# Automatic Concrete Pipe Material Detection from Mobile Laser Scanning Data Using Transfer Learning

**Kartika Nur Rahma Putri[1,2], Khalegh Barati[2], Xuesong Shen[2], James Linke[3]**

[1]Department of Civil and Environmental Engineering, Universitas Gadjah Mada, Indonesia
[2]School of Civil and Environmental Engineering, University of New South Wales, Australia
[3]GeoAI, Sydney, NSW 2019, Australia

kartikanurrahmaputri@ugm.ac.id, khalegh.barati@unsw.edu.au, x.shen@unsw.edu.au, j.linke@geoai.au

**Abstract**

**The current prefabricated material tracking method in construction projects usually uses RFID or GPS technology. This technique requires tagging materials, which can be labour-intensive for large volumes of materials on a construction site. Deep learning-based object detection can provide a solution for material tracking in construction projects without the need for material tagging. To obtain this result, a deep learning architecture must be trained to automatically detect construction materials. However, there is a limited dataset of construction material objects in outdoor environments for training. Therefore, this study proposes a framework for material detection at construction sites using a modified dataset and transfer learning from other LiDAR data mainly used for autonomous driving (e.g., cars, pedestrians, and cyclists). The method includes workflows for data acquisition using a Mobile Laser Scanner (MLS) system, data preprocessing, training, and testing. A site-specific case study using concrete pipe material was used to validate the proposed approach. The technical performance of the material object detection model trained with transfer learning increased from 35.40% F1 score to 80.56% F1 score. Furthermore, with an additional modified dataset, the F1 score increased to 87.18%.**

**Keywords**

**Construction material detection; 3D point cloud; mobile laser scanning; transfer learning; deep learning**

## 1 Introduction

Tag-based material-tracking technologies, such as Radio Frequency Identification (RFID), require significant manual effort and are often prone to errors and inefficiencies. For instance, workers must manually attach tags to materials and traverse construction sites to scan and extract information at close range. On large construction sites, this process becomes time-consuming and error-prone because of the widespread distribution of materials [1]. Furthermore, positioning technologies such as RFID, ultra-wideband (UWB), infrared, and Global Positioning System (GPS) often depend on additional infrastructure such as Wi-Fi, cellular networks, or Bluetooth to function effectively. This dedicated infrastructure requirement increases the complexity and cost of tag-based methods.

Considering the current challenges of using RFID, UWB, and GPS, which require manual tagging and dedicated infrastructure for data transfer, the application of visual perception technology has gained significant interest. Some technologies, such as image-based technology and point cloud processing, do not require dedicated infrastructure [2]. Although image-based methods for object detection are promising, they only capture 2-dimensional (2D) data, making it difficult to obtain information about the material's 3-dimensional (3D) size. To address this, point cloud data is a possible alternative, as it provides detailed 3D information about the material at a construction site. However, processing large point cloud datasets often requires significant computational power, which may delay real-time applications of the technology.

The current point cloud data used in construction projects is mostly generated by Terrestrial Laser Scanners (TLS), which produce highly accurate, dense point clouds. The TLS method is considered more suitable for scanning small and dense areas. However, for an infrastructure project with a large area, such as a toll road or airport, TLS data generation can be too slow. On the other hand, Mobile Laser Scanning (MLS) provides faster data acquisition for large areas [3], but faces challenges with occlusion and sparse point clouds.

To address the challenges of detecting construction materials from mobile laser scanning (MLS) data, this study proposes a transfer-learning-based framework. A real dataset from a large infrastructure project was used to evaluate the proposed approach's feasibility. The data were collected using an MLS system and pre-processed through subsampling, segmentation, and data labelling.

A Point Voxel Region-based Convolutional Neural Network (PV-RCNN) [4] was used for 3D object detection. Transfer learning was applied by initialising the model with weights pre-trained on an autonomous driving dataset and then fine-tuning it using the construction MLS data. The results were compared with a model trained from scratch using the same dataset to evaluate the benefit of transfer learning.

## 2 Related Work

This section provides an overview of current material object detection in construction projects. It highlights relevant studies on the processing of 3D point cloud data and on the application of deep learning methods for the detection of construction materials.

### 2.1 Construction Material Recognition Method

To support material recognition in construction, several technologies, including 3D object detection and vision-based scene understanding, have been applied in construction projects. One study demonstrated a vision-based method for material recognition to automate construction progress monitoring and generate Building Information Models (BIM) from unordered site image collections [5]. This approach utilised 3D point cloud models created from site photo logs using Structure-from-Motion (SfM) techniques. The primary aim was to classify material types based on appearance to facilitate progress tracking. This approach primarily targets materials that have already been installed on site, which limits its applicability for detecting and managing on-site material inventory.

In terms of material quantification, one recent research developed a method for automatic measurement of corrugated pipes and rebars for large precast concrete beams from a TLS [6]. The segmentation framework was based on the application of radius nearest-neighbours, covariance, and a density-based spatial clustering algorithm to detect pipes and rebar and to extract their sizes, including radius and length. Although the algorithm accurately extracted the object size with 98% accuracy, the method still required a manual clutter-removal stage and relied on dense TLS data.

Previous work introduced a vision-based machine learning method for counting steel pipes on construction sites [7]. This method utilised a Convolutional Neural Network (CNN) combined with image processing technologies to enhance counting accuracy and reduce labour intensity. The primary aim was to automate the counting process to improve efficiency and accuracy. Despite its effectiveness, the method may face challenges in highly cluttered environments where steel pipes overlap significantly.

3D point cloud object detection methods in the construction industry, including vision-based recognition, semantic segmentation, and volumetric measurement, have significantly enhanced monitoring and material management. However, challenges such as occlusion, noise, and data completeness, especially for the MLS point cloud, remain unresolved. Therefore, further research and development to improve the robustness of these techniques are needed to ensure their applicability across diverse construction environments.

### 2.2 3D Point Cloud Processing

In automatic 3D object detection for point cloud processing, three methods are generally used. The methods are simple geometric models, feature descriptors, and deep learning [8]. The first method assumes that objects can be detected by utilising simple geometric models, such as planes and cylinders, from point cloud data. This method usually uses random sample consensus (RANSAC), which computes the least-squares error between the known object and the points in the point cloud, or the Hough transform, which searches for the object using its parametric shapes, such as the radius or slope-intercept.

The second method uses feature descriptors divided into local and global categories. The local descriptor tries to summarise statistics of the object's features, such as curvature, density, and normal, to describe the neighbourhoods around interest points. Meanwhile, the global descriptors define objects using their lengths, areas, and angles in a single feature vector. Examples of research using global descriptors include [9] that uses the Principal Axes Descriptor to automatically detect construction equipment from a TLS point cloud. In the context of material detection, recent studies have explored the use of geometric feature extraction from mobile laser scanning point clouds to enable tracking and volume estimation of precast concrete pipes [10]. While this method achieved 78.5% accuracy, its performance indicates room for improvement in material object detection.

The last method of object detection uses deep learning to find a direct mapping from the 3D point cloud input to its object labels. The deep learning method requires extensive training on a large number of objects before it can correctly predict the type of object to be detected. Compared to traditional geometric feature extraction, deep learning offers advantages in robustness. One key benefit of deep learning is that it minimises the need for manually engineered feature descriptors and employs more flexible yet robust classification rules [8].

However, there are some concerns about the applicability of machine learning data, especially related to the amount of data required to train the model.

In the construction industry, the most publicly available point cloud dataset with its labelling is an indoor dataset, such as the Stanford Large-Scale 3D Indoor Spaces Dataset (S3DIS) [11] and the ScanNet [12] dataset. Meanwhile, another point cloud dataset in outdoor settings is usually used for the autonomous driving process, such as the Karlsruhe Institute of Technology and the Toyota Technological Institute (KITTI) [13]. This dataset class usually contains cars, pedestrians, trees, and cyclists.

Despite growing interest in point cloud technology for construction applications, its use for automated material tracking remains limited. Most existing studies focus on progress monitoring using TLS. While MLS, which is faster and more practical for large infrastructure, generates sparse and occluded point clouds that are more challenging to process. Furthermore, most existing 3D object detection algorithms and publicly available datasets are developed for autonomous driving, covering common classes such as cars and pedestrians, rather than construction-specific materials such as pipes, concrete, and steel. Existing MLS-based approaches for pipe detection have also shown limited accuracy. As a pilot project, this study aims to develop and test a deep learning framework trained on custom MLS data for automated concrete pipe detection under the challenging conditions of single-pass MLS scanning.

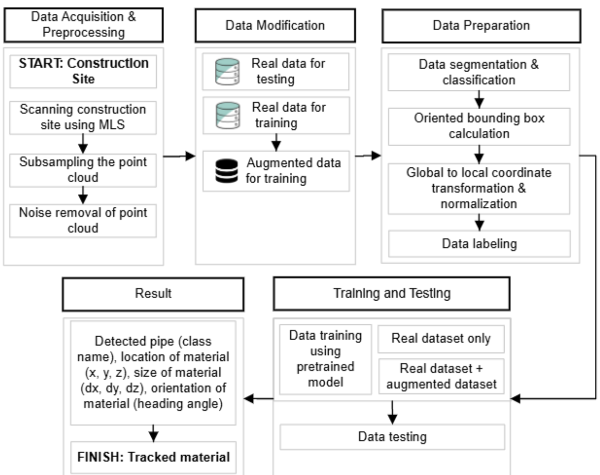



Figure 1. A step-by-step method for 3D point cloud object detection in a construction project.

# 3 Methods

The proposed methodology aims to detect prefabricated concrete pipes from MLS point cloud data. Figure 1 illustrates the 3D material object detection step-by-step method used in this paper. Point cloud data from a real construction site was used as a case study. From this dataset, only 8 areas can be used. To increase the amount of training data, a preprocessing stage was introduced. Then, a transfer learning approach for 3D object detection was used to detect the pipe.

## 3.1 Data Acquisition

The 3D point cloud data was obtained using MLS in a single pass with the specifications mentioned in Table 1. MLS was selected in this study due to its fast data acquisition capability, which allows large areas to be scanned in a single continuous pass. Since the MLS system performs only one scan along the road, many of the data is dense on one side and sparse on the other. The resulting data was large, noisy, and mostly occluded, as can be seen in Figure 2. Although multi-pass scanning could improve point density, it was not possible due to this constraint. Furthermore, single-pass data collection reflects how MLS is used in practice, where fast data acquisition is one of its main advantages. Testing our method under these conditions shows that it can work well in real inspection scenarios. To obtain data for labelling and training, the area containing prefabricated concrete pipes was segmented into distinct regions. The initial data for each area includes concrete pipes, ground points, other materials around the pipes, and the scene.

Table 1. Mobile laser scanner specification.

| Criteria | Specification |
|---|---|
| Laser Sensor | XT32 |
| Range Accuracy | ±3cm |
| Vertical Field of View | $-16^{o} \sim 15^{o}$ |
| Horizontal Field of View | $0^{o} - 360^{o}$ |
| Maximum Range | 120 m |

Figure 2. 3D point cloud data.

Point cloud subsampling was performed to reduce the number of points and reduce computational requirements. The subsampling method was obtained using open-source cloud-compare software. To ensure that point features are not removed, the subsampling method used a spatial mode with a minimum distance of 0.02 meters between points. This subsampling method reduced up to 50% of the points for each area.

### 3.2 Data Preprocessing and Modification

The dataset preparation was obtained to increase the number of areas for data training. From the data-acquisition stage, 8 areas containing prefabricated concrete pipes formed the real dataset for this case study as can be seen in Figure 3. The real dataset was partitioned into training and test sets, with six areas allocated to training and two to testing. To ensure the integrity of the evaluation process, the testing data were entirely excluded from both the training phase and the generation of modified data. This approach prevents the model from acquiring any prior knowledge of the testing dataset's properties.

The additional dataset was obtained by modifying the current dataset. The ground points of each dataset were segmented and deleted to get the modified dataset (Figure 4). Next, only the good pipe from each dataset is segmented without the ground points and scene (Figure 5). Both modifications to the datasets aim to train the model to recognise the correct pipe in a clean environment. From this stage, the training data increased from 6 to 18 areas.

Figure 3. Example of the current dataset.

Figure 4. Current dataset without ground point.

Figure 5. Good pipes only dataset.

### 3.3 Dataset Preparation

All areas were prepared to match the algorithm. The data was segmented into pipes and scenes. A bounding box was generated for each pipe and the scene. An oriented bounding box (OBB) was used to capture the size and orientation of each pipe. Since the concrete pipes are located on the ground, only the heading angle (also known as the yaw angle) was calculated. The pitch and roll angles were assumed to be zero [13]. The heading angle describes the rotation of the object around the vertical (z) axis and indicates the direction the object faces horizontally. The illustration of the object bounding box, coordinates and heading angle can be seen in Figure 6. Each object was labelled with its centre coordinate (x, y, z), 3D size generated from its bounding box (length: dx, width: dy, height: dz), heading angle, and class name.

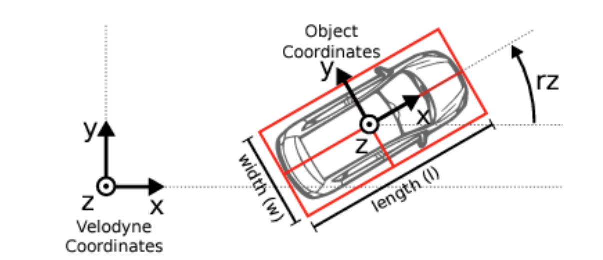



Figure 6. Object coordinates illustration.

Given that the PV-RCNN algorithm utilises both point and voxel features, the intensity data must also be included in the point cloud dataset. Additionally, the data needs to be verified for compliance with the point cloud range and voxel size specified in the OpenPCDET custom dataset information. Each point from the laser scanner data contains information about its global coordinates. Since the area is divided into regions containing pipes, the coordinate information was transformed to local coordinates. To ensure the point cloud range was not too large, each data point was transformed again so that the first point in each point cloud was located around the 0, 0, 0 coordinate (normalisation).

### 3.4 Dataset Training and Testing

The methodology described was implemented in a Python environment on a Personal Computer with an Intel(R) Core (TM) i7-6700 CPU@ 3.4 GHz processor and installed RAM of 64 GB. The model training time ranged from 20 minutes to 1 hour depending on the dataset size and configuration and to evaluate a single scan took approximately 30 to 40 seconds to process. The primary library used for 3D processing was Open3D. From 8 real datasets, areas 1 and 5 were selected as the test datasets. The baseline for object detection was calculated by training only on a real dataset (6 areas) and testing on 2 areas. The next scenario involved adding a pre-trained model to the dataset and an additional modified dataset for training.

The model's performance was evaluated using precision, recall, and F1 scores as defined in equations 1, 2, and 3. Precision measures the proportion of detected objects that are correct, calculated as the ratio of true positives (TP) to the sum of true positives (TP) and false positives (FP). Recall measures how many of the actual objects are correctly detected, defined as the ratio of true positives to the sum of true positives and false negatives (FN). In this context, TP represents correctly detected objects, FP refers to incorrect detections, and FN indicates objects that were present but missed by the model. The F1 score combines precision and recall into a single metric by computing their harmonic mean.

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3)$$

## 4 Results and Discussion

Each dataset configuration was trained separately, and performance was measured at each epoch. This analysis highlights variations in results across different data types and demonstrates the impact of pre-trained models on 3D object detection performance.

### 4.1 Real Dataset

From 8 real datasets, Areas 1 and 5 were selected as the test datasets. Area 1 represents a region with high-quality pipe data and minimal noise, while Area 5 contains more occlusions and noise. As a baseline, the model was trained on only six real datasets and evaluated on the designated test data. The Ground Truth (GT) bounding box, shown in Figure 7 indicates that Area 1 contains 14 pipes and Figure 8 indicates that Area 5 contains 19 pipes.

Figure 7. Ground truth for Area 1

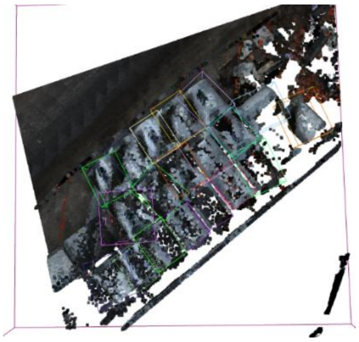


Figure 8. Ground truth for Area 5

In the first experiment, PV-RCNN was trained using six real MLS datasets without pretraining. The baseline detection results, shown in Figure 9 (a) and Figure 9 (b), indicate poor performance due to the limited number of training samples. The model failed to accurately detect pipe objects, with predicted bounding box centres and orientations misaligned from the ground truth. In addition, clutter was frequently misclassified as pipes, and several actual pipes were missed, leading to unreliable detections.

Intersection over Union (IoU) measures the overlap between a predicted bounding box and the ground truth to evaluate object detection accuracy. The IoU 50% means detection is considered correct if at least 50% of the predicted box overlaps with the ground truth. In terms of quantitative performance, object detection using only real data resulted in a precision of 25% and a recall of 60.6% for IoU 50%. Although no pipes were correctly detected according to the visualisation criteria, the precision, recall, and F1 scores at the 50% threshold were non-zero. This is because some predicted bounding boxes partially overlapped with ground-truth pipes and were counted as true positives under the relatively loose 50% IoU criterion, despite their inaccurate geometry.

After incorporating a pre-trained model, detection performance improved substantially. In Area 1, the model correctly detected 13 out of 14 pipes, with bounding box locations, dimensions, and orientations closely aligned with the ground truth, as shown in Figure Figure 9 (c). In the more challenging Area 5, which contains higher noise and clutter, the model detected 17 out of 19 pipes, with two clutter objects misclassified as pipes as can be seen in Figure 9 (d). When the pre-trained model was introduced, precision improved to 72.73%, while recall increased to 90.28%. The recall of 90.28% indicates that the model correctly identified most of the actual pipes, though some remained undetected. These results highlight the effectiveness of integrating a pre-trained model with real data to improve object detection performance compared to the baseline approach.

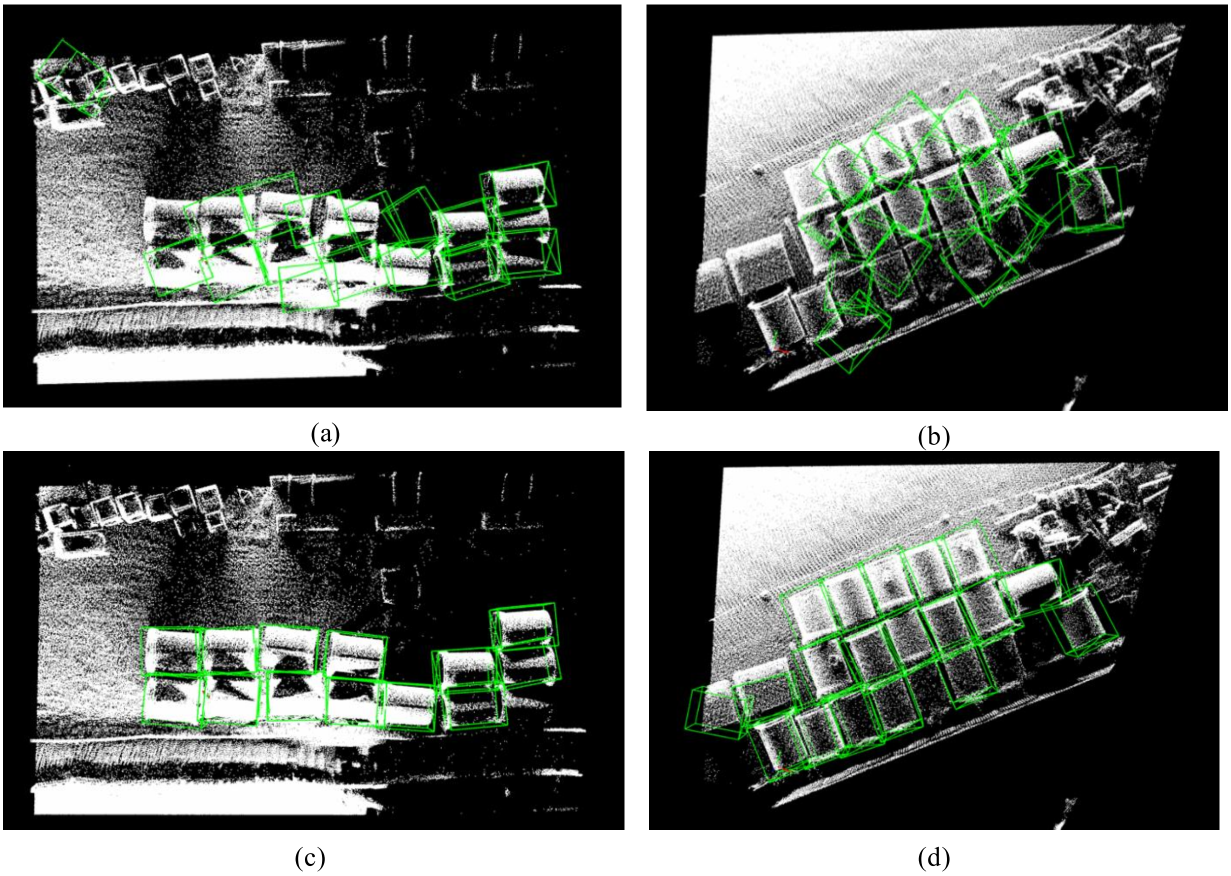



Figure 9. Predicted bounding box from 6 real data: (a) Area 1 without pre-trained model; (b) Area 5 without pre-trained model; (c) Area 1 with pre-trained model; and (d) Area 5 with pre-trained model.

### 4.2 Real Dataset plus Modified Dataset

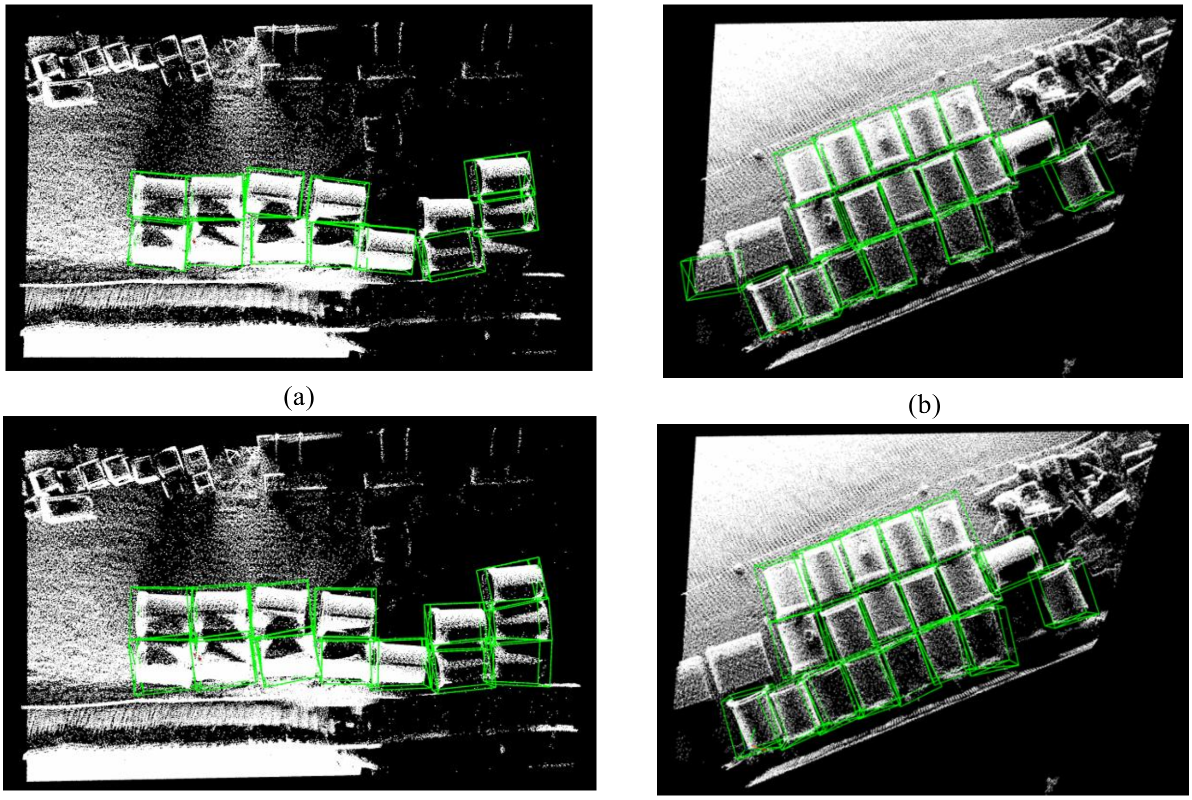
The modified datasets were generated to increase the amount of training data. As a result, an additional 12 data points were added to the real data. The remaining areas were then trained using a pre-trained PV-RCNN model for 80 epochs. As shown in Figure 10, the introduction of additional modified datasets significantly improved the results. The algorithm successfully detected 100% of pipes in both Area 1 (14 out of 14 pipes) and Area 5 (19 out of 19 pipes), with no misclassifications. The incorporation of modified datasets into the object detection model increased the F1 score from 80.56% to 87.18%. While the results for Area 1 showed minimal differences between using only real datasets and real plus modified datasets in terms of the number of detected pipes and bounding box size, the improvements are more evident in Area 5. The bounding box locations and orientations in Area 5 became more aligned with the ground truth, and fewer objects were misclassified. The result of object detection performance for all datasets can be seen in Table 23.

The results at IoU 70% are slightly lower than those at IoU 50%, which is expected since a stricter threshold requires more precise bounding box localization. Nevertheless, both thresholds show a consistent improvement in detection performance. At IoU 50%, the F1 score improves from 35.40% in the baseline to 87.18% in the best configuration. Similarly, at IoU 70%, the F1 score increases from 0.00% to 81.97%. This result demonstrates the effectiveness of integrating modified data in refining the model’s performance. The poor performance of the baseline model highlights the sensitivity of the proposed methods to the availability of labelled training data. Detection performance also varied across test areas, depending on scene complexity.

Table 2. 3D object detection performance for all datasets.

| Real Data | Modified Data | Training type | Precision (50%) | Recall (50%) | F1 (50%) | Precision (50%) | Recall (50%) | F1 (50%) |
|---|---|---|---|---|---|---|---|---|
| 6 real data | - | - | 25.00% | 60.60% | 35.40% | 0.00% | 3.03% | 0.00% |

| Real Data | Modified Data | Training type | Precision (50%) | Recall (50%) | F1 (50%) | Precision (50%) | Recall (50%) | F1 (50%) |
|---|---|---|---|---|---|---|---|---|
| (baseline) | | | | | | | | |
| 6 real data | - | Pretrained model | 72.73% | 90.28% | 80.56% | 50.17% | 81.81% | 62.20% |
| 6 real data | 6 modified data | Pretrained model | 72.42% | 93.93% | 81.78% | 36.02% | 69.69% | 47.49% |
| 6 real data | 12 modified data | Pretrained model | 77.27% | 100.00% | 87.18% | 77.27% | 87.27% | 81.97% |



(a) (b) (c) (d)

Figure 10. Predicted bounding box from real and modified data with pre-trained model: (a) Area 1 from real and 6 modified data; (b) Area 5 from real and 6 modified data; (c) Area 1 from real and 12 modified data; and (d) Area 5 from real and 12 modified data.

## 5 Conclusion and Future Work

This research focuses on the application of 3D point clouds for on-site detection of construction materials. The main challenge was the limited actual data and occluded point cloud generated from MLS. This paper addresses the gap by proposing a material-detection method based on transfer learning and modified datasets.

A case study was conducted to verify the proposed method. Two datasets were used: the real dataset and the modified dataset. From the case study results, 3D point cloud object detection improves from 35.40% F1 score on the real dataset without a pre-trained model to 80.56% with a pre-trained model. The results demonstrate that transfer learning combined with dataset preparation can substantially improve pipe detection performance when only a limited amount of real MLS data is available. This improvement is largely due to the use of pre-trained models trained on autonomous driving datasets, which provide strong geometric priors due to their relatively similar scanning conditions. Additionally, the introduction of modified datasets further enhances the F1 score to 87.18%. This result suggests that using additional modified datasets can improve the model's performance.

Although the method shows promising results for material object detection, this study has several limitations. First, the number of testing areas is limited. To comprehensively assess the model's accuracy, additional areas, including those without pipes, should be tested. Second, the detected material is restricted to a single type and size. Future research should explore different material types to evaluate the algorithm's performance across a wider range of conditions. Third, the parameters used in this study, such as subsampling threshold, model hyperparameters, and other preprocessing settings, were not fully optimized. Only a single configuration was tested, and future work should explore the sensitivity of the method to different parameter settings to better understand their impact on detection performance. Additionally, the data used in this study is limited to 3D point clouds from MLS only. Further research could incorporate other data sources, such as images, videos, and 3D point clouds from airborne laser scanners, to assess whether integrating multiple data types improves detection accuracy.

## Acknowledgement

The authors are grateful to BPPT (Center for Higher Education Funding), Ministry of Education, Culture, Research, and Technology, Republic of Indonesia, LPDP (Indonesia Endowment Fund for Education) [Grant No. 1253/J5/KM.01.00/2021]. This research is also supported by funding from the Australian Research Council (ARC) Research Hub for Resilient and Intelligent Infrastructure Systems (IH210100048), and the ARC Research Hub for Human-Robot Teaming for Sustainable and Resilient Construction (IH240100016). The laser-scanning data provided by Linke & Linke Surveys is greatly appreciated.

## References

[1] S. Kang, M. W. Park, and W. Suh, "Feasibility study of the unmanned-aerial-vehicle radio-frequency identification system for localizing construction materials on large-scale open sites," *Sensors Mater.*, vol. 31, no. 5, pp. 1449–1465, 2019, doi: 10.18494/SAM.2019.2266.

[2] A. Alarifi *et al.*, "Ultra wideband indoor positioning technologies: Analysis and recent advances," *Sensors (Switzerland)*, vol. 16, no. 5, pp. 1–36, 2016, doi: 10.3390/s16050707.

[3] B. Yang, Y. Zang, Z. Dong, and R. Huang, "An automated method to register airborne and terrestrial laser scanning point clouds," *ISPRS J. Photogramm. Remote Sens.*, vol. 109, pp. 62–76, 2015, doi: 10.1016/j.isprsjprs.2015.08.006.

[4] S. Shi *et al.*, "PV-RCNN: Point-voxel feature set abstraction for 3D object detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2020, pp. 10526–10535. doi: 10.1109/CVPR42600.2020.01054.

[5] A. Dimitrov and M. Golparvar-Fard, "Vision-based material recognition for automated monitoring of construction progress and generating building information modeling from unordered site image collections," *Adv. Eng. Informatics*, vol. 28, no. 1, pp. 37–49, 2014, doi: 10.1016/j.aei.2013.11.002.

[6] J. Shu, X. Zhang, W. Li, Z. Zeng, H. Zhang, and Y. Duan, "Point cloud and machine learning-based automated recognition and measurement of corrugated pipes and rebars for large precast concrete beams," *Autom. Constr.*, vol. 165, no. June, 2024, doi: 10.1016/j.autcon.2024.105493.

[7] Y. Li and J. Chen, "Computer Vision–Based Counting Model for Dense Steel Pipe on Construction Sites," *J. Constr. Eng. Manag.*, vol. 148, no. 1, pp. 1–16, 2022, doi: 10.1061/(asce)co.1943-7862.0002217.

[8] J. Chen, Z. Kira, and Y. K. Cho, "Deep Learning Approach to Point Cloud Scene Understanding for Automated Scan to 3D Reconstruction," *J. Comput. Civ. Eng.*, vol. 33, no. 4, pp. 1–10, 2019, doi: 10.1061/(asce)cp.1943-5487.0000842.

[9] J. Chen, Y. Fang, Y. K. Cho, and C. Kim, "Principal Axes Descriptor for Automated Construction-Equipment Classification from Point Clouds," *J. Comput. Civ. Eng.*, vol. 31, no. 2, p. 04016058, 2017, doi: 10.1061/(asce)cp.1943-5487.0000628.

[10] K. N. R. Putri, X. Shen, M. Aleksandrov, K. Barati, and J. Linke, "Acquiring Semantic Information of Precast Concrete Pipe Using Geometric Feature Extraction from Mobile Laser Scanning Point Cloud," in *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2024, vol. X–4, no. October, pp. 263–270.

[11] I. Armeni *et al.*, "3D Semantic Parsing of Large-Scale Indoor Spaces," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1534–1543. doi: 10.1109/CVPR.2016.170.

[12] A. Dai, A. X. Chang, M. Savva, M. Halber, T. Funkhouser, and M. Nießner, "ScanNet: Richly-annotated 3D reconstructions of indoor scenes," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017, vol. 2017-Janua, pp. 2432–2443. doi: 10.1109/CVPR.2017.261.

[13] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset. The International Journal of Robotics Research," *Int. J. Rob. Res.*, no. October, pp. 1–6, 2013.