# Semantic segmentation of 3D point cloud data acquired from robot dog for scaffold monitoring

**Juhyeon Kim[a], Duho Chung[a], Yohan Kim[a], and Hyoungkwan Kim[a]**

[a]Department of Civil and Environmental Engineering, Yonsei University, Korea
E-mail: kah5125@yonsei.ac.kr, jungduho1@yonsei.ac.kr, homez815@yonsei.ac.kr, hyoungkwan@yonsei.ac.kr

**Abstract –**

**Many of the fatalities and injuries in the construction industry occur in scaffolding accidents, and monitoring the scaffolding process and checking compliance are critical. However, monitoring scaffolds is labor-intensive and inefficient because it is done manually. To address this issue, we propose an advanced 3D reconstruction method for detecting and monitoring scaffolds. Deep learning-based RandLA-Net architecture is used to perform scene segmentation. RandLA-Net is trained based on transfer learning, using the knowledge of the model learned with the Semantic3D dataset. RandLA-Net uses 3D point cloud data that are matched and registered by LIO-SAM, a laser slam algorithm. By attaching a LiDAR to a quadruped robot, it is possible to obtain data frequently in a manner suitable for construction sites. The proposed methodology has demonstrated good performance in monitoring scaffolds.**

**Keywords –**

**Scaffold; Mobile Laser Scanning (MLS); Robot Dog; 3D Semantic Segmentation; Transfer Learning**

## 1 Introduction

According to the Korea Occupational Safety and Health Agency, 292 fatalities occurred at construction sites in 2019, in which 162 workers died during work related to temporary structures [1]. The main factor among the causes of fatalities and injuries is scaffolds, one of the representative temporary structures. To secure the safety, guidelines for installation and use are provided, and they check the spacing, angle, presence, etc. of the scaffold members. However, the process of checking the regulation of scaffolds is done manually by humans and takes a very long time.

Previous studies have attempted vision-based safety monitoring [2, 3] and sensor-based monitoring such as strain gauge or accelerometer [4, 5] to perform automated management of scaffolds. Wang [6] classified the point clouds of scaffolds acquired by Terrestrial Laser Scanning (TLS) using histogram of data, M-estimator Sample Consensus (MSAC), and Random Sample Consensus (RANSAC) algorithm. Xu et al. [7] proposed a reconstruction procedure for scaffolds using a 3D local feature descriptor, Linear Straight Signature Histogram of Orientations (LSSHOT), for the photogrammetric point cloud. For effective safety management of scaffolds, the actual geometry of the scaffold should be checked. However, not much research has been done on point cloud data containing geometric information of scaffolds. Moreover, the recognition method of scaffolds via a deep learning-based 3D segmentation model has not yet been proposed.

To address the above issue, this paper proposes a point cloud data reconstruction method of scaffolds acquired with a robot dog, as shown in Figure 1. In general, acquisition and post-processing of point cloud data take a lot of time. However, this study uses the robot dog and a Mobile Laser Scanning (MLS) method with a Simultaneous Localization and Mapping (SLAM) algorithm to improve the efficiency of data collection.

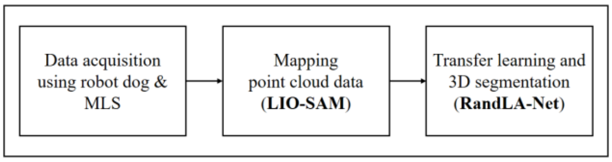



Figure 1. Overview of the proposed method

## 2 Methodology

### 2.1 Data acquisition system using robot dog

A data collection system using a robot dog (A1 robot dog of Unitree) was developed for this study. An Inertial Measurement Unit (IMU) sensor and Velodyne VLP-16 are mounted on the robot dog; the robot dog provides power for the sensors. Figure 2 shows the overall hardware configuration of the robot dog. The software development kit provided by Unitree, drivers for scanning instruments, and customized teleoperation code were installed on the on-board computer (NVIDIA Jetson

TX2). Main computer and on-board computer communicated via Secure Shell (SSH) using the robot's internal Wireless Access Point (WAP). By manipulating the on-board computer through SSH, control, scan, and data transmission became possible remotely.

Since the robot dog's field of view was limited when collecting data, this study developed appropriate scan motions and used them for data acquisition. The robot dog's maximum pitch and roll angle were 20 degrees, respectively, and the maximum yaw angle was 28 degrees. Therefore, the combination of angles of roll, pitch, and yaw has made the robot dog's field of view as wide as possible in the scan position. A few examples of posture at the scanning point of the robot dog are shown in Figure 3. Depending on the relative position of the scaffold, the robot's scanning posture was changed, and data were continuously acquired by wandering around the structure.

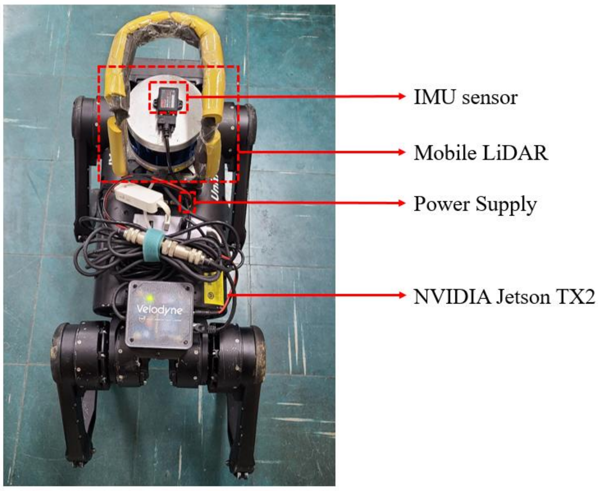
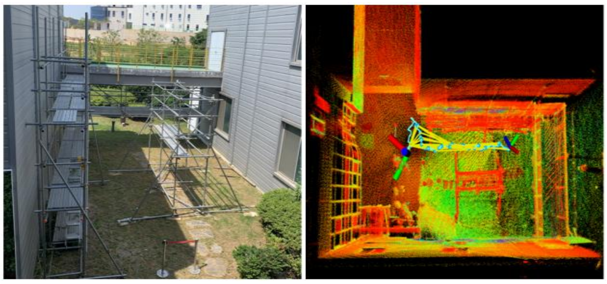



Figure 2. Hardware configuration of robot dog

Figure 3. Examples of robot dog scanning postures

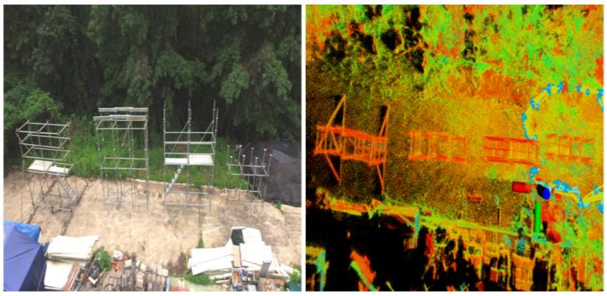
## 2.2 Mapping point cloud data

To create the point cloud map and estimate the odometry of the robot dog, a SLAM algorithm called LIO-SAM [8] was used. The LIO-SAM algorithm uses high-frequency IMU data to predict tightly-coupled lidar inertial odometry and enables precise and fast mapping in complex environments [8]. The SLAM algorithm was implemented with IMU data and LiDAR data. The data acquired from the robot's on-board computer were transferred to the laptop, the main computer. Figure 4(b) and Figure 5(b) show the registered point cloud data acquired at Site A and Site B, respectively.



(a) (b)

Figure 4. Site A dataset; (a) photograph of Site A, (b) registered point clouds



(a) (b)

Figure 5. Site B dataset; (a) photograph of Site B, (b) registered point clouds

## 2.3 Transfer learning and 3D segmentation

To achieve high learning performance from less training data, we propose a transfer learning-based semantic segmentation method for training a new model using a pre-trained model. The Semantic3D dataset [9] was used to obtain the pre-trained model. Semantic3D is one of the typical datasets used in 3D classification, 3D object detection and tracking, and 3D segmentation. This dataset was chosen because it is similar in size to our dataset, acquired outdoors, and is a registered data unlike SemanticKITTI [10].

The registered 3D point cloud data were segmented by RandLA-Net [11]. It achieved the best performance for 3D segmentation method on the Semantic3D benchmark [12]. RandLA-Net uses random sampling and local feature aggregators to classify large-scale point cloud data in a short period of time. The network follows the commonly used encoder-decoder structure and consists of five encoder-decoders. In this study, we attempted a fine-tuning method through changing the frozen layers of encoders and decoders of the pre-trained model. Consequently, the training performance was best when re-training three encoders and decoders in the middle of the network.

## 3 Experiments and Results

### 3.1 Dataset

Point clouds were acquired at Site A and Site B as shown in Figure 4 and Figure 5. Point cloud data acquired at Site A were used to train the model, and data acquired at Site B were used for testing and validation. Of the four structures at Site B, only the second structure was the validation data and the others were the test data. On Site A, point cloud data were acquired through the robot dog and a handheld method. Sensor data were stored in a Robot Operating System (ROS)-based bag file format, with 16 files acquired in the handheld method and 20 files acquired using the robot dog.

### 3.2 Implementation

RandLA-Net was trained with a batch size of 4 and there were two class labels: background and scaffold. Epoch was set to 100 by default, early stopping was used, and learning rate decay was implemented. The training process of the 3D segmentation model on data acquired by the robot dog is as follows: 1) Find the most suitable modified structure through fine-tuning using training and validation data acquired by the handheld method; 2) Train the model using the found structure in the above step with the training data acquired by the robot dog.

### 3.3 Evaluation metrics

The performance of the semantic segmentation model was computed by Eqs. (1), (2), and (3). Precision (P) indicates how many actual positive points are included among the positive points predicted by the model. Recall (R) calculates the proportion of points that are predicted to be positive properly among the points that are actually positive. F1-score (F1) is the weighted average of Precision and Recall.

$$\mathrm{P} = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (1)$$

$$R = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (2)$$

$$F1 = 2 * \frac{P * R}{P + R} \quad (3)$$

### 3.4 Results and Discussion

Table 1 demonstrates the performance of a model trained through transfer learning using data acquired by the robot dog. The model achieved an 84.96% F1-score on scaffolds. The F1-score of the scaffolds is always relatively lower than that of the background, because the number of points corresponding to the scaffolds is small. Figure 6 shows the worst and best examples of scaffold prediction. In Figure 6, pink is true positive, cyan is true negative, yellow is false negative, and orange is false positive. In the figure, the white area is caused by cropping the point cloud data corresponding to the validation data.

Table 1. The performance of the 3D segmentation model trained with the robot dog dataset

| | Precision | Recall | F1 score |
|---|---|---|---|
| Scaffolds | 97.15% | 76.45% | 84.96% |
| Background | 98.56% | 99.86% | 99.20% |

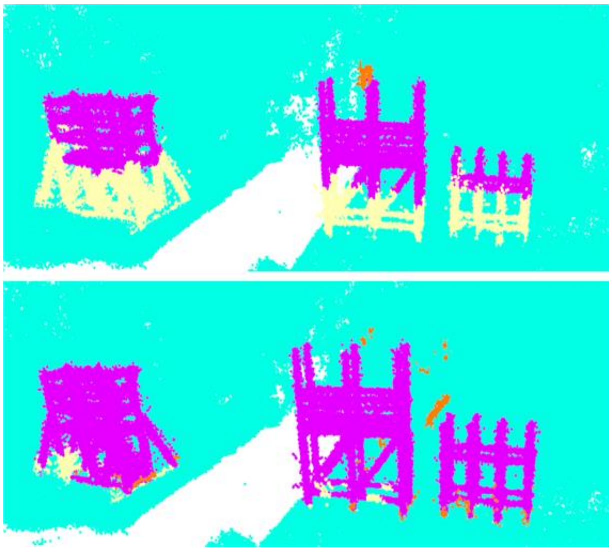


Figure 6. Examples of semantic segmentation results (top: F1 score is 63.16%; bottom: F1 score is 96.31%)

Overall, the lower parts of scaffolds were poorly predicted, as shown in Figure 7. It is assumed that this is due to the effect of noise caused by the surrounding environment or people, or to the difficulty of distinguishing between scaffolds and ground. In future studies, we will systematically compare the results of the data acquired by the handheld method with those of the data acquired using the robot dog. Furthermore, we will analyze the cause of poor prediction on the data acquired by the robot dog. Although the point cloud data of scaffolds were not perfectly classified with the proposed method, they were segmented more effectively than when using the RANSAC or histogram-based method attempted in previous studies.

Figure 7. Examples of failure of semantic segmentation

## 4 Conclusion

This study proposed a novel pipeline to segment 3D point cloud data acquired from a robot dog through transfer learning. The LIO-SAM-based data acquisition system was implemented and tested using a robot dog with various patterns of scanning postures. By the proposed data acquisition system, point cloud data can be effectively collected by the robot moving around the scaffold. With the transfer learning, the RandLA-Net algorithm was efficiently trained on the scaffold data of the two sites for scaffold segmentation. In the experiment, the F1-score from the handheld method (97.31%) was relatively higher than that from the robot dog (84.96%). The cause of this phenomenon will be analyzed in future research. The scaffold point clouds classified by the segmentation model can be used for 3D modeling through post-processing steps. Using the 3D model, safety managers can monitor the scaffolding installation and dismantling procedure and automatically check compliance with safety regulations. If the proposed method is further developed, it is expected to efficiently monitor scaffolds and reduce mortality and accident rates at construction sites.

## Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Ministry of Science and ICT (No. 2021R1A2C2004308) and the "National R&D Project for Smart Construction Technology (No.21SMIP-A158708-02)" funded by the Korea Agency for Infrastructure Technology Advancement under the Ministry of Land, Infrastructure and Transport, and managed by the Korea Expressway Corporation.

## References

[1] KOSHA (Korea Occupational Safety and Health Agency). Industrial Accident Analysis for 2019. On-line: https://www.kosha.or.kr/kosha/data/industrialAccidentStatus.do?mode=view&articleNo=419465&article.offset=0&articleLimit=10, Accessed: 16/07/2021.

[2] Chi, H. L., Chai, J., Wu, C., Zhu, J., Wang, X., and Liu, C. Scaffolding progress monitoring of LNG plant maintenance project using BIM and image processing technologies. In *International Conference on Research and Innovation in Information Systems (ICRIIS)*, pages 1-6, Langkawi, Malaysia, 2017.

[3] Khan, N., Saleem, M. R., Lee, D., Park, M. W., and Park, C. Utilizing safety rule correlation for mobile scaffolds monitoring leveraging deep convolution neural networks. *Computers in Industry*, 129: 103448, 2021.

[4] Yuan, X., Anumba, C. J., and Parfitt, M. K. Cyber-physical systems for temporary structure monitoring. *Automation in Construction*, 66: 1-14, 2016.

[5] Cho, C., Kim, K., Park, J., and Cho, Y. K. Data-driven monitoring system for preventing the collapse of scaffolding structures. *Journal of construction engineering and management*, 144(8): 04018077, 2018.

[6] Wang, Q. Automatic checks from 3D point cloud data for safety regulation compliance for scaffold work platforms. *Automation in Construction*, 104:

38-51, 2019.

[7] Xu, Y., Tuttas, S., Hoegner, L., and Stilla, U. Reconstruction of scaffolds from a photogrammetric point cloud of construction sites using a novel 3D local feature descriptor. *Automation in Construction*, 85: 76-95, 2018.

[8] Shan, T., Englot, B., Meyers, D., Wang, W., Ratti, C., and Rus, D. LIO-SAM: Tightly-coupled lidar inertial odometry via smoothing and mapping. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5135-5142, Las Vegas, USA, 2020.

[9] Hackel, T., Savinov, N., Ladicky, L., Wegner, J. D., Schindler, K., and Pollefeys, M. Semantic3D. net: A new large-scale point cloud classification benchmark. *arXiv preprint arXiv:1704.03847*, 2017.

[10] Behley, J., Garbade, M., Milioto, A., Quenzel, J., Behnke, S., Stachniss, C., and Gall, J. SemanticKITTI: A dataset for semantic scene understanding of lidar sequences. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9297-9307, Seoul, South Korea, 2019.

[11] Hu, Q., Yang, B., Xie, L., Rosa, S., Guo, Y., Wang, Z., ... and Markham, A. RandLA-Net: Efficient semantic segmentation of large-scale point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11108-11117, Seattle, USA, 2020.

[12] Guo, Y., Wang, H., Hu, Q., Liu, H., Liu, L., and Bennamoun, M. Deep learning for 3D point clouds: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 2020.