A Mixed Reality Dataset and Benchmarking of Deep Learning Models for MEP Classification

Boan Tao[1], Jiajun Li[1] and Frédéric Bosché [1]

[1]School of Engineering, University of Edinburgh, Edinburgh, Scotland, UK

boan.tao@ed.ac.uk, Jiajun.Li@ed.ac.uk, f.bosche@ed.ac.uk,

Abstract -

This paper introduces a Mixed Reality (MR)-captured dataset and a benchmark evaluation of deep learning models for Mechanical, Electrical, and Plumbing (MEP) component classification. The dataset, collected using Microsoft HoloLens 2, comprises eight distinct building environments containing pipes and ducts, presenting unique challenges characteristic of MR-captured data, including lower point density and increased noise compared to laser scans. In addition to presenting the dataset, this study provides a comparative evaluation of nine state-of-the-art deep learning architectures, assessing their performance and computational efficiency through rigorous cross-validation. Our analysis reveals that PointNN, a non-parametric approach, achieves superior classification accuracy (90.21%) while maintaining minimal computational requirements (1.34 GFLOPs). This dual contribution of this paper aims to advance automated MEP inspection systems and provide insights for developing efficient deep learning models for construction applications.

Keywords -

Mixed Reality; Point Cloud Classification; Deep Learning; MEP Systems; Construction Inspection; Building Information Modeling;

1 Introduction

Mechanical, Electrical, and Plumbing (MEP) systems form intricate networks essential for providing services such as heating, ventilation, air conditioning (HVAC), water distribution, and waste management. The inherent complexity of these systems introduces significant challenges in their design, installation, and maintenance, which result in inefficiencies, conflicts, and costly rework during construction [1]. Recent advancements in digital technologies have offered promising solutions to address these challenges. Notably, the integration of Mixed Reality (MR) and Artificial Intelligence (AI) can contribute to improving MEP system construction, monitoring and management.

Mixed Reality (MR) devices, such as the Microsoft HoloLens 2, provide unprecedented capabilities to overlay digital information onto the physical world, enabling real-time visualisation and interaction with virtual MEP components in their actual context [2]. These visualisation capabilities can be enhanced by deep learning algorithms that efficiently process and analyse complex three-dimensional data, including point clouds representing MEP systems [3]. For example, mobile scanning solutions like Apple's RoomPlan API [4] integrate device sensors with machine learning to create parametric 3D models of interior spaces. Such systems leverage camera feeds and LiDAR sensors to automatically detect and classify structural elements like walls, windows, and openings, while simultaneously recognising room features and fixtures, providing a comprehensive digital representation of the built environment. However, while such tools exist for capturing and modelling architectural elements, there is currently no comparable automated solution for MEP system mapping. Even with high-quality laser scanning data, the automated detection and classification of MEP components remains a significant technical challenge. This is due to several factors, including the inherent complexity and diversity of MEP systems, as well as occlusions commonly encountered in densely packed service areas.

This paper presents two key contributions:

1. It introduces a novel MR-captured dataset for MEP component classification, detailing its acquisition, preprocessing, and characteristics.
2. It benchmarks multiple 3D deep learning architectures on this dataset, providing insights into their classification performance and computational efficiency.

The remainder of this paper is organised as follows: Section 2 provides a comprehensive review of existing 3D datasets in construction, challenges in MR-based data capture, and recent advances in point cloud classification methods. Section 3 presents our dataset acquisition and preparation methodology, including detailed scanning protocols, dataset statistics, and cross-validation strategies. Section 4 introduces the selected deep learning models and evaluation metrics, along with implementation specifications. In Section 5, we present a detailed analysis of model performance and computational efficiency, accompanied by insights into the relative strengths and limitations of different architectural approaches. Within this section, we also discuss the implications of our findings for practical deployment and outline future research directions.

Finally, Section 6 summarises our key contributions and their significance for advancing MR-based MEP inspection in the Architecture, Engineering, and Construction (AEC) industry.

2 Literature Review

2.1 3D Datasets in Construction

Many point cloud datasets used in research are not fully representative of the challenges faced in construction environments. For instance, the ModelNet dataset [5] is a popular benchmark for 3D shape classification and segmentation, but it consists of CAD models converted to point clouds, lacking the noise and incompleteness typical of real-world scans. Similarly, ShapeNet [6] is a large-scale dataset of 3D shapes, but it also consists of clean, complete 3D models rather than actual scanned point clouds. These datasets are valuable for developing and testing algorithms, but they don't capture the complexity, imperfections and overall challenges of real-world construction scenes.

Recognising this limitation, Uy et al. [7] introduced ScanObjectNN, a real-world point cloud dataset for object classification. ScanObjectNN includes point clouds extracted from scanned indoor scenes, preserving real-world characteristics like background, occlusion, and missing parts. While more realistic than synthetic datasets, ScanObjectNN focuses on common indoor objects (e.g., chairs, tables, beds, cabinets, displays) rather than construction-specific or MEP components.

For construction-specific applications, several datasets have been developed to address the unique challenges of the built environment. The S3DIS dataset [8] provides large-scale, real-world point clouds of indoor spaces, including structural elements and some building components (e.g., doors, windows, columns, beams). The Paris-Lille-3D dataset [9] offers large-scale outdoor point clouds of urban environments, which can be relevant for certain construction tasks.

In more direct relation to MEP components (the focus of our paper), Yin et al. [10] established a benchmark industrial LiDAR dataset containing approximately 80 million data points from four industrial scenes. Their dataset, captured using a Leica BLK360 scanner, achieved accuracies of 4 mm to 7 mm and included five categories of MEP components: pipes, pumps, tanks, I-shape, and rectangular beams. Agapaki and Brilakis [11] also introduce construction and MEP point cloud datasets: CLOI (Channels, L-shapes, circular sections, I-shapes). It is a large-scale repository of labelled point clusters specifically designed for industrial and MEP components. It contains over 140 million hand-labelled points from five industrial facilities, covering ten classes of the most frequent construction and MEP shapes. The dataset was collected using static terrestrial laser scanners, capturing the real-world noise and occlusions typical in complex industrial environments.

2.2 Challenges in MR-based Data Capture

Despite the potential of MR devices for on-site data capture, several challenges remain. Hübner et al. [12] evaluated the HoloLens (version 1) for indoor mapping applications, providing insights into its depth sensing capabilities and limitations. They found that the device's accuracy was highly dependent on distance and surface inclination, with optimal performance at distances less than 2.5 m and inclinations below 20 degrees. Beyond these limits, the sensor's accuracy deteriorated significantly, with noise exceeding 2 cm at greater distances or inclinations.

MR devices like the HoloLens produce point clouds that are significantly less dense and more prone to noise compared to traditional LiDAR scans. This is particularly problematic in MEP environments, where occlusions, reflective surfaces, and hard-to-reach installation locations can further degrade data quality.

The accuracy of MR-captured point clouds decreases with distance, which poses challenges for capturing tall or expansive spaces often encountered in industrial and commercial buildings. These limitations have implications for the applicability of existing deep learning models, which are typically trained on high-quality LiDAR datasets, as discussed in Section 2.1.

2.3 Point cloud Classification

Point cloud classification has advanced significantly with deep learning techniques. PointNet [13] pioneered direct processing of unordered point sets using MLPs but struggled with capturing local features. PointNet++ [14] improved upon this by introducing hierarchical feature learning to model local geometric structures and handle noise.

Recent works have built on this foundation with various improvements. PointNeXt [15] modernises PointNet++ with better training strategies, while PointMetaBase [16] introduces a modular meta-architecture to optimize computational efficiency. PointVector [17] enhances local feature learning through vector-based representations.

Graph-based models like DGCNN [18] apply dynamic edge convolutions to capture local relationships and adapt to structural changes, which is especially useful for noisy MR data. DeepGCNs [19] extend this concept with deeper architectures and residual connections for more complex feature hierarchies.

Attention mechanisms, such as those used in Point Transformer [20], help models focus on the most informative regions of sparse or noisy point clouds, improving robustness.

Further, PointMLP [21] simplifies the architecture using residual MLPs while achieving strong performance on real-world datasets like ScanObjectNN, making it suitable for MR data.

Finally, Point-NN [22] introduces a non-parametric approach using simple operations (FPS, k-NN, pooling) with trigonometric spatial encoding. This method excels in MR environments due to its inherent robustness to varying point densities and distributions without requiring trainable parameters.

2.4 Gap in Existing Research

While there has been significant progress in both point cloud analysis and the application of MR in construction, there remains a critical gap in datasets with characteristics representative of data acquisition using typical MR-mounted sensing solutions, in particular to support applications related to MEP mapping. Existing datasets predominantly focus on high-precision LiDAR scans, which do not adequately capture the unique challenges posed by MR-based data collection. The sparse, noisy, and gap-prone nature of MR-captured meshes is not represented in these datasets. There is a need for datasets that reflect the real-world conditions encountered in MR-based inspections, including lower point density, increased noise, and more frequent occlusions due to the limitations of MR devices and the challenging nature of MEP environments.

3 Dataset Acquisition and Preparation

3.1 Research Framework

To clearly present the overall workflow of this research, we summarise the main steps, objectives, methods, and equipment involved:

- **1. MR Data Acquisition**
  Objective: Capture raw 3D spatial meshes of real-world MEP systems (pipes and ducts) under realistic MR conditions, while managing challenges such as noise and sparsity.
  Device: Microsoft HoloLens 2
  Method: On-site scanning of building environments with multi-angle coverage and careful motion control. The operator tilted the device upwards to capture ceiling-mounted MEP components, ensuring maximum possible coverage.

- **2. Dataset Preparation**
  Objective: Generate a structured and labelled MR point cloud dataset (MR-MEP) suitable for training and evaluating deep learning models.
  Process: Convert raw spatial meshes to point clouds, clean the data (removing large architectural elements and noise), segment MEP components using BIM-derived Oriented Bounding Boxes (OBBs), label points as ‘pipe’ or ‘duct’, and apply data augmentation to balance class distribution.

- **3. Model Benchmarking**
  Objective: Assess and compare the classification performance and computational demands of state-of-the-art models on MR-captured MEP data.
  Models: Nine deep learning architectures, including PointNet, PointNet++, DGCNN, PointMLP, PointNeXt, PointTransformer, PointMetaBase, PointVector, and PointNN.
  Method: Perform 6-fold cross-validation with scene-based splits, applying pre-training on ScanObjectNN where applicable and balancing training data with augmented samples.
  Evaluation Metrics: Overall Accuracy (OA), Mean Class Accuracy (mAcc), class-specific accuracy (pipes and ducts), and computational efficiency (parameters, FLOPs, throughput).

- **4. Performance Analysis**
  Objective: Determine the optimal model for practical MR-based MEP classification, balancing high accuracy with computational efficiency for real-world deployment, including potential on-device processing.
  Process: Analyse cross-validation results, evaluate accuracy and stability, and compare computational efficiency. Identify models most suitable for MR data constraints, such as sparsity and noise.

3.2 Scanning Procedure and Protocols

The Microsoft HoloLens 2 MR headset was selected as the primary data collection device for this study, owing to its sophisticated sensing and processing capabilities [23]. This advanced device incorporates a comprehensive sensor array specifically designed for spatial mapping. Hübner et al. [12] conducted an extensive evaluation of the HoloLens 2 for indoor mapping tasks, revealing a correlation between sensor-to-object distance and measurement accuracy. Their findings indicate that as the distance between the sensor and the object increases, the associated measurement noise also tends to rise. Specifically, for distances under 2.5 m, the noise level remains below 5 mm. However, beyond this threshold, the error exhibits a rapid escalation, with noise levels reaching 2 cm or more at greater distances.

In our study reported here, the scene mapping functionality of the HoloLens 2 (HL2) was utilised to capture the environment as a triangle mesh. The mesh resolution was set to the finest level, and a comprehensive scan of the environment was conducted. Given that the HL2’s capture sensors are front-facing, it was necessary to tilt the device upwards to scan certain areas. Additionally, since some Mechanical, Electrical, and Plumbing (MEP) components were located on the ceiling, which exceeded a height of 2.5 m, at times the device had to be handheld at arm’s length above the operator’s head, with both hands gripping the sides of the headset for stability. The camera was oriented upward to directly face the MEP components, and the operator maintained a steady, controlled motion with movement speed kept below approximately 0.3 meters per second to ensure complete coverage.

Data acquisition was conducted across 8 distinct building environments to ensure a diverse and representative dataset. A standardised scanning protocol was implemented to maintain consistency and data quality across all sites. This protocol comprised a preliminary site walkthrough to identify and delineate key MEP areas of interest, followed by systematic and comprehensive scanning of the identified MEP areas. Particular attention was given to capturing each component thoroughly from multiple angles to optimise coverage. The duration of each scanning session was approximately 15-20 minutes, with variations depending on the complexity and scale of the individual construction site.

3.3 Dataset Overview and Statistics

The dataset primarily focuses on MEP components like pipes and ducts, featuring both mixed reality captured environments and their corresponding BIM models. Data were collected from eight distinct scenes, six of which are classified as mixed environments containing both pipes and ducts: bike parking, apartment room, Garage 1, Garage 2, Garage 5, and Garage 6. The remaining two scenes, Garage 3 and Garage 4, are pipe-only environments.

To establish the ground truth, BIM was employed to generate precise digital representations from laser scan data acquisition of those environments.

3.4 Dataset Preparation for Point Cloud Classification Use Case

In the context of scan-vs-BIM comparison for automated construction progress monitoring, we developed a component-wise detection strategy utilising oriented bounding boxes (OBBs) derived from the corresponding BIM elements. Similarly to [24, 25], the idea is to explore whether the point cloud in the vicinity of each BIM object confirms the presence of that element. This systematic decomposition approach resulted in 382 distinct OBB regions across all environments, where each OBB region encompasses point cloud data that potentially corresponds to a specific MEP component segment.

Table 1 presents the key characteristics of the eight scenes in our dataset, detailing the number of vertices, segmented pipes, segmented ducts, and the dimensions of OBB enclosing each MEP component in each scene.

3.5 Class Balancing Strategy

Analysis of the dataset composition reveals a significant class imbalance between ducts and pipes. Of the total 382 elements, 44 (11.5%) represent ducts, while the remaining 338 (88.5%) correspond to pipes. This imbalance is arguably representative of the reality of building environments [26], where piping networks often dominate due to their essential role in various building services including water supply, drainage, heating, cooling, and fire protection systems.

To mitigate the issue of class imbalance, a data augmentation pipeline consisting of rotation, scaling, and mirroring was applied to the duct class within the mixed scenes. This augmentation process was repeated until a 1:1 ratio between duct and pipe samples was achieved, ensuring a balanced dataset for model training.

3.6 Cross-Validation Strategy

A 6-fold cross-validation strategy was employed to robustly evaluate the performance of the selected and deployed classification models. Each fold ensures that every mixed scene serves as the test set once, while the model is evaluated on a combination of pipes and ducts. During training, the model is exposed to both mixed scenes and pipe-only scenes.

The cross-validation folds were structured such that in each iteration, one scene was held out for testing while all others were used for training. The test scenes for the six folds were: Bike Parking, Apartment Room, Garage 1, Garage 2, Garage 5, and Garage 6, respectively.

4 Methods and Implementation

4.1 Selected Deep Learning Models

We select a diverse set of classification models, including traditional deep learning architectures, graph-based methods, transformer-inspired models, and a non-parametric method. This range enables an assessment of trade-offs between model complexity, computational efficiency, and performance, particularly for point cloud processing tasks involving noisy or sparse data captured by MR devices. These models have been pre-trained on ScanObjectNN, which closely resembles Mixed Reality captured data as discussed in Section section 2.1, making them particularly suitable for real-world MR applications. The selected architectures include:

- **PointNet** [13] serves as a fundamental architecture for point cloud classification and segmentation tasks.
- **PointNet++** [14] builds on PointNet and introduces hierarchical feature learning through Set Abstraction modules, enhancing the model's capability to capture local structures.
- **DGCNN** [18] takes a graph-based approach, implementing dynamic edge convolution to address noise and sparsity challenges by effectively leveraging local neighborhood structures.
- **PointMLP** [21] advances the field by implementing a sophisticated residual MLP framework designed for robust performance.

Scene	Vertices	Pipes	Ducts	Size (m^2)
Bike Parking	32,962	51	3	10.4 × 8.6
Apartment Room	82,139	96	11	12.5 × 10.8
Garage 1	95,416	24	7	18.6 × 14.8
Garage 2	85,814	40	4	20.9 × 10.9
Garage 3	47,177	67	0	12.2 × 4.3
Garage 4	14,029	24	0	13.2 × 8.0
Garage 5	35,967	3	5	12.7 × 3.9
Garage 6	39,447	33	14	30.5 × 15.9
Total	432,951	338	44	131.0 × 77.2

Table 1. Scene-level characteristics of the captured environments

- **PointNeXt** [15] revisits the PointNet++ architecture, incorporating enhanced training strategies and introducing a more scalable model design.
- **PointMetaBase** [16] offers a unified analysis framework that enables efficient point cloud processing.
- **PointVector** [17], which employs vector-based processing techniques to capture spatial relationships within point clouds;
- **PointTransformer** [20], which adapts transformer principles to emphasise spatial relationships in point cloud data;
- **Point-NN** [22] takes a unique approach as a non-parametric, training-free method that utilises a point-memory bank for similarity matching.

4.2 Evaluation Metrics

We employ a diverse set of metrics across six-fold validation to rigorously assess model performance and efficiency.

The evaluation of classification accuracy encompasses three distinct metrics:

- **Class-specific Accuracy (Acc)** quantifying the classification accuracy for pipe and ducts independently, enabling granular performance assessment and identification of potential classification biases.
- **Overall Accuracy (OA)** quantifies the proportion of correctly classified samples across the entire dataset, providing a global performance indicator.
- **Mean Class Accuracy (mAcc)** averages individual class accuracies, thereby ensuring equitable evaluation across all categories, regardless of class imbalance issues.

The evaluation of computational efficiency also comprises three fundamental metrics:

- **M, the total number of trainable parameters** quantifying the model complexity and expressed in millions.
- **G, the FLOPs** quantifying the computational demand during inference and expressed in billions of floating-point operations.
- **Throughput measurements** quantifying the system's processing speed by measuring the number of complete inference requests the model can process per second under realistic deployment conditions.

For statistical robustness, we report the mean and standard deviation of each accuracy and efficiency metrics across all six folds, thereby providing comprehensive insights into both performance magnitude and consistency.

4.3 Hardware and Software Specifications

The experiments were conducted on a system equipped with an NVIDIA GeForce RTX 3090 GPU and an AMD Ryzen 9 3900X 12-core processor. The software environment included Ubuntu 20.04.6 LTS as the operating system. The deep learning framework used was PyTorch 2.0.0, which was integrated with CUDA 11.8 for GPU acceleration. Python version 3.9.0 was used as the programming language to run the experiments.

5 Results and Discussion

The experimental results are presented in table 2 and table 3 that provide a comprehensive evaluation of the selected models across both accuracy and computational efficiency metrics. Table 2 summarises the cross-validation accruacy results.Among all models, PointNN demonstrates clear superior performance, achieving by far the highest scores across all accuracy metrics (90.21% overall accuracy), and with lower standard deviations, indicating more stable performance across different test scenes. Table 3 presents the computational efficiency metrics.The results show that PointNN achieves optimal efficiency with minimal parameters (~0.0M[1]) and highest processing speed (11,932.60 instances per second), while PointMetaBase requires the least computational operations at 0.55G FLOPs, demonstrating its theoretical computational efficiency.

To visualise the trade-off between model performance and computational demands, fig. 1 plots the Overall Accuracy against FLOPs for all evaluated models. The figure clearly demonstrates that PointNN (marked with a star) achieves an optimal balance, reaching the highest accuracy

[1] because it is a non-parametric method

Model	Overall Accuracy (%)	Mean Class Accuracy (%)	Pipe (%)	Duct (%)
PointNN	**90.21 ± 2.12**	**89.45 ± 2.34**	**91.33 ± 2.45**	**85.67 ± 3.56**
PointMLP	81.55 ± 8.23	74.00 ± 12.45	67.76 ± 37.91	80.23 ± 21.54
PointNet++	80.85 ± 13.21	78.22 ± 13.45	77.42 ± 23.45	79.02 ± 24.12
PointTransformer	80.52 ± 13.45	77.49 ± 14.56	89.20 ± 11.23	75.78 ± 17.89
PointNext	80.45 ± 14.23	75.37 ± 17.34	80.53 ± 34.56	70.22 ± 31.23
DGCNN	78.57 ± 7.89	72.36 ± 9.21	70.44 ± 31.82	74.28 ± 25.85
PointNet	78.43 ± 8.56	70.24 ± 9.78	62.80 ± 33.45	77.68 ± 24.56
PointMetaBase	78.35 ± 13.65	74.79 ± 11.83	81.80 ± 18.32	67.78 ± 32.41
PointVector	68.79 ± 15.67	63.79 ± 12.45	71.85 ± 27.89	55.74 ± 37.89

Table 2. Cross-validation Results (Mean ± Std)

Model	Parameters (M)	FLOPs (G)	Throughput (ins./sec)
PointNN	**~0.0**	1.34	**11,932.60**
PointMetaBase	1.417	**0.55**	2,313.80
PointNext	1.364	1.64	2,776.64
PointNet++	1.463	1.71	2,373.52
PointVector	1.544	1.68	1,069.91
DGCNN	1.800	4.82	802.59
PointNet	3.470	0.91	5,620.63
PointMLP	13.228	31.35	199.37
PointTransforme	22.647	29.16	432.74

Table 3. Computational Efficiency Metrics

while maintaining moderate computational complexity. Most deep learning models cluster in the middle range of both metrics, with FLOPs between 0.5G and 5G and accuracies between 75% and 85%. Notably, more complex models like PointMLP and PointTransformer, despite their substantially higher computational demands (>25G FLOPs), do not yield proportional improvements in accuracy.

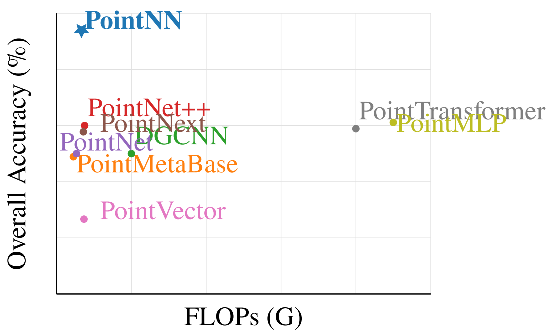

Figure 1. Model accuracy versus computational complexity (FLOPs). PointNN (marked with a star) achieves the best accuracy while maintaining minimal computational requirements.

5.1 Performance Analysis

PointNN achieves superior performance across all accuracy metrics, with an overall accuracy of 90.21% (±2.12%) and mean class accuracy of 89.45% (±2.34%). PointNN's consistently low standard deviations (±2.12% to ±3.56% across different metrics) strongly suggest superior generalization capability and resistance to overfitting, particularly valuable for MR applications where data distributions can vary significantly between deployments. Most notably, it maintains remarkably consistent performance across both Pipe (91.33% ±2.45%) and Duct (85.67% ±3.56%) classes, indicating stable learning across different geometric features.

The second tier of performers exhibits concerning signs of potential overfitting, as evidenced by their high cross-validation standard deviations. PointMLP (±8.23%), PointNet++ (±13.21%), PointTransformer (±13.45%), and PointNext (±14.23%) show comparable mean accuracy but their high variances across different folds suggest that these models may be overspecialising to training set characteristics rather than learning generalisable features, with their class-specific performance showing dramatic variations (up to ±37.91% for PointMLP's pipe classification) indicating inconsistent feature learning.

More traditional architectures like DGCNN (78.57% ±7.89%) and the original PointNet (78.43% ±8.56%) demonstrate moderate performance with relatively lower cross-validation standard deviations compared to the second tier. While their balanced standard deviations suggest more stable learning, their lower overall accuracy makes them less suitable for precision-critical MR applications. These models' moderate standard deviations indicate better generalisation than more complex models but still fall short of PointNN's stability.

From a computational perspective, PointNN's performance is particularly noteworthy for MR applications, given its minimal parameter count (~0.0M) and high throughput (11,932.60 instances/second). The absence of learnable parameters inherently prevents traditional overfitting, as reflected in the consistent cross-validation results. This stands in stark contrast to more complex models like Point-Transformer and PointMLP, which require substantially more parameters (22.647M and 13.228M respectively) and show signs of overfitting in their high standard deviations despite achieving reasonable mean accuracy.

Overall, the non-parametric approach of PointNN provides several key advantages for MR point cloud processing: First, the trigonometric encoding effectively captures high-frequency geometric patterns, which remain stable even when surface details are inconsistently captured by MR devices. Second, the point-memory bank's similarity-based classification provides flexible matching rather than rigid decision boundaries, crucial for handling varying point densities and object appearances. Third, PointNN shows remarkable immunity to common MR data challenges including partial occlusions and sensor noise through its local geometry aggregation mechanism. Combined with its computational efficiency and inherent regularization from the non-parametric design, these characteristics make PointNN especially valuable for real-world MR deployments where robustness, consistency, and real-time processing are critical requirements.

6 Conclusion

This work presented a new dataset MR-MEP and an investigation into the performance of point cloud classification deep learning models on that type of data. Specifically, we present MR-MEP, a specialised dataset that pairs mixed reality captured point clouds with corresponding BIM models for MEP components in building environments. The dataset was collected using Microsoft HL2 across 8 diverse building scenes, containing $432,951$ vertices spanning a total area of 131.0×77.2 m^2. What distinguishes MR-MEP is its dual-source nature - combining precisely captured mixed reality point cloud data with corresponding BIM models - making it valuable for multiple point cloud understanding tasks including classification, object detection, and semantic/instance segmentation. For our specific investigation into MEP component classification, the dataset is split into 382 annotated MEP components (338 pipes and 44 ducts), segmented using BIM-derived oriented bounding boxes (OBBs). To ensure robust evaluation, we implemented a 6-fold cross-validation strategy and addressed the natural class imbalance through careful data augmentation.

Our experimental analysis of the different point cloud classification models demonstrates that the recently proposed non-parametric approach, PointNN, achieves state-of-the-art performance with 90.21% classification accuracy while maintaining minimal computational overhead of 1.34 GFLOPs. The model's ability to maintain high accuracy despite the quality limitations of MR-captured data suggests potential for robust deployment in real-world construction scenarios. Moreover, the minimal computational requirements make it particularly suitable for on-device processing in MR devices, addressing a critical need in mobile construction inspection applications.

Although the proposed approach demonstrates promising results in MEP component classification from MR-acquired point cloud data, several critical areas merit further investigation. The current implementation, while effective, has been primarily evaluated on a binary classification task distinguishing between pipes and ducts. This should be extended to support multi-class classification capabilities, in order to accommodate the diverse range of MEP components encountered in real-world construction environments. Additionally, while our work focuses on classification, further evaluation across other crucial point cloud understanding tasks is needed. For instance, in real construction scenarios, MEP components need to be precisely localized through object detection and semantically segmented for detailed analysis. For such tasks, while Point-NN's non-parametric approach demonstrates strong performance in classification, its effectiveness in more complex scenarios like instance or semantic segmentation remains to be thoroughly investigated. It is possible that other architectures with learned parameters (including the *Point-PN* approach proposed by the same authors [22]) might prove more suitable for tasks requiring fine-grained spatial understanding or complex feature hierarchies, as these tasks may benefit from learned, task-specific feature representations. To bridge the gap between theoretical advances and practical implementation, future work should focus on integrating deep learning-based detection and classification models like those into comprehensive MEP construction monitoring frameworks such as the one proposed by [24]. Such integration presents several technical challenges that warrant investigation, including: the optimisation of real-time processing capabilities within MR-based progress monitoring systems, and the development of robust verification mechanisms for construction geometry inspection.

References

[1] Jun Wang, Xiangyu Wang, Wenchi Shou, Heap-Yih Chong, and Jun Guo. Building information modeling-based integration of mep layout designs and constructability. *Automation in Construction*, 61:134–146, 2016. ISSN 0926-5805. doi:10.1016/j.autcon.2015.10.003.

[2] Ahlem Assila, Amira Dhouib, Ziad Monla, and Mourad Zghal. Integration of augmented, virtual and mixed reality with building information modeling: A systematic review. In Jessie Y. C. Chen and Gino Fragomeni, editors, *Virtual, Augmented and Mixed Reality: Design and Development*, pages 3–19, Cham, 2022. Springer International Publishing.

ISBN 978-3-031-05939-1. doi:10.1007/978-3-031-05939-1_1.

[3] Hongzhe Yue, Qian Wang, Hongxiang Zhao, Ningshuang Zeng, and Yi Tan. Deep learning applications for point clouds in the construction industry. *Automation in Construction*, 168:105769, 2024. ISSN 0926-5805. doi:10.1016/j.autcon.2024.105769.

[4] Apple Inc. *RoomPlan*, 2024. URL `https://developer.apple.com/documentation/roomplan`. Accessed: 2024-11-09.

[5] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3D ShapeNets: A deep representation for volumetric shapes . In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1912–1920, Los Alamitos, CA, USA, June 2015. IEEE Computer Society. doi:10.1109/CVPR.2015.7298801.

[6] Angel X. Chang, Thomas A. Funkhouser, Leonidas J. Guibas, Pat Hanrahan, Qi-Xing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, Jianxiong Xiao, L. Yi, and Fisher Yu. Shapenet: An information-rich 3d model repository. *ArXiv*, abs/1512.03012, 2015.

[7] Mikaela Angelina Uy, Quang-Hieu Pham, Binh-Son Hua, Thanh Nguyen, and Sai-Kit Yeung. Revisiting point cloud classification: A new benchmark dataset and classification model on real-world data. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 1588–1597, 2019. doi:10.1109/ICCV.2019.00167.

[8] Iro Armeni, Ozan Sener, Amir R. Zamir, Helen Jiang, Ioannis Brilakis, Martin Fischer, and Silvio Savarese. 3d semantic parsing of large-scale indoor spaces. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1534–1543, 2016. doi:10.1109/CVPR.2016.170.

[9] Xavier Roynard, Jean-Emmanuel Deschaud, and François Goulette. Paris-lille-3d: A large and high-quality ground-truth urban point cloud dataset for automatic segmentation and classification. *Int. J. Rob. Res.*, 37(6):545–557, May 2018. ISSN 0278-3649. doi:10.1177/0278364918767506.

[10] Chao Yin, Boyu Wang, Vincent J.L. Gan, Mingzhu Wang, and Jack C.P. Cheng. Automated semantic segmentation of industrial point clouds using respointnet++. *Automation in Construction*, 130:103874, 2021. ISSN 0926-5805. doi:https://doi.org/10.1016/j.autcon.2021.103874.

[11] Eva Agapaki and Ioannis Brilakis. Cloi-net: Class segmentation of industrial facilities' point cloud datasets. *Advanced Engineering Informatics*, 45:101121, 2020. ISSN 1474-0346. doi:https://doi.org/10.1016/j.aei.2020.101121.

[12] Patrick Hübner, Kate Clintworth, Qingyi Liu, Martin Weinmann, and Sven Wursthorn. Evaluation of hololens tracking and depth sensing for indoor mapping applications. *Sensors*, 20(4):1021, 2020. doi:10.3390/s20041021.

[13] R. Qi Charles, Hao Su, Mo Kaichun, and Leonidas J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 77–85, 2017. doi:10.1109/CVPR.2017.16.

[14] Charles R. Qi, Li Yi, Hao Su, and Leonidas J. Guibas. Pointnet++: deep hierarchical feature learning on point sets in a metric space. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 5105–5114, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964.

[15] Guocheng Qian, Yuchen Li, Houwen Peng, Jinjie Mai, Hasan Abed Al Kader Hammoud, Mohamed Elhoseiny, and Bernard Ghanem. Pointnext: revisiting pointnet++ with improved training and scaling strategies. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS '22, Red Hook, NY, USA, 2024. Curran Associates Inc. ISBN 9781713871088. doi:10.5555/3600270.3601955.

[16] Haojia Lin, Xiawu Zheng, Lijiang Li, Fei Chao, Shanshan Wang, Yan Wang, Yonghong Tian, and Rongrong Ji. Meta Architecture for Point Cloud Analysis . In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 17682–17691, Los Alamitos, CA, USA, 2023. IEEE Computer Society. doi:10.1109/CVPR52729.2023.01696.

[17] Xin Deng, WenYu Zhang, Qing Ding, and XinMing Zhang. Pointvector: A vector representation in point cloud analysis. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9455–9465, 2023. doi:10.1109/CVPR52729.2023.00912.

[18] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E. Sarma, Michael M. Bronstein, and Justin M. Solomon. Dynamic graph cnn for learning on point clouds. *ACM Trans. Graph.*, 38(5), October 2019. ISSN 0730-0301. doi:10.1145/3326362.

[19] Guohao Li, Matthias Muller, Ali Thabet, and Bernard Ghanem. DeepGCNs: Can GCNs Go As Deep As CNNs? . In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9266–9275, Los Alamitos, CA, USA, November 2019. IEEE Computer Society. doi:10.1109/ICCV.2019.00936.

[20] Hengshuang Zhao, Li Jiang, Jiaya Jia, Philip Torr, and Vladlen Koltun. Point transformer. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 16239–16248, 2021. doi:10.1109/ICCV48922.2021.01595.

[21] Xu Ma, Can Qin, Haoxuan You, Haoxi Ran, and Yun Fu. Rethinking network design and local geometry in point cloud: A simple residual mlp framework, 2022.

[22] Renrui Zhang, Liuhui Wang, Yali Wang, Peng Gao, Hongsheng Li, and Jianbo Shi. Starting from non-parametric networks for 3d point cloud analysis. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5344–5353, 06 2023. doi:10.1109/CVPR52729.2023.00517.

[23] Dorin Ungureanu, Federica Bogo, Silvano Galliani, Pooja Sama, Xin Duan, Casey Meekhof, Jan Stühmer, Thomas J. Cashman, Bugra Tekin, Johannes L. Schönberger, Pawel Olszta, and Marc Pollefeys. Hololens 2 research mode as a tool for computer vision research. *CoRR*, abs/2008.11239, 2020. doi:10.48550/arXiv.2008.11239.

[24] Boan Tao, Frédéric Bosché, and Jiajun Li. Mixed reality-based mep construction progress monitoring: Evaluation of methods for mesh-to-mesh comparison. *Automation in Construction*, 168:105852, 2024. ISSN 0926-5805. doi:10.1016/j.autcon.2024.105852.

[25] Zhiqi Hu and Ioannis Brilakis. Matching design-intent planar, curved, and linear structural instances in point clouds. *Automation in Construction*, 158:105219, 2024. ISSN 0926-5805. doi:https://doi.org/10.1016/j.autcon.2023.105219.

[26] Viktor Drobnyi, Zhiqi Hu, Yasmin Fathy, and Ioannis Brilakis. Construction and maintenance of building geometric digital twins: State of the art review. *Sensors*, 23 (9):4382, 2023. doi:10.3390/s23094382.