowing users to choose the most suitable method for their engineering project. These are prioritized based on accuracy (mAP), construction phase, and detection frequency. This approach enables the provision of distinct update principles based on the condition of the engineering project.

#### 3.2.3 Visual Presentation and the User Interface

Through the automated progress update function, it can reduce the human resources required for project management. On the other hand, visual presentation allows operators to understand the construction progress through screen presentation, thereby effectively improving work efficiency. This study proposes two visual presentation methods, explained as follows:

Users can select components through the Physic. Raycast function in the operation interface. This will query and display different colors corresponding to the construction phases detected by image recognition, simulating the on-site construction situation.

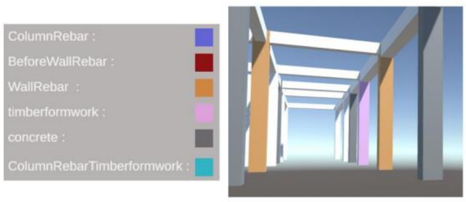


Figure 7. Query and display different colors corresponding to the construction phase

Another visual presentation method focuses on the overall progress control. Users can update component progress through the progress update interface. When this function is enabled, the system will compare and analyze the planned and actual completion dates for each component, presenting the results through the 3D model display area.

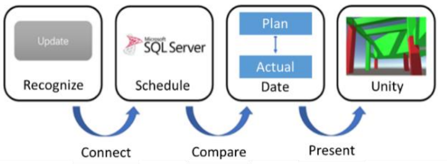


Figure 8 Process of schedule update and model presentation

## 4 Application Scenarios

Based on the prototype system developed, this section will conduct tests through various scenarios to showcase the functionalities developed to achieve the research objectives. The following sections will categorize the system's operations and provide detailed demonstrations and explanations for each application mode.

### 4.1 The Camera Deployment of the Prototype System

In the initial phase, this study utilized a simulated construction site in an interior parking floor of the building to validate whether the camera deploy functionality of the prototype system aligns with the requirements of typical construction scenarios. The red-highlighted area in the layout represents the scope of the research tests. Four surveillance cameras with corresponding field-of-view lenses were strategically installed both on-site and within the system, facilitating subsequent progress detection through the detection function and verifying its capability to cover the entire testing area.

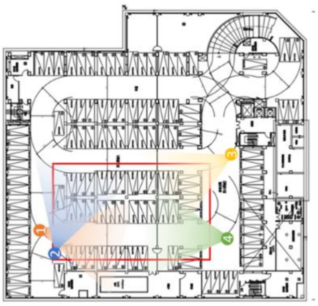

Figure 9. The layout of the research tests scope and camera deploys

The test has already pre-installed multiple perspective cameras in the system, and the desired camera position data will be obtained within the prototype system. In the system interface's 3D model display area, the selected perspective camera position will be shown. Users can choose a reference component closest to this camera by clicking with the mouse and calculating the X and Y-axis displacement. Finally, at the site, using this reference component as the origin, surveillance cameras will be set up using relative displacement.


Figure 10. Demonstration of Camera set up on site and in model

### 4.2 Recognize Various Construction Phases

This study utilized an existing building to simulate the site environment and whether the detection model can recognize various construction phases. The study modified the final images to depict scenarios such as rebar tying of columns, formwork assembly of walls, and concrete pouring.

Through the execution of image recognition functions, the prototype system of this study successfully recognized the construction phases of components, except for components that have been obstructed.


Figure 11. Simulate the completion phase of rebar tying and formwork assembly


Figure 12. Simulate the completion phase of concrete pouring

### 4.3 Multi-angle View Detection and Results Integration

After the AI conducts object detection and outputs the results, the prototype system automatically reads the information. Subsequently, from four different perspectives, the BIM model is interactively selected using the Unity Physic. Raycast function. Based on these selected components, progress data is updated. The BIM model, representing the defined testing area, adjusts its color presentation according to the different construction phases. The extraction of the updated results from the selected components confirms the effective coverage of the entire construction area using multiple cameras.

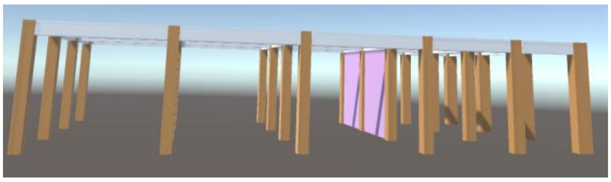

Figure 13. The completion phase visualization of rebar tying and formwork assembly

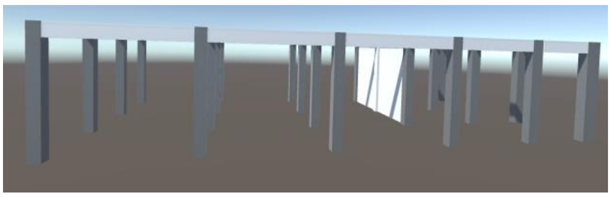

Figure 14. The completion phase visualization of concrete pouring

### 4.4 Use Surveillance Cameras for Recognize Testing during the Construction Stage

After validating the relevant functions indoors, the system proceeded to on-site testing in a construction setting. In this case, focusing on a construction project, CCTV images from the construction site were used as the image source, and the BIM model for the project was constructed for system testing. Initially, the prototype system deployed surveillance cameras. After measuring and calculating the data for the coordinates of the system's perspective camera at the site, the system completed the installation of the perspective camera. The comparison between the system's perspective view and the construction site image after deployment is shown in the following figure.

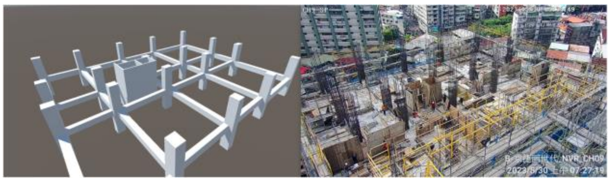

Figure 15. Comparison of system perspective and construction site images

Subsequently, the construction image recognition function was executed. The prototype system, after detection, retrieved the results and utilized the functionality to select the components to be updated. The comparison between the recognition results and the system interface is illustrated in the following figure. This presentation of results demonstrates the feasibility of various functions of the prototype system in practical cases.

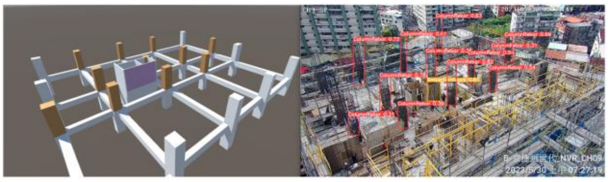

Figure 16. Comparison of recognized results and visualization component screen

## 5 Conclusions

This study proposes a prototype system that integrates AI image recognition, BIM, visualization technology, and on-site construction image monitoring. It not only provides real-time monitoring of construction site progress but also establishes an automated and visual management system. The AI detection and recognition integration mode, based on BIM, realizes an automated and visual construction progress management platform.

The system automates the integration of detection results from multiple surveillance cameras, ensuring comprehensive progress control over the entire construction area. It also utilizes a BIM model to integrate construction schedules, achieving automated updates to the schedule, thus reducing the operational loading on project managers.

In terms of functionality, the system integrates a BIM model and 4D construction management, using visualization technology to present different construction phases according to the schedule. Additionally, the system uses different colors on the component to show detection results, simultaneously comparing planned schedules with actual schedules. This color-coded representation indicates whether the construction progress of components is ahead or behind, providing project managers with a more concrete and intuitive understanding of construction progress.

With the continuous development of mixed reality technology, this study plans to integrate wearable mixed reality devices in the future. This involves presenting the model on MR devices to assist less-experienced on-site engineers in quickly familiarizing themselves with the site conditions. Additionally, by replacing camera installations with wearable mixed-reality devices, the aim is to achieve real-time image detection and updates.

## 6 References


[1] Son H. and Kim C., 3D structural component recognition and modeling method using color and 3D data for construction progress monitoring. *Automation in Construction*, 19(7):844-854, 2010.

[2] Golparvar-Fard M., Peña-Mora F., Savarese S., Application of D4AR - A 4-Dimensional augmented reality model for automating construction progress monitoring data collection, processing and communication. *Journal of Information Technology in Construction*, 14:129-153, 2009.

[3] Volk R., Stengel J., Schultmann F., Building Information Modeling (BIM) for existing buildings — Literature review and future needs. *Automation in Construction*, 38:109-127, 2014.

[4] Elbeltagi E. and Dawood M., Integrated visualized time control system for repetitive construction projects. *Automation in Construction*, 20(7):940-953, 2011.

[5] LeCun Y., Bengio Y., Hinton G., Deep learning. *Nature*, 521:436–444, 2015.

[6] Schmidhuber J., Deep learning in neural networks: An overview. *Neural Networks*, 61:85-117, 2015.

[7] Alizadehsalehi S., Yitmen I., The Impact of Field Data Capturing Technologies on Automated Construction Project Progress Monitoring. *Procedia Engineering* 161:97-103, 2016.

[8] Golparvar-Fard M., Peña-Mora F., Savarese S., Application of D4AR - A 4-Dimensional augmented reality model for automating construction progress monitoring data collection, processing and communication. *Journal of Information Technology in Construction* 14:129-153, 2009.

[9] Costin A., Pradhananga N., Teizer J., Leveraging passive RFID technology for construction resource field mobility and status monitoring in a high-rise renovation project. *Automation in Construction* 24:1-15, 2012.

[10] Shahi A., West J. S., Haas C. T., Onsite 3D marking for construction activity tracking. *Automation in Construction* 30:136-143, 2013.

[11] Pučko Z., Šuman N., Rebolj D., Automated continuous construction progress monitoring using multiple workplace real time 3D scans. *Advanced Engineering Informatics* 38:27-40, 2018.

[12] Baduge S. K., Thilakarathna S., Perera J. S., Arashpour M., Sharafi P., Teodosio B., Shringi A., Mendis P., Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications. *Automation in Construction* 141: 104440, 2022.

[13] Reja V. K., Varghese K., Ha Q. P., Computer vision-based construction progress monitoring. *Automation in Construction* 138:104245, 2022.

[14] Zheng Z., Zhang Z., Pan W., Virtual prototyping- and transfer learning-enabled module detection for modular integrated construction. *Automation in Construction* 120:103387, 2020.

[15] Cheng J. C. P., Wong P. K. Y., Luo H., Wang M., Leung P. H., Vision-based monitoring of site safety compliance based on worker re-identification and personal protective equipment classification. *Automation in Construction* 139:104312, 2022.

[16] Martinez P., Barkokebas B., Hamzeh F., Al-Hussein M., Ahmad R., A vision-based approach for automatic progress tracking of floor paneling in offsite construction facilities. *Automation in Construction* 125:103620, 2021.

[17] Bochkovskiy A., Wang C. Y., Liao H. Y. M., Yolov4: Optimal speed and accuracy of object detection. *arXiv* 2004.10934, 2020.

[18] Lin T. Y., Maire M., Belongie S., Hays J., Perona P., Ramanan D., Dollar P., Zitnick L., Microsoft COCO: Common Objects in Context. *Proceedings of the 2014 European Conference on Computer Vision*, 740-755, 2014.

[19] cocodataset.org. COCO dataset. On-line:[https://cocodataset.org/](https://cocodataset.org/), Accessed:16/02/2024