Virtual Prototyping-Enabled Pseudo-Lidar Point Cloud Dataset for 3D Module Detection in Modular Integrated Construction

Jiayi Xu[1], Wei Pan[1]
[1]Department of Civil Engineering, The University of Hong Kong, China
xiaoyi12@connect.hku.hk

Abstract. Onsite activities of modular integrated construction (MiC) mainly involve repetitive module installation processes. Automatic three-dimensional module detection (3DMD) can facilitate effective MiC site management with enhanced safety, efficiency, and quality by determining module postures through oriented 3D bounding boxes. However, the lack of and difficulty in creating datasets with real point clouds of module installation impede the exploration of 3DMD. This paper thus proposes a virtual prototyping (VP)-enabled method to develop a pseudo-lidar point cloud dataset for 3DMD. Building information modelling and 3D simulation engine were combined to establish simulative MiC construction scenes. Then, diverse scenarios were developed for generating and annotating pseudo-lidar point clouds. The dataset was validated using the Frustum-PointNet model with a maximum average precision of 89%. This paper thus provides a methodological foundation for the 3D detection of other construction objects and an innovative approach to strengthen image/point cloud datasets in construction.

1. Introduction

Modular integrated construction (MiC) is gaining increasing attention in the construction industry with demonstrated benefits of enhanced productivity, quality, safety, and sustainability (Pan and Hon, 2020). It employs multiple three-dimensional (3D) units that are fully finished in the factory and then transported to the site for assembly. Onsite activities of MiC are thus predigested to involve repetitive module installation processes. 3D perception of these processes will facilitate effective MiC site management by fully understanding what is happening on the construction site. For example, safety will be enhanced by knowing the actual distances between modules and workers/equipment to prevent collision accidents, progress will be monitored by analysing module locations, and quality will be assured by identifying horizontal/vertical errors between adjacent modules. 3D object detection (3DOD) is a promising computer vision technology to realise 3D perception by detecting the presence of objects via labels, denoting the shapes via oriented 3D bounding boxes (bboxes), and determining the locations via transformed coordinates (Qian et al., 2022). This present paper selects the module as the target object due to its essential role in MiC site management.

Methods for 3DOD have evolved from traditional machine-learning techniques that require human-defined features to end-to-end deep-learning techniques that learn feature representations automatically from data (Liu et al., 2020). Although deep learning approaches to 3DOD have aroused significant attention, relevant research in construction is still in its infancy mainly due to two challenges: the difficulty of 3D data collection and the paucity of intelligent algorithms. This paper focuses on addressing the first challenge. A comprehensive dataset with sufficient data samples is imperative for training and testing deep learning models. Generally, there are two commonly used 3DOD modalities, i.e., images and point clouds. Some researchers have developed image datasets by collecting construction data using surveillance cameras or smartphones for detecting dynamic construction site objects (Lee et al., 2022, Otgonbold et al., 2022). However, image-based 3DOD performs significantly poorly as it is an ill-posed inverse problem by recovering 3D information from 2D input (Qian et al., 2022). Comparatively, point clouds are more reliable data sources that can provide precise 3D spatial information. However, developing a point cloud dataset is much more difficult than creating an image dataset, mainly suffering from three aspects, i.e., data collection, annotation, and sensor cost. First, it is time-consuming and cumbersome because lidar sensors need to be set up at several locations to obtain comprehensive data from different viewpoints. Moreover, lidar sensors should be located close to the target object to avoid large numbers of sparse points, which has potential safety risks and may affect ongoing construction activities. Thus, it is difficult to obtain point clouds of practical module installation processes. Second, the accuracy of data annotation may be impaired due to the uncertainties in calibrating the ground truth of 3D bboxes of modules. Researchers can quickly identify 2D bboxes of target objects in images, whereas the identification of 3D bboxes in point clouds depends on several factors, e.g., point cloud density, object similarity, and expert experience. Lastly, lidar sensors' prohibitive deployment cost in the current market also impedes the development of point cloud datasets.

To tackle these challenges, researchers have proposed pseudo-lidar point clouds, which convert input images into point cloud representations through depth estimation (Weng and Kitani, 2019). It has effectively bridged the performance gap between point cloud-based and image-based 3DOD (Weng and Kitani, 2019, Wang et al., 2020). Despite some preliminary efforts to adopt pseudo-lidar point clouds for 3DOD of construction equipment (Shen et al., 2021), relevant research in construction is still scarce, and well-recognised public point cloud datasets are lacking owing to unique construction projects and challenging data collection. This present paper aims to address these challenges by developing a pseudo-lidar point cloud dataset for 3D module detection (3DMD) during module installation processes. However, although real-life image data can be captured from MiC sites, the accuracy of depth estimation and uncertainties in data annotation significantly hinder the construction of a high-quality pseudo-lidar point cloud dataset. Virtual prototyping (VP) is a promising alternative to generate realistic and reliable data by simulating construction sites with the integration of Building Information Modelling (BIM) (Zheng et al., 2020), which enables the generation of precise and sufficient pseudo-lidar point clouds to facilitate model training. Therefore, this paper proposes a VP-enabled method to develop the pseudo-lidar point cloud dataset for 3DMD on MiC sites. The developed dataset involved 1024 pseudo-lidar point clouds and was validated using Frustum-PointNet (F-PointNet), a mainstream 3DOD algorithm. The proposed VP-based method is an effective innovation to strengthen point cloud datasets for 3DOD in construction. This study also facilitates construction automation by achieving point cloud-based 3DMD.

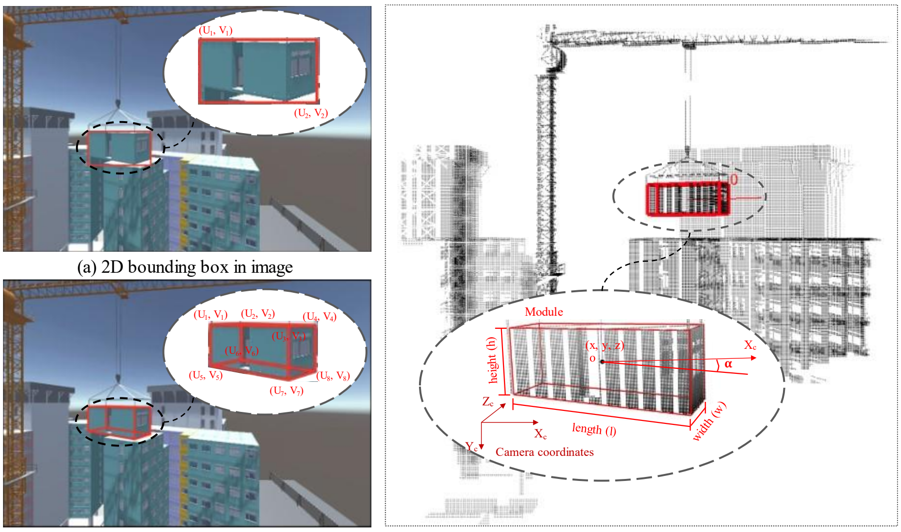
2. Literature Review

2.1 Module Installation Process Monitoring

MiC transforms most onsite works into factory prefabrication for producing complete modules, and module installation becomes one of the critical onsite activities for achieving successful project delivery. Monitoring module installation processes will enhance MiC site management, e.g., progress control, safety management, and quality assurance. However, current manual practices remain labour-intensive and time-consuming. To facilitate MiC automation, a classification model was proposed by Zhang et al. (2019) to automatically recognise different module installation stages for progress analysis and control. Furthermore, Zheng et al. (2020) developed a deep-learning model to detect the presence and determine the location of modules in images. The developed model proved effective for automatic progress monitoring, whereas the limited information provided by 2D bboxes was insufficient for accurate geometry assessment or reliable collision detection. Thus, it is crucial to explore 3DMD to monitor module installation processes comprehensively.

2.2 3D Object Detection in Construction

Computer vision techniques have been increasingly applied in construction to automate tedious and laborious tasks. Object detection is a fundamental application of computer vision to detect various construction objects. Construction site images collected using surveillance cameras or smartphones have proved to be effective in detecting instances of typical classes, e.g., helmets (Otgonbold et al., 2022), and excavators (Lee et al., 2022). However, most image-based studies have focused on 2D object detection using a 4-parameter bbox (Figure 1(a)) to locate objects in images. Some researchers have attempted to generate 3D bboxes in images (Figure 1(b)) by identifying critical vertices based on geometric constraints (Yan et al., 2020) or estimating the image depth (Shen et al., 2021). Nevertheless, compared with the detected data obtained from a 3D bbox in point clouds (Figure 1(c)), image-based 3DOD is still an ill-posed inverse problem by recovering 3D information from 2D input (Qian et al., 2022). Moreover, problems induced by image characteristics, e.g., object occlusion due to overlapping, 3D distortion due to 2D projection, and distance misjudgement due to different camera viewpoints, have potential construction safety risks.



(a) 2D bounding box in image

(b) 3D bounding box in image

(c) 3D bounding box in point cloud

Figure 1: Illustration of bounding boxes in image and point cloud

Point clouds have thus been utilised to detect specific construction objects to address the limitation of 2D images. Since point clouds are essentially irregular and unordered clusters of points without semantics, features are important means of object detection. Initially, geometry- and knowledge-based detectors were developed for recognising specific building elements from point clouds. For example, a local curvature-based shape descriptor was developed to extract pipe spools (Czerniawski et al., 2016) and a rule-based detector was devised to recognise walls, ceilings, and doors building on the common knowledge of orientations, openings, and spatial relationships (Wang et al., 2015). To further expand the targets to diverse object classes, machine learning was adopted to detect construction objects (e.g., construction equipment (Chen et al., 2018)) using manually defined 3D feature descriptors, which requires a small-scale dataset but suffers from low accuracy. The detection performance strongly depends on the manual feature selection. Consequently, the deep learning approach has attracted attention as it performs end-to-end detection without defining the features and can achieve high accuracy (Liu et al., 2020). However, it requires a larger-scale dataset and thus has yet to be explored in MiC projects because of data deficiency. Therefore, this paper proposes a VP-enabled method to address the challenge of collecting point clouds of module installation through simulation.

2.3 Dataset Generation Methods in Construction

As a high-quality dataset is fundamental to high-performance object detection tasks, many researchers have developed special image datasets for diverse construction objects. Apart from the direct use of public datasets, the identified dataset generation methods can be divided into three categories: site collection, online searching, and virtual modelling. Site collection aims to capture data from real-life construction sites using sensors (e.g., cameras and lidar sensors), while collecting point clouds of practical module installation processes is problematic as mentioned above. Online searching is also not applicable for this study as there is no such open-source point cloud data for onsite construction activities of MiC projects. Virtual modelling generates virtual data by establishing simulation models of varied construction objects. This approach was mostly adopted to supplement insufficient real-life data (Zheng et al., 2020, Lee et al., 2022). However, in some cases where real-life data were unavailable, virtual modelling was also used to generate a simulated dataset to support object detection, e.g., an image dataset of building structural components derived from a scale model (Hou et al., 2020) and a point cloud dataset of building components derived from a BIM model (Chuang and Sung, 2021). The virtual datasets were effective in achieving reliable object detection results. Therefore, this paper selects the virtual modelling method to develop a simulated point cloud dataset for 3DMD.

3. Research Methodology

This study adopted a four-step approach to develop the VP-enabled pseudo-lidar point cloud dataset for 3DMD on MiC sites. The overall research framework is presented in Figure 2. First, a VP model was established to simulate realistic module installation processes. Second, various parameters (e.g., module location, module size, and camera attitude) were considered to create different scenarios, from which colour and depth images were captured to generate pseudo-lidar point clouds. Third, the created dataset was automatically annotated, building on a popular dataset format. Finally, the dataset was trained and validated using a mainstream 3DOD algorithm. The steps and outcomes are described below.

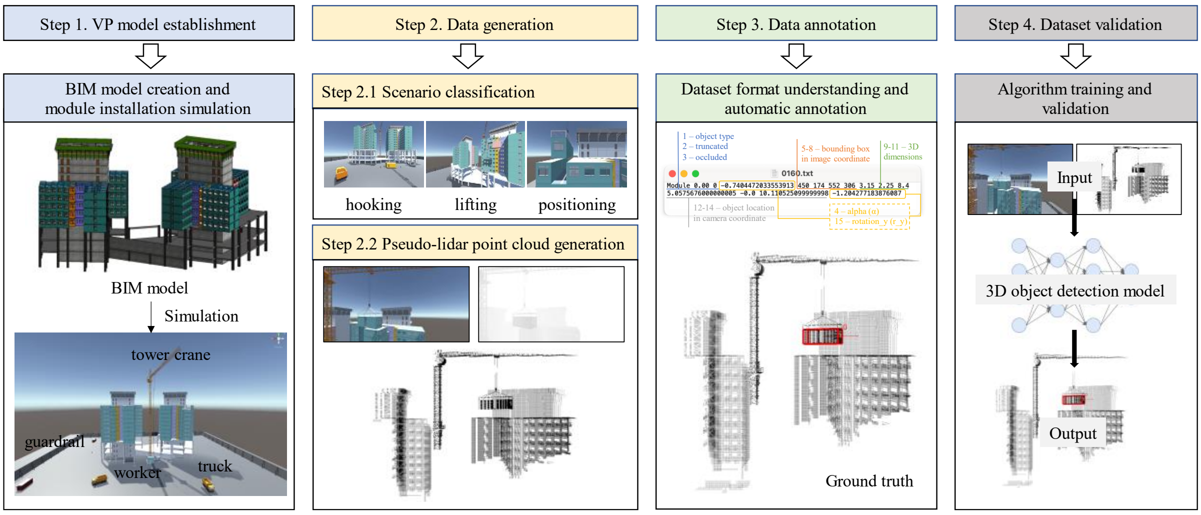


Figure 2: Overall research framework

3.1 Simulation Model Establishment

BIM offers informative 3D digital representations of construction projects with its powerful physical and functional characteristics and has been extensively adopted in construction (Xu et al., 2022). A 3D digital model of MiC buildings is imperative to simulate module installation. Therefore, a BIM model was established to represent a real-life MiC project in Hong Kong. Revit was selected for model creation in this study owing to its popularity and applicability. However, a static BIM model was insufficient to simulate the dynamic module installation processes, and a platform that provides development functions was required. Unity was thus selected, which is a popular game engine that can create 3D interactive experiences. The created BIM model was imported into the Unity environment, together with other construction objects, to simulate the real construction scene, e.g., workers, tower cranes, trucks, and guardrails.

3.2 Data Generation

Scenario Classification.

Different MiC scenarios were determined, considering various parameters to reflect real-life module installation processes, increasing the diversity and comprehensiveness of the developed dataset. Table 1 shows the parameters that constitute different scenarios, i.e., module type, module installation stage, module posture, and camera attitude, and Figure 3 illustrates the development process of a scenario. First, module-related variables were selected, including module type and installation stage. The selected MiC project adopted four types of modules with a height of 3.15m, a width of 2.25m, and a varied length of 5.55-8.4m. Different colours were assigned to denote the module types. Besides, the three typical module installation stages, namely, module hooking, lifting, and positioning (Zhang and Pan, 2020), were covered for each module type. Once these two variables were determined, a random point was selected based on the planned crane lifting path to obtain module location and rotation. Finally, eight virtual cameras were randomly distributed from different viewpoints (0° to 360°) and distances (5 to 60 m) to capture images. Consequently, eight data samples could be simultaneously generated from one scenario, accelerating the data collection and increasing the data diversity.

Table 1: Identified parameters to constitute different scenarios.

Parameters	**Values**
Module type	Type A, Type B, Type C, Type D
Module installation stage	Hooking, Lifting, Positioning
Module posture (location and rotation)	Random point selection in the planned path
Camera attitude (location and rotation)	5-60m, 0-360°

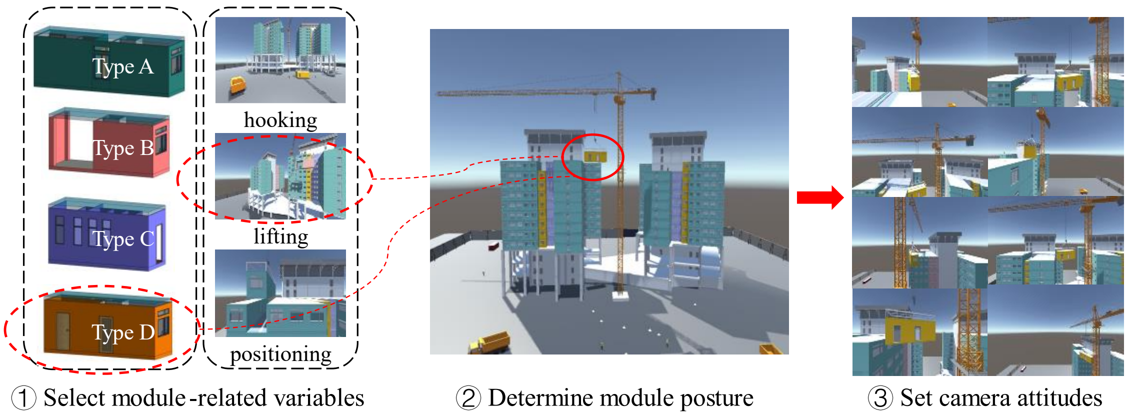


Figure 3: Development process of a scenario

Figure 4 depicts the profile of scenarios by module type and installation stage. A total of 128 scenarios were eventually set up, with 32 for each module type. Regarding installation stages, module lifting accounted for the largest portion of scenarios (47.66%) due to its longest distance, followed by module positioning (37.5%) and module hooking (14.84%). Module positioning sometimes takes a long time to ensure the assembly quality and was thus more considered in the scenarios.

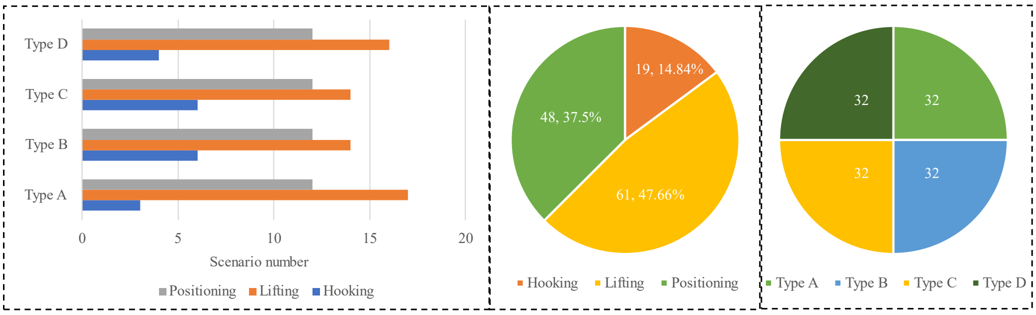


Figure 4: Profile of scenarios by module type and module installation stage

Pseudo-Lidar Point Cloud Generation.

A data generation pipeline (Figure 5) was proposed based on the established scenarios to generate pseudo-lidar point clouds. Colour and depth images captured from Unity were inputted to merge into RGBD images using the open-source library Open3D. Then, a multi-coordinate transformation was performed to obtain the pseudo-lidar point cloud. A clipping box was used in Unity to define the field of view of cameras, and the 3D objects were initially in the camera coordinate system (CCS) (Figure 5(a)). Then, a homogeneous transformation matrix was utilised to change the points from CCS to the clip coordinate system (CLCS) (Figure 5(b)). Before projecting the points onto the screen, the computer automatically transformed the CLCS to a normalised device coordinate (NDC) (Figure 5(c)) by projective division to remain points that were included in the clipping box. Finally, the points were distributed in the RGBD image coordinate system (ICS) (Figure 5(d)). Two specific points (A and B) are highlighted in Figure 5 to illustrate the multi-coordinate transformation, and the equations used are presented in Table 2. A script was developed to reversely transform points from ICS to CCS for generating pseudo-lidar point clouds. A total of 1024 data samples were finally captured and generated.

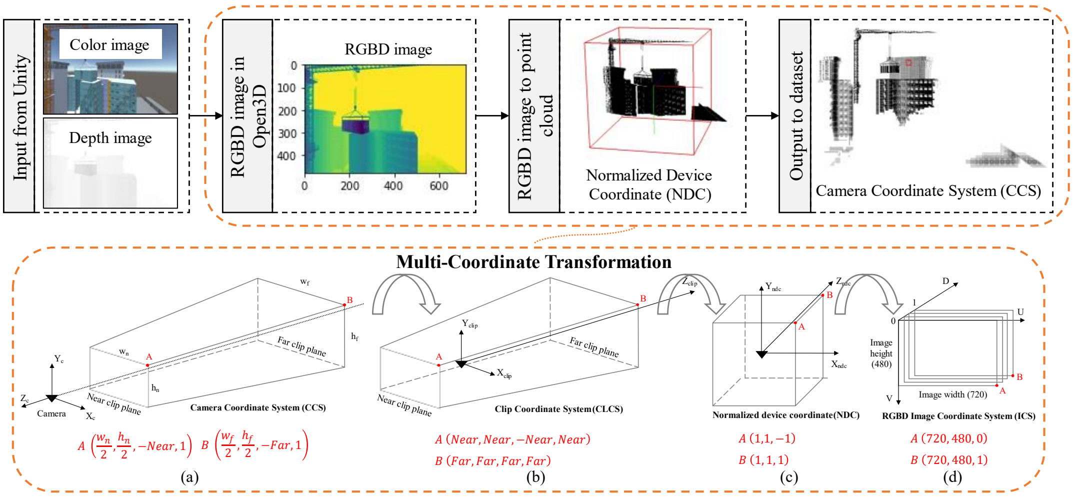


Figure 5: Data generation pipeline for pseudo-lidar point clouds

Table 2: Multi-coordinate transformation

Coordinates	Point expressions
Camera Coordinate System	(X_c, Y_c, Z_c): ground truth in Unity
Clip Coordinate System	$\begin{bmatrix} X_{clip} \\ Y_{clip} \\ Z_{clip} \\ W_{clip} \end{bmatrix} = \begin{bmatrix} \frac{2Near}{w} & 0 & 0 & 0 \\ 0 & \frac{2Near}{h} & 0 & 0 \\ 0 & 0 & \frac{-(Far+Near)}{Far-Near} & \frac{-2Far \times Near}{Far-Near} \\ 0 & 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} \quad (1)$
Normalised Device Coordinate	$\begin{bmatrix} X_{ndc} \\ Y_{ndc} \\ Z_{ndc} \end{bmatrix} = \begin{bmatrix} X_{clip} \\ Y_{clip} \\ Z_{clip} \end{bmatrix} / (-Z_c) = \begin{bmatrix} \frac{-2Near}{w \times Z_c} X_c \\ \frac{-2Near}{h \times Z_c} Y_c \\ \frac{(Far+Near)Z_c + 2Far \times Near}{(Far-Near)Z_c} \end{bmatrix} \quad (2)$
RGBD Image Coordinate System	$\begin{bmatrix} U \\ V \\ D \end{bmatrix} = \begin{bmatrix} (X_{ndc}+1)\frac{width}{2} \\ (Y_{ndc}+1)\frac{height}{2} \\ \frac{Z_{ndc}+1}{2} \end{bmatrix} \quad (3)$

Note: $Near$ and Far refer to the distance between the near/far clip plane and camera, respectively; w and h denote the width and height of the clip plane; $width$ and $height$ represent the width and height of the image.

3.3 Data Annotation

Data annotation is a fundamental task of object detection, and its accuracy will significantly influence the performance of deep learning models. Unlike data annotation for 2D object detection, which requires the object class and bbox coordinates in images, 3D data annotation is more complicated due to numerous irregular and unordered points. Therefore, a well-recognised dataset format is essential to help 3D data annotation. The format of the KITTI (Karlsruhe Institute of Technology and Toyota Technological Institute) dataset (Geiger et al., 2013) was adopted in this study, which is one of the most popular datasets used for 3DOD in autonomous driving. A label file involving 15 parameters is used to annotate data for 3DOD tasks (Figure 6). The first three parameters describe the object type, truncation, and occlusion degree of the object. The 5^{th}-8^{th} parameters depict the 2D bbox in the image. The 3D dimensions and location of the object are provided in the 9^{th}-11^{th} and 12^{th}-14^{th} parameters, respectively. Besides, two parameters denote the object orientation, namely the observation angle (alpha) and rotation angle (rotation_y) of the object in CCS, as illustrated in Figure 6. The label file will be processed to obtain the ground truth value of 3D bboxes of the objects in point clouds.

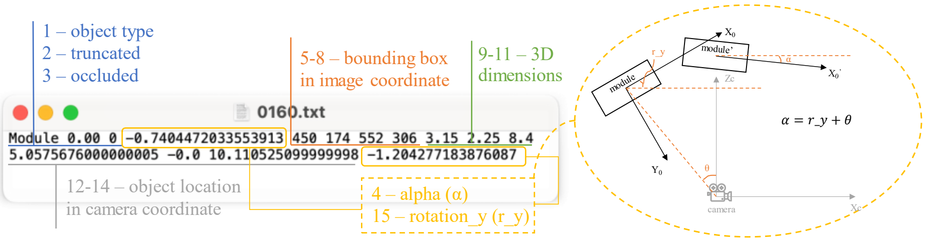


Figure 6: Data annotation file based on the KITTI dataset

Based on a complete understanding of the KITTI dataset format, our developed dataset was automatically annotated in conformity with the defined parameters. A Python-based script was developed to derive the 2D bbox and orientations from Unity's ground truth values of 3D dimensions and location. An essential benefit of the proposed VP-based dataset generation method was that the ground truth was available, and a 100% accurate annotation could be made regardless of complexities such as occlusion and truncation. Figure 7 provides examples of the 3D annotation results in both normal and special cases.

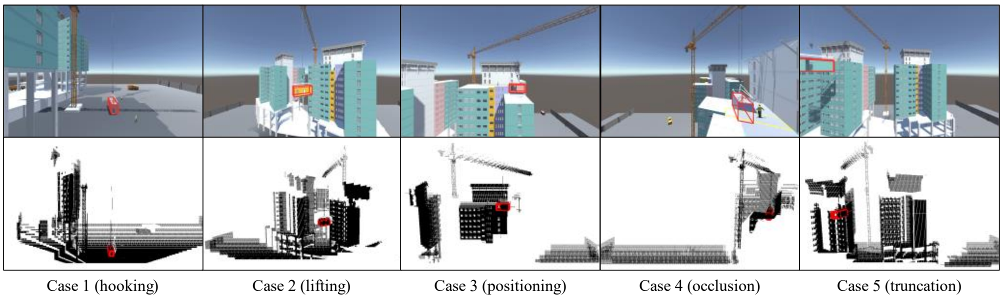


Figure 7: Examples of the 3D annotation results in different cases

3.4 Dataset Validation

Experiments were conducted using the F-PointNet model to validate the feasibility of the proposed VP-enabled pseudo-lidar point cloud dataset. F-PointNet was selected as it is one of the most popular and well-known 3DOD models, which creates frustum point clouds based on the 2D proposal from the image and then estimates the 3D bbox in the frustum (Qi et al., 2018). To evaluate the model performance, the average precision (AP) value was adopted as an overall detection performance indicator, which could be calculated from the precision-recall (P-R) curves. The experiment was conducted in a Linux system of a desktop computer with an Intel(R) Core (TM) i9-10900X CPU and an ASUS ROG RTX3090 GPU. Due to the small size of the dataset, it was randomly split into 80% training samples and 20% validation samples empirically. The model was trained with parameters as follows: the number of points was 1024; the learning rate and momentum were 0.0001 and 0.9, respectively; the batch size and the number of epochs were set to 32 and 150, respectively. To evaluate the impact of dataset size, the model was further trained on an enhanced dataset with data augmentation methods such as randomly shifting box centre and scaling width and height.

4. Results and Discussion

The experimental results are presented in Figure 8. The 3D bird view refers to the top view of 3D bboxes in point clouds, which is usually adopted in autonomous driving. The intersection of union (IoU) was set to 0.5 in this study. The developed dataset achieved maximum AP values of 89% and 57.11% from the 3D bird and 3D views, respectively. Table 3 compares the F-PointNet model's performance in the previous study and this study. Several findings can be derived from the results. First, the comparison showed a 6.5% increase in AP from the 3D bird view while a 19.5% decrease in AP from the 3D view. Meanwhile, the AP of module detection in the 3D bird view was higher than that in the 3D view, indicating that larger deviations were found in the vertical dimension (axis Y in CCS). This finding is mainly attributed to the ignorance of axis Y since the 3D bird view is, in essence, the 2D view from the top. The F-PointNet model was proposed for 3DOD in autonomous driving, in which the objects (e.g., pedestrians, cars, and cyclists) all stand on the road with few variations in the vertical dimension. Therefore, axis Y was less restricted in the loss functions, leading to significant errors in the case of dynamic module installation. Further investigation on model mortification is required to adapt the model to module installation scenarios and improve the overall detection performance. Second, data augmentation is an important and useful method to improve object detection performance with a small-scale dataset. The AP of module detection in the 3D bird and the 3D view increased by 68% and 71%, respectively. Since only a small dataset was developed in this study, it is believed that the model could be further improved with more data produced.

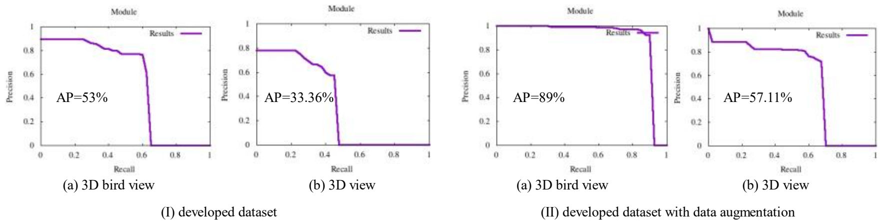


Figure 8: Experimental results of model training and validation

Table 3: Object detection performance of F-PointNet model in different datasets

Dataset	**Ours**	**Ours with data augmentation**	**KITTI (Qi et al., 2018)**
AP (3D bird view)	53%	89%	83.53%
AP (3D view)	33.36%	57.11%	70.92%

5. Conclusions

This paper proposes a VP-enabled method to develop a pseudo-lidar point cloud dataset for 3DMD during module installation. The dataset was created through three steps of simulation model establishment, data generation, and data annotation and was tested using the F-PointNet model. The results demonstrate the feasibility and effectiveness of using pseudo-lidar point clouds for 3DMD, with achieved AP as high as 89%.

The contributions of this paper are threefold. First, it proposes a VP-enabled approach to addressing the challenges of point cloud data collection and annotation in construction and develops a dataset for 3DMD as an example. This approach can serve as an innovative method to strengthen the comprehensiveness of image/point cloud datasets in construction. Second, this paper should contribute to the 3D perception of the module installation process by realising 3DMD with an oriented 3D bbox, which can facilitate effective MiC site management with enhanced safety, efficiency, and quality. Third, this paper successfully extends the object detection task in the construction sector from 2D to 3D, providing a methodological foundation for the 3D detection of other construction objects (e.g., workers, tower cranes, and trucks).

Nevertheless, there are limitations of this study which warrant future research. First, the developed small-scale dataset falls short to achieve a very high performance of module detection. The dataset should thus be enhanced by including more module types (e.g., L-shape modules) from diverse MiC projects. Real point cloud data should also be captured as data samples despite the difficulty in collection and annotation. Second, more 3DOD models should be selected to validate and benchmark the dataset by adapting to MiC scenarios.

References

Chen, J., Fang, Y. and Cho, Y. K. (2018). Performance evaluation of 3D descriptors for object recognition in construction applications, Autom. Constr. 86**,** pp. 44-52.

Chuang, T. and Sung, C. (2021). Learning-guided point cloud vectorization for building component modeling, Autom. Constr. 132, pp. 103978.

Czerniawski, T., Nahangi, M., Haas, C. and Walbridge, S. (2016). Pipe spool recognition in cluttered point clouds using a curvature-based shape descriptor, Autom. Constr. 71, pp. 346-358.

Geiger, A., Lenz, P., Stiller, C. and Urtasun, R. (2013). Vision meets robotics: The kitti dataset, Int. J. Rob. Res. 32, pp. 1231-1237.

Hou, X., Zeng, Y. and Xue, J. (2020). Detecting Structural Components of Building Engineering Based on Deep-Learning Method, J. Constr. Eng. Manag. 146, pp. 04019097.

Lee, J., Hwang, J., Chi, S. and Seo, J. (2022). Synthetic Image Dataset Development for Vision-Based Construction Equipment Detection, J. Comput. Civ. Eng. 36, pp. 04022020.

Liu, L., Ouyang, W., Wang, X., Fieguth, P., Chen, J., Liu, X. and Pietikäinen, M. (2020). Deep Learning for Generic Object Detection: A Survey, Int. J. Comput. Vis. 128, pp. 261-318.

Otgonbold, M.-E., Gochoo, M., Alnajjar, F., Ali, L., Tan, T.-H., Hsieh, J.-W. and Chen, P.-Y. (2022). SHEL5K: An Extended Dataset and Benchmarking for Safety Helmet Detection, Sensors. 22, pp. 2315.

Pan, W. and Hon, C. K. (2020). Briefing: Modular integrated construction for high-rise buildings, Proceedings of the Institution of Civil Engineers - Municipal Engineer. 173, pp. 64-68.

Qi, C. R., Liu, W., Wu, C., Su, H. and Guibas, L. J. (2018). Frustum pointnets for 3d object detection from rgb-d data, Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 918-927.

Qian, R., Lai, X. and Li, X. (2022). 3D Object Detection for Autonomous Driving: A Survey, Pattern Recognit. 130, pp. 108796.

Shen, J., Yan, W., Li, P. and Xiong, X. (2021). Deep learning-based object identification with instance segmentation and pseudo-LiDAR point cloud for work zone safety, Comput.-Aided Civ. Infrastruct. Eng. 36, pp. 1549-1567.

Wang, C., Cho, Y. K. and Kim, C. (2015). Automatic BIM component extraction from point clouds of existing buildings for sustainability applications, Autom. Constr. 56, pp. 1-13.

Wang, Y., Chao, W.-L., Garg, D., Hariharan, B., Campbell, M. and Weinberger, K. Q. (2020). Pseudo-lidar from visual depth estimation: Bridging the gap in 3d object detection for autonomous driving, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8445-8453.

Weng, X. and Kitani, K. (2019). Monocular 3d object detection with pseudo-lidar point cloud, Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. pp. 1-14.

Xu, J., Teng, Y., Pan, W. and Zhang, Y. (2022). BIM-integrated LCA to automate embodied carbon assessment of prefabricated buildings, J. Clean. Prod. 374, pp. 133894.

Yan, X., Zhang, H. and Li, H. (2020). Computer vision-based recognition of 3D relationship between construction entities for monitoring struck-by accidents, Comput.-Aided Civ. Infrastruct. Eng. 35, pp. 1023-1038.

Zhang, Z. and Pan, W. (2020). Lift planning and optimization in construction: A thirty-year review, Autom. Constr. 118, pp. 103271.

Zhang, Z., Pan, W. and Zheng, Z. (2019). Transfer Learning Enabled Process Recognition for Module Installation of High-rise Modular Buildings, Modular and Offsite Construction (MOC) Summit Proceedings. pp. 268-275.

Zheng, Z., Zhang, Z. and Pan, W. (2020). Virtual prototyping- and transfer learning-enabled module detection for modular integrated construction, Autom. Constr. 120, pp. 103387.

Acknowledgement

This work was supported by the Research Impact Fund of the Hong Kong Research Grants Council (Project No. HKU R7027-18). Also acknowledged are the Estates Office of The University of Hong Kong and Paul Y. Engineering Group Limited for providing data access.