AS-BUILT DETECTION OF STRUCTURES BY THE SEGMENTATION OF THREE-DIMENSIONAL MODELS AND POINT CLOUD DATA

Nobuyoshi Yabuki & Tomohiro Fukuda
Osaka University, Japan

Ryu Izutsu
Kajima Corporation, Japan

ABSTRACT: *At construction sites, as-built management is generally conducted by taking pictures or surveying with total stations and comparing the images or survey data with design drawings or Building Information Modeling (BIM) models. Since this work is time-consuming and error-prone, more efficient and accurate methods using advanced Information and Communication Technology (ICT) are desired. Therefore, this research proposes a method that can efficiently capture the progress of construction by detecting each constructed structural member, such as beams, columns, connections, etc. In this proposed method, construction engineers first take many pictures of the construction site and conduct automatic image segmentation using a pre-trained Convolutional Neural Network (CNN) model. Next, point cloud data is generated from taken pictures by using Structure from Motion (SfM). Then, the point cloud data is semantically segmented by overlapping the segmented images and point cloud data using the pin-hole camera technique. Finally, the design BIM model and segmented point cloud data are overlapped, and constructed parts of the BIM model can be detected, which can be reported as as-built parts. A prototype system was developed and applied to an actual railway construction project in Osaka, Japan for testing the accuracy and performance of the system.*

KEYWORDS: *Construction progress management, Instance segmentation, Point cloud, Building Information Modeling.*

1. INTRODUCTION

Construction site management involves inspecting the completed parts of a construction project to ensure that the work is within specifications and contractual requirements. This task requires construction workers to compare the actual construction with the provided drawings and documents. The goal is to ensure that the construction is performed correctly and to calculate the corresponding contract price. Traditionally, construction management relied on drawings, but the use of 3D models has become more prevalent. These models enable better visualization and consensus building among stakeholders. While image data and laser scanners have been used in previous studies to create 3D models, large-scale structures and deep learning techniques have not been fully utilized for construction site monitoring. The succession of technical skills in the construction industry has been identified as an issue, prompting the need for changes in the construction production system. Leveraging technology advancements, such as 3D models and sensor information, has improved efficiency and contributed to various aspects of construction, including design, management, and maintenance. Building Information Modeling (BIM) is a lifecycle management system that facilitates efficient building maintenance. However, the process of collating 3D models with 2D drawings is time-consuming and prone to human error. Structure from motion (SfM) is a method used to acquire 3D data of existing structures, but converting point cloud models to polygon models presents challenges such as removing unnecessary details and setting appropriate thresholds. Efforts are needed to develop more efficient methods for capturing the current 3D model of a structure.

Recent advancements in deep learning and object detection technology have automated tasks such as construction site inspections, including identifying deformations and damages from images. The availability of large image datasets, such as ImageNet, has greatly improved object recognition accuracy using deep learning algorithms. In addition, recently, much research has been done for classifying point cloud data using deep learning (Charles et al. 2017). However, much research is required to classify civil infrastructure members.

Thus, this research has adopted a more simple 2D object detection method using deep learning and a pin-hole camera method and combined it with 3D BIM models to reproduce the construction situation on a 3D model and calculate construction costs. A training dataset specific to construction members was created to fine-tune existing deep-learning models. The proposed method enables efficient shape detection and attribute identification of construction elements and should contribute to the integration of detection information into 3D models, facilitating the creation of as-built models.

Referee List (DOI: 10.36253/fup_referee_list)
FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)
Nobuyoshi Yabuki, Tomohiro Fukuda, Ryu Izutsu, *As-Built Detection of Structures by the Segmentation of Three-Dimensional Models and Point Cloud Data,* pp. 1117-1124, © 2023 Author(s), CC BY NC 4.0, DOI 10.36253/979-12-215-0289-3.111

2. RELATED WORKS

Before the advent of deep learning-based object detection, selecting features for object detection was challenging, especially in complex construction sites with various members and intricate structures. Past research in the field of construction has focused on automating tasks such as progress and productivity management using deep learning. One study proposed a system to automatically recognize completed parts in construction site images, but the detection results were not applicable to other systems, and accuracy was limited for complex-shaped structural members (Fathi et al., 2015). Research combining automation technology and BIM has aimed for efficient work (Kropp et al., 2018; Park et al., 2018), but perfect automation remains elusive due to the need for human intervention.

Another study developed a management system using a 3D model and proposed a method for constructing original models by detecting structural members in existing bridges from point cloud data (Lu et al., 2019). However, the method faced challenges in detecting complex geometric structures such as concrete or truss bridges. A laser scanner is used to create detailed BIM models of existing facilities but encountered difficulties with complex structures and occlusion (Tang et al., 2010). Various approaches have been attempted to create 3D models of existing buildings (Bosche et al., 2009; Brilakis et al., 2010).

Recent advancements in computer vision technology have enabled the automation of tasks performed by the human visual system. One study developed a system that automatically detects construction members in a room using 2D image data (Hamledari et al., 2017). Other studies have attempted to capture construction status and shape from image data (Gidaris et al., 2015; Khaloo et al., 2015). Perez-Perez et al. (2021) developed a method for the segmentation of indoor point clouds via joint semantic and geometric features for 3D modeling of the built environment. Pan et al. (2022) proposed geometric digital twins of buildings with small objects by fusing laser scanning and AI-based image recognition. However, the detection of different material members and multiple structure types for outdoor civil infrastructures remains challenging. Therefore, this research aims to fill this gap to improve the performance of as-built detection of civil infrastructures under construction for better construction site management.

3. PROPOSED METHOD

The proposed method aims to recreate the construction status in an as-built 3D model by incorporating the shape information from the detection result images obtained through a deep learning model. This allows for cost calculations without the need to match 2D drawings with the construction progress. The positional relationship between the detection result images, the completed 3D model, and a point cloud model generated using Structure from Motion (SfM) are matched.

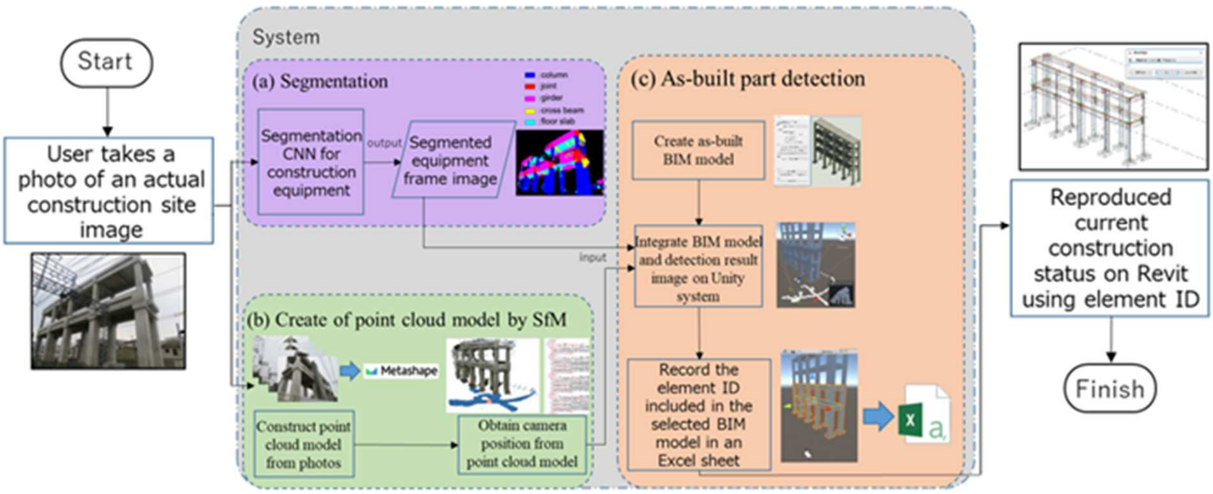

Fig. 1: Method overview.

As shown in Fig. 1, the method consists of three main steps: (1) performing segmentation detection using a fine-tuned deep learning model to identify structural members in the construction images, (2) creating a point cloud model using Agisoft Metashape to determine the 3D positions of the images, and (3) integrating the completed 3D model, detection result images, and point cloud model in a volume detection system using the Unity game engine. The positional relationship is established, and the identified structural members are recorded in an Excel sheet. Finally, the construction status is reproduced in BIM software (Revit), and the attribute information of the structural members is used to calculate the construction cost at the time of shooting.

3.1 SEGMENTATION

In this study, the weights of a U-Net model trained on the Cityscapes Dataset (Cityscapes Dataset, n.d.) were adjusted to distinguish structural members and the background. By updating the weights of the 37 layers, the positions and attributes of the structural members in the captured images could be identified. These detection results were treated as the finished form, providing insights into the construction site's situation. Since there were no published trained models for construction members such as columns, beams, and ducts, a training dataset was created using interior photographs of buildings under construction. The dataset was manually annotated using Adobe Photoshop CS4, creating mask images for each target member. The existing trained model was then fine-tuned using the mask images and the corresponding color changes in the original images.

In this study, a U-Net model trained on the Cityscapes Dataset was fine-tuned and used as a CNN for object detection to detect construction structural members from images, specifically targeting the five main structural members (Fig. 2).

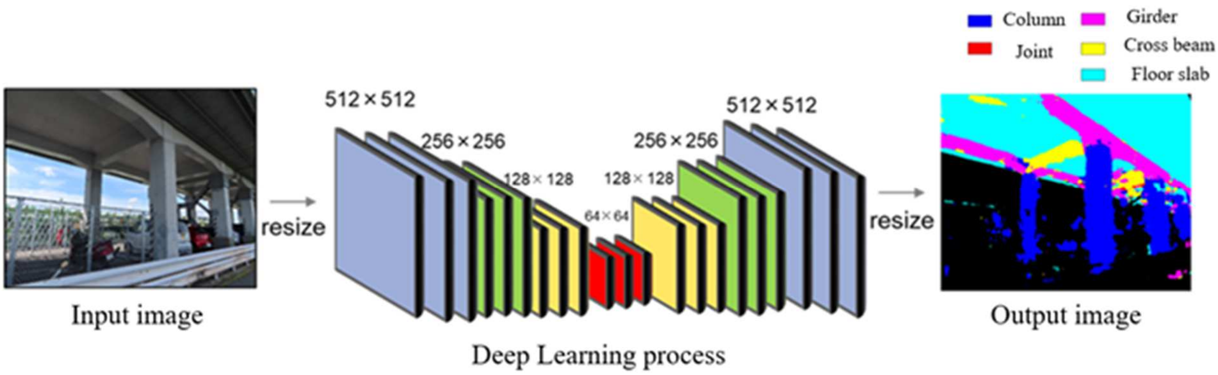

Fig. 2: U-Net-based CNN structure and learning results (example).

Meanwhile, a three-dimensional model representing the real space, including the camera position and target structure, was created using Structure from Motion (SfM) and Agisoft Metashape, a software for photogrammetric processing and 3D spatial data generation. Fig. 3 shows the results of fine-tuning U-Net using the created training dataset, where IoU stands for Intersection over Union.

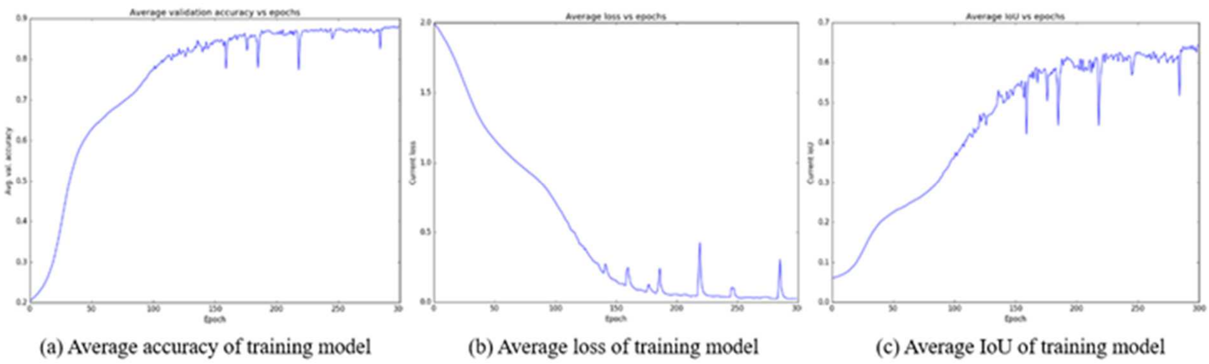

Fig. 3: The results of fine-tuning U-Net using the created training dataset.

The deep learning model in this study does not exhibit over-learning. The average accuracy for the training dataset increases, but the average loss for the test training dataset does not decrease. The detection accuracy of the fine-tuned model is evaluated using the average IoU value, which is 0.6428 at the 300th epoch. The IoU value measures the overlap between the correct answer area and the predicted area, indicating the model's performance. The evaluation index IoU indicates the numerical value obtained by dividing the overlapping part of the correct answer area and the prediction area by the union part of both areas, as explained in equation (1).

$$IoU\text{ (Intersection over Union) } = \text{(Intersection of detection areas)}/\text{(Union of detection areas)} \tag{1}$$

3.2 POINT CLOUD MODEL CONSTRUCTION

It is difficult to reproduce the positional relationship of the construction status when each image and model is imported into the game engine Unity in the lack of coordinate information. Therefore, we use SfM and Agisoft Metashape software to create a three-dimensional model that replicates the camera position and target structure in virtual space. Metashape allows us to process digital images and generate 3D spatial data, enabling the scanning of both small objects and large buildings. By analyzing the overlapping shooting locations in the photographs, we can calculate the distance to the subject in each photo.

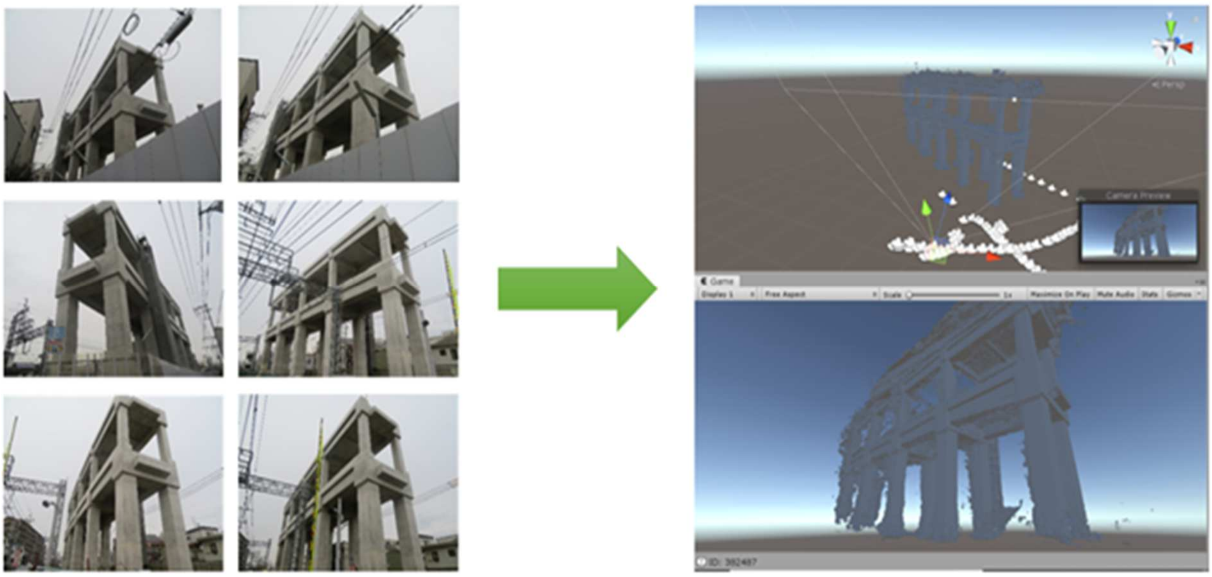

Fig. 4: Transfer from real frame to SfM model.

The gray model consists of a mesh overlaying a point cloud model created in Metashape, while the white object represents a virtual camera. As shown in Fig. 4, we can confirm the accurate reproduction of the camera's position and the target structure in Unity. Valid values for camera parameters such as position and rotation were confirmed in Unity, indicating successful reproduction of the real-world positions of the target structure and the camera in the game space. By overlaying the SfM model with the expected BIM model, a work detection system was created. The deep learning model performs object detection using the created mask image, adjusting the position and rotation coordinates of the virtual camera based on the mask image's parameters. The field of view on the virtual camera side is also adjusted to match the mask image. These preparations enable the replication of the construction situation in the game engine and the reflection of the deep learning model's detection results onto the BIM model.

4. EVALUATION

A case study was conducted on a building under construction at Osaka University to verify the proposed method. The system was tested by creating a BIM model from construction drawings and capturing photographs at the construction site, allowing for the verification of volume detection using a point cloud model. The system was implemented and tested using Unity on a standard PC with an Intel Core i7-3770K CPU and 32 GB RAM.

The detection result image from the deep learning model is utilized as a filter to extract and choose the completed portion from the generated BIM model. By aligning the aspect ratio of the Unity camera with the actual image size, the system excludes the undetected background area that is still under construction, ensuring only completed members are selected such as shown in Fig. 5.

In order to apply the image filter to the application, multiple angled frames are captured and went through a deep learning model with different weights. Fig. 6 shows part of the detection results

Fig. 7 shows the generated model in Unity and Revit. The detection results on the left if it indicates the accurate detection of completed ducts. However, upon examining the member model based on element ID, it was found that four beams on the front side of the third-floor slab were missing. Additionally, low detection accuracy is observed for elements not included in the learning dataset, such as overhead poles, scaffolds, and multiple electric wires, depending on the viewing angle of the target structure. On its right, the screen displays the selection of the element ID obtained from the viewpoint, with the selected member highlighted by the blue wireframe line and the unselected part shown by the black wireframe line.

Table 1 displays the results of member detection from multiple viewpoints in the case study, including detection accuracy for each member and overall detection accuracy calculations.

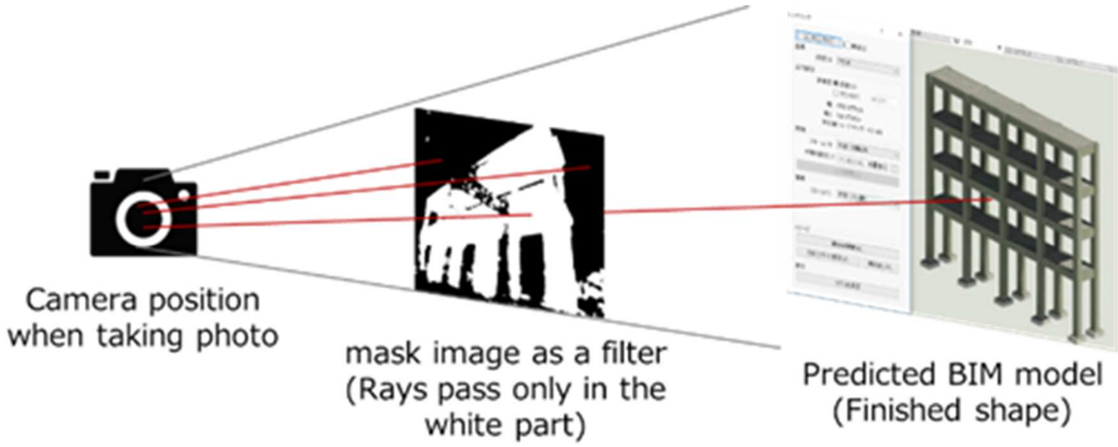

Fig. 5: Image filter in BIM model construction.

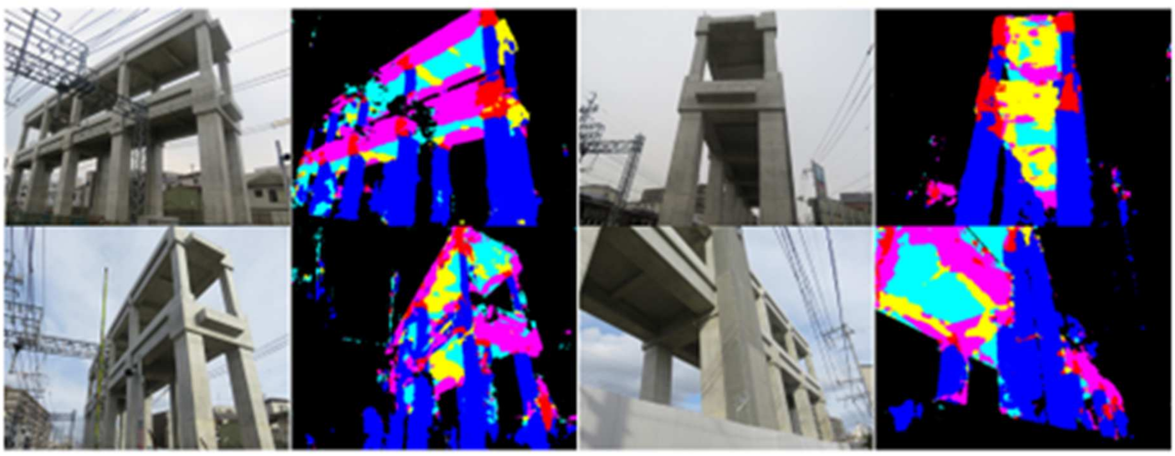

Fig. 6: Part of the results in image processing.

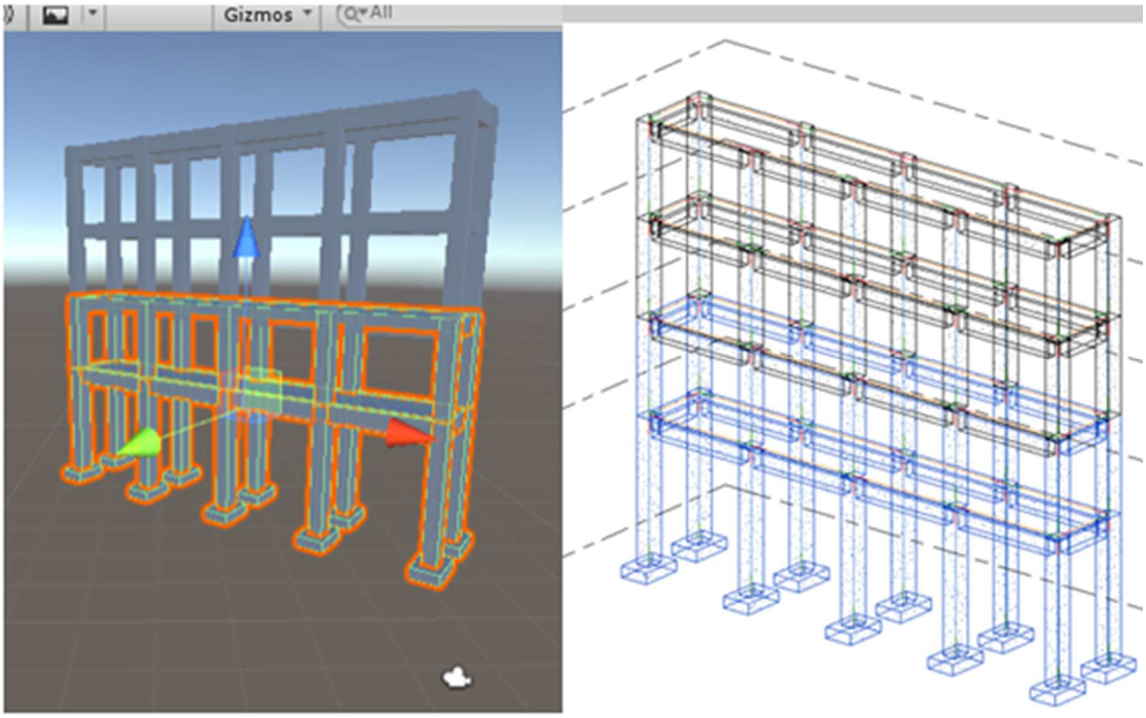

Fig. 7: Generated 3D model in Unity and Revit.

Table 1: Structure member detection accuracy.

Shooting viewpoint	Overall detection accuracy (%)	Column detection accuracy (%)	Floor slab detection accuracy (%)	Foundation detection accuracy (%)	Beam detection accuracy (%)
Viewpoint 1	92.31	100	100	100	80
Viewpoint 2	98.08	95	100	100	100
Viewpoint 3	83.33	90	100	100	75
Viewpoint 4	86.79	90	100	100	80

5. CONCLUSION

This study aimed to verify the effectiveness of using deep learning for detecting structural members and construction equipment at a construction site. To overcome the limitation of existing training datasets, a verification experiment was conducted using images created from photographs of other construction sites. The existing convolutional neural network (CNN) was fine-tuned with different learning weights to detect structural members from actual construction site photos. The following results were obtained:

- The shape of the target structure could be detected from the construction site photographs by considering the detection result image.
- A volume calculation system was constructed using the deep learning model's segmentation results, enabling volume calculation on a three-dimensional model based on the shape information from two-dimensional images.
- The 3D model that reproduces the construction site was displayed on BIM software like Revit by acquiring the element ID from the 3D model using the Unity game engine.

- The recall of the constructed 3D model at each viewpoint showed an average of 90%, demonstrating high accuracy by combining detection results from multiple viewpoints.
- By assigning attribute information of construction unit prices to each member, it was possible to calculate the work volume based on the work form.

To improve the accuracy of the volume detection system, the deep learning model's detection accuracy needs enhancement. The training dataset should include information on detecting obstacles in front of the target. The applicability of the proposed method to other structures and the diversification of the system needs further investigation.

REFERENCES

Bosche, F., Haas, C. T., & Akinci, B. (2009). Automated recognition of 3D CAD objects in site laser scans for project 3D status visualization and performance control. Journal of Computing in Civil Engineering, 23(6), 311-318. https://doi.org/10.1061/(ASCE)0887-3801(2009)23:6(311)

Brilakis, I., Lourakis, M., Sacks, R., Savarese, S., Christodoulou, S., Teizer, J., & Makhmalbaf, A. (2010). Toward the automated generation of parametric BIMs based on hybrid video and laser scanning data. Advanced Engineering Informatics, 24(4), 456-465. https://doi.org/10.1016/j.aei.2010.06.006

Charles, R. Q., Su, H., Mo, K., & Guibas, L. J. (2017). PointNet: Deep learing on Point Sets for 3D Classification and Segmentation. IEEE Conference on Computer Vision and Pattern Recognition, DOI: 10.1109/CVPR.2017.16

Fathi, H., Dai, F., & Lourakis, M. (2015). Automated as-built 3D reconstruction of civil infrastructure using computer vision: Achievements, opportunities, and challenges. Advanced Engineering Informatics, 29(2), 149-161. https://doi.org/10.1016/j.aei.2015.01.012

Gidaris, S., & Komodakis, N. (2015). Object detection via a multi-region and semantic segmentation-aware CNN model. In Proceedings of the IEEE International Conference on Computer Vision, 1134-1142. https://doi.org/10.48550/arXiv.1505.01749

Hamledari, H., McCabe, B., & Davari, S. (2017). Automated computer vision-based detection of components of under-construction indoor partitions. Automation in Construction, 74, 78-94. https://doi.org/10.1016/j.autcon.2016.11.009

Khaloo, A., & Lattanzi, D. (2015). Extracting structural models through computer vision. In Structures Congress 2015, 538-548.

Kropp, C., Koch, C., & König, M. (2018). Interior construction state recognition with 4D BIM registered image sequences. Automation in Construction, 86, 11-32. https://doi.org/10.1016/j.autcon.2017.10.027

Lu, R., Brilakis, I., & Middleton, C. R. (2019). Detection of structural components in point clouds of existing RC bridges. Computer-Aided Civil and Infrastructure Engineering, 34(3), 191-212. https://doi.org/10.1111/mice.12407

Pan, Y., Braun, A., Brilakis, I., & Borrmann, A. (2022). Enriching geometric digital twins of buildings with small objects by fusing laser scanning and AI-based image recognition. Automation in Construction, 140, 104375. https://doi.org/10.1016/j.autcon.2022.104375

Park, J., Cai, H., & Perissin, D. (2018). Bringing information to the field: automated photo registration and 4D BIM. Journal of Computing in Civil Engineering, 32(2), 04017084. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000740.

Perez-Perez, Y., Golparvar-Fard, M., & El-Rayes, K. (2021). Segmentation of point clouds via joint semantic and geometric features for 3D modeling of the built environment. Automation in Construction, 125, 103584. https://doi.org/10.1016/j.autcon.2021.103584

Tang, P., Huber, D., Akinci, B., Lipman, R., & Lytle, A. (2010). Automatic reconstruction of as-built building information models from laser-scanned point clouds: A review of related techniques. Automation in Construction, 19(7), 829-843. https://doi.org/10.1016/j.autcon.2010.06.007

Cityscapes Dataset: Semantic Understanding of Urban Street Scenes. (n.d.). Retrieved May 19, 2023, from https://www.cityscapes-dataset.com/