NUS3D: A BIM-Synthetic, Mobile-Scan and TLS Point Cloud Dataset for Robotic 3D Scene Understanding in Built Environment

Vincent J.L. Gan[1], Mingkai Li[1], Jingxuan Li[1], Chao Yin[2], Boyu Wang[3]

[1] National University of Singapore, Singapore
[2] Hong Kong University of Science and Technology, HK
[3] New York University (Abu Dhabi), UAE

Abstract –
Representation learning for 3D scenes in the architecture, engineering and construction domains is becoming important, as it facilitates Scan-to-BIM, 3D perception and robotic deployment in built environments. This study presents NUS3D, a dataset that consolidates BIM-synthetic, mobile-scan and terrestrial laser scan (TLS) point clouds for building interiors. Originated in 2022, we attempted to integrate heterogeneous data sources to capture the spectrum of 3D data characteristics including structural and M&E (mechanical and electrical) components. BIM-synthetic point clouds provide noise-free geometry that serves as a design-intent reference, mobile-scan contributes dynamic observations collected in real-world conditions, TLS offers high-precision and dense measurements of as-built scenes. Such mixed-source composition enables training and evaluation of algorithms for representation learning, allowing models to acquire generalizable spatial and geometric features that are transferable across synthetic and real, static and dynamic, as well as design and as-built domains. This paper includes a semantic segmentation test using several BIM and TLS M&E point clouds. Experiment demonstrates the potential of NUS3D for advancing future research on 3D scene understanding, Scan-to-BIM, and robotic perception in the built environment.

Keywords –
Point Cloud; 3D Scene Understanding; Scan-to-BIM; Robotic Perception.

1 Introduction

Robots are increasingly integrated into built environments [13], yet autonomous perception in these settings remains highly challenging. Such environments range from unstructured construction sites, where layouts change over time, to semi-structured spaces with known geometry but movable elements like furniture, equipment and human activity present. Unlike controlled conditions, these spaces are in constant flux such as clutter, moving obstacles, and human presence alter scene geometry, complicating environment perception.

Point cloud semantic segmentation serves as a key enabler for 3D scene understanding. In practice, 3D point clouds can be obtained through passive reconstruction or active sensing. Passive reconstruction typically relies on cameras to infer geometric information. This includes monocular approaches that estimate depth from a single image, often supported by deep learning techniques [7]. Stereo or multi-view approaches exploit disparity or multi-view geometry (structure-from-motion, multi-view stereo) to generate point clouds [25]. In contrast, active sensing relies on external sensors to directly acquire geometry of the object. For ground-based measurements, LiDAR (Light Detection and Ranging) is more capable of producing dense point clouds in large-scale environments. Among these systems, terrestrial laser scanners (TLS) are particularly effective, offering superior accuracy and measurement range. However, TLS does not support real-time point loud registration that are usually required in scanning complex, occluded environments where the sensor should move around the built environment [13].

As such, researchers are looking into mobile scanning (MLS) in which LiDAR are leveraged with robot to improve the mobility of 3D scanning [13,28]. In this regard, Simultaneous Localization and Mapping (SLAM) has gained increasing attention. SLAM enables robots to determine their position within an environment while simultaneously generating a 3D map of that environment. It allows real-time tracking of the scanning system while reconstructing the 3D scene [10,14]. This approach enhances the efficiency of mobile LiDAR mapping and point cloud acquisition, particularly in indoor building and construction environments where cluttered and occluded spaces are commonly encountered.

In recent years, several 3D point cloud datasets have been released (see **Table 1**). Notably, ModelNet [26] is one of the earliest collections, comprising CAD models primarily intended for 3D object classification, recognition and shape analysis. Semantic3D [12] provides a benchmark for outdoor semantic segmentation, featuring extensive terrestrial laser scans of streets, buildings, vegetation, and other urban structures. For autonomous driving and navigation, KITTI [11] remains a widely used dataset, containing road scenes captured with LiDAR and stereo cameras for benchmarking 3D detection and SLAM. Extending this effort, nuScenes [4] incorporates a rich set of LiDAR, camera, and radar recordings across urban traffic environments.

The TUM RGB-D dataset [23] focuses on visual SLAM evaluation, offering indoor RGB-D sequences with ground truth trajectories, and has become a de facto standard for benchmarking SLAM. Meanwhile, ScanNet [6] introduces richly annotated RGB-D indoor scans, including reconstructed meshes and semantic labels, which have since become a foundation for indoor scene parsing tasks. Similarly, the Stanford Large-Scale 3D Indoor Spaces (S3DIS) dataset [2] offers labeled point clouds of complex building interiors. Both the ScanNet and S3DIS were employed in Scan-to-BIM studies.

Table 1. Representative 3D datasets used in computer vision, robotics and built environment research

Dataset		Description	Reference
ModelNet	CAD (indoor objects)	CAD models for object classification, recognition, and shape analysis.	[26]
Semantic 3D	Outdoor urban scenes	Large-scale outdoor laser point cloud dataset including streets, buildings, churches, and vegetation.	[12]
KITTI	Urban scenes	Classic dataset for autonomous driving with LiDAR, GPS/IMU for 3D detection.	[11]
nuScenes	Urban scenes	Dataset covering roads, including LiDAR, cameras, and radars.	[4]
TUM RGB-D SLAM	Indoor SLAM	RGB-D sequences with ground-truth trajectories.	[23]
ScanNet	Indoor scenes (RGB-D scans)	Real indoor scene with RGB-D sequences, reconstructed meshes, and annotations.	[6]
S3DIS	Indoor large building spaces	Point clouds of large-scale indoor buildings with semantic labels.	[2]

Despite the breadth of existing resources, several limitations remain when applying these datasets. Built environments, particularly indoor spaces and building-street interfaces with indoor-outdoor transitions, are characterized by clutter, occlusions, narrow passages, mixed lighting, and weakly textured surfaces, posing substantial challenges that differ from existing datasets. Particularly, such environments often contain dense mechanical, electrical and plumbing components and spatial relationships with structural components.

Figure 1. Prospect for built environment 3D dataset

Another limitation lies in the method used to generate these datasets. The type and quality of point cloud data available from passive and active sensing, as well as static or dynamic mobile approaches involve different trade-offs in terms of accuracy, completeness, noise and operational efficiency. Looking ahead, built environment datasets should include multiple data sources such as BIM-synthetic point clouds, which represent noise-free design intent; mobile scans that capture dynamic observations under real-world conditions; and TLS data, which offer dense, high-precision ground truth. This mixed-source composition supports robust training and evaluation of 3D representation learning models with improved generalization. Specifically, there is growing momentum toward robotic data acquisition (see **Figure 1**). Integrating mobile robotic LiDAR mapping with

BIM-synthetic and TLS data enables a holistic dataset for BIM-to-Scan and Scan-to-BIM verification, supporting reliable structural semantic segmentation, surface-level feature extraction, and enhances 3D scene understanding.

This paper presents our ongoing efforts to develop NUS3D, a hybrid dataset that consolidates BIM-synthetic, mobile-scan, and high-precision TLS point clouds to capture the diverse characteristics of 3D data in built environments, particularly with indoor and indoor-outdoor transitional spaces that are characterized by clutter, occlusions, semi-structured layouts and weakly textured surfaces. This composition supports the development and evaluation of geometric representation learning pipelines, enabling them to acquire generalizable geometric features transferable across synthetic and real data as well as design-intent and as-built domains. Such capabilities are expected to advance scene understanding for environment perception, robotic deployment and Scan-to-BIM research.

2 Related Work

2.1 Data Acquisition for Indoor Environments

In cluttered changing environment, conventional static terrestrial laser scanning remains accurate but it is often time-consuming and limited in flexibility for large or continuously evolving spaces. There is a growing deployment of mobile robots in clutter, unstructured environments, such as scanning building geometry with UAV-UGV [14] and automated mobile robot navigation to produce survey-quality 3D point clouds of construction sites.

Ibrahim, Golparvar-Fard and El-Rayes [17] introduced metrics and methods for assessing the quality of reality capture, particularly in scenarios where unmanned ground rovers are deployed for construction monitoring and inspection. Cheng et al. [5] proposed an improved coverage planning that optimizes waypoint ordering and sequences for robotic data acquisition. Such integration opens the possibility of quality-assurance-driven scanning, where robots adapt their movement trajectories to optimize coverage and improve point cloud quality.

2.2 Point Cloud Quality Assessment

Research has been undertaken to explore how errors affect scene understanding, including those arising from SLAM drift, LiDAR occlusions, and inaccuracies in BIM models. Esfahani et al. [9] provide groundwork by conducting a quantitative investigation into how levels of automation and modeler training affect Scan-to-BIM accuracy, offering predictive models of geometric uncertainty under various workflows.

The establishment of point cloud quality metrics, such as density, coverage and spatial distribution, remains important for defining minimum data requirements for Scan-to-BIM outcomes. Liu et al. [20] contributes toward this goal by introducing BIMNet, an IFC-based large-scale dataset and corresponding benchmark focused on as-built BIM reconstruction from. Furthermore, the deviation analysis method proposed by Anil et al. [1] demonstrates how geometric deviations between point clouds and BIMs can be systematically analysed to locate, characterize, and diagnose errors in as-built models.

2.3 Representation Learning

Another research avenue lies in the evaluation of representation learning and scene understanding, including transformer-based architecture, contrastive learning, and recent 3D Gaussian splatting approaches. For instance, Perez-Perez et al. [22] proposed a learning-based pipeline combining semantic labels and geometric descriptors, leveraging classifiers (SVM, AdaBoost) and Markov Random Field optimization, to segment point clouds for automated 3D model fitting. Semantic-geometric pipeline was demonstrated in industrial contexts enriching RGB-point clouds with instance segmentation for geometry fitting to yield structural and M&E models [21]. Another study [18] proposes an instance-level clustering method for road markings from mobile LiDAR, robust to occlusion and complex geometric conditions across 28 datasets. Complementing these, Hu and Brilakis [16] introduced instance segmentation anchored in design-intent via matching IFC models to as-built scenes using PROSAC-based detection and DBSCAN clustering, which achieves high precision in cluttered, occluded scans.

Another promising direction is the development of geometry-aware AI, in which BIM-derived priors, such as the verticality of walls or the planarity of slabs, are incorporated to regularize predictions in real-world scans. New representation learning pipelines can leverage BIM-synthetic point clouds as training data, combined with domain adaptation to bridge discrepancies between synthetic and real scans [15]. These approaches can be extended to zero- or few-shot scenarios by applying large vision-language models (e.g., OpenMaskDINO3D, SegPoint, Reason3D).

Beyond data capture and representation learning, geometric synchronization is another direction where point cloud data streams are continuously used for geometric updates. Bosché et al. [3] have combined Scan-to-BIM and Scan-vs-BIM techniques in a unified approach for automated comparison of as-built and as-planned M&E works. Future work seeks to explore automatic as-built model updating, where deviations between reality capture and the design model are detected and translated into structured updates, enabling self-correcting digital representations. Erişen et al. [8] proposed a multi-task domain-adapted deep learning framework capable of generating semantically annotated 3D reconstructions directly from a single RGB image, delivering mesh accuracy and semantic fidelity.

These developments indicate a research trajectory that integrates sensor-agnostic reconstruction (especially mobile scans) with 4D spatiotemporal geometric updates. Together, the research paves the way toward enriched and adaptable 3D scene understanding that evolves with real-world environments.

3 Developing NUS3D Dataset

The proposed NUS3D dataset aims to integrate multiple sources of 3D point clouds to support the testing and verification of scene understanding and environment perception. It comprises three subsets shown in **Table 2**.

BIM-synthetic point clouds are derived from BIM by sampling geometric primitives and surfaces directly from building elements, producing noise-free and complete synthetic representations. Mobile-scan point clouds are captured using active or passive 3D scene reconstruction systems supported by SLAM algorithms. TLS point clouds are obtained through high-resolution time-of-flight laser scanning from ground-based stations, providing dense and accurate data. It includes structural and mechanical and electrical (M&E) components.

Table 2. Structure of NUS3D dataset

Subset	Method	Description
BIM-Synthetic Point Clouds	BIM 3D model	This subset comprises synthetic 3D point clouds generated from BIM providing complete and noise-free geometric representations of building elements.
Mobile-Scan Point Clouds	Handheld or robotic scans with SLAM	This subset comprises real-world 3D point clouds using handheld or robotic scan (e.g., RealSense L515 or SHARE SLAM S20), capturing practical variations in density and noise.
TLS Point Clouds	Terrestrial laser scanner	This subset comprises high-density 3D point clouds using terrestrial laser scanners (e.g., FARO®), providing accurate as-built geometry while reflecting challenges such as occlusion in real environments.

Note: As different sensing devices exhibit varying levels of accuracy, density, and noise characteristics, this paper does not differentiate such sensor-specific variations at a finer level.

3.1 BIM-Synthetic Point Clouds

BIM-synthetic point clouds are synthetic datasets produced by sampling geometric primitives and surfaces directly from building elements within a BIM model. This process ensures that representations are both noise-free and comprehensive, providing ideal ground truth for downstream learning. A notable application of this subset is its use in BIM-to-scan semantic segmentation, where synthetic point clouds serve as training data to overcome the domain gap between as-designed and as-built data. For example, a domain adaptation network, DawNet, augments segmentation performance by combining hybrid attention mechanisms with feature whitening, significantly enhancing accuracy when adapting to noisy real-world scans [15].

3.2 Mobile-Scan Point Clouds

Figure 2 illustrates mobile-scan point clouds conducted since 2022. The environments present dense furniture, narrow corridors, reflective surfaces and severe occlusions, combined with semi-structured layouts and transitions between indoor and outdoor domains.

Handheld active or passive 3D scene reconstruction systems are deployed for data acquisition with the aid of robots for some scenes. The 3D scene reconstruction are undertaken with SLAM to support real-time mobile 3D mapping under real-world conditions. The workflow begins with data capture, as investigated in previous studies [14,24]. LiDAR measures depth by emitting laser pulses and calculating distances based on time-of-flight, generating depth maps with spatial geometry. Then, the data captured are matched via SLAM to incrementally reconstruct dense colorized 3D point clouds along the robot or handheld sensor trajectory.

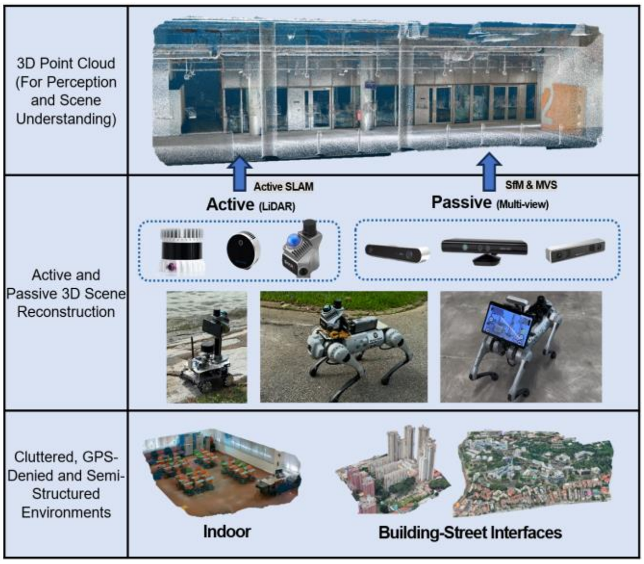

Figure 2. Mobile-scan point cloud acquisition

3.3 TLS Point Clouds

TLS point clouds are generated using high-resolution time-of-flight laser ranging, which captures dense and highly accurate 3D point clouds from ground-based scanning stations. Each point encodes a precise spatial coordinate, enabling faithful reconstructions of architectural and structural elements such as walls, floors, ceilings, columns, and detailed furnishings, including desks, chairs, and equipment. This results in a comprehensive and true-to-life digital representation that supports tasks ranging from spatial measurement and geometric analysis to the built facilities.

Compared with mobile-scan or BIM-synthetic point clouds, TLS exhibits higher density and accuracy, but their quality can still be influenced by environmental conditions. Variations in lighting, weakly textured surfaces and reflective materials across indoor and outdoor environments can introduce challenges in point cloud registration.

4 Cross-Domain Point Cloud Semantic Segmentation and Evaluation

4.1 Semantic Segmentation

In this paper, a segmentation experiment is undertaken as an illustrative example to demonstrate the potential benefit of NUS3D dataset. The evaluation focuses on augmenting TLS data with BIM-synthetic point clouds to explore cross-domain learning. The comparison evaluates TLS-only as well as TLS and BIM-synthetic training, using M&E components. In this study, ResPointNet++ [27] is adopted as a representative backbone for controlled evaluation, while broader benchmarking across additional architectures and tasks will be investigated in future work.

Semantic segmentation aims to assign each point a meaningful interpretation, either in the form of a semantic category or an object-level instance label. Classical geometry-based approaches typically utilize local surface descriptors, such as normals, curvature, and smoothness, to group points with similar differential properties, while model-fitting techniques identify structural primitives (e.g., planes, cylinders) through parametric estimation and consensus-based optimization [3]. In contrast, learning-based methods leverage neural architectures designed for irregular point sets to learn hierarchical feature representations and perform dense, point-wise classification under supervised or weakly supervised settings. The network is trained for 200 epochs on an NVIDIA RTX 4090 GPU.

The experiments are conducted on four real TLS scan M&E scenes (P1-P4) [19] together with four BIM-synthetic scenes (V1-V4). The synthetic point clouds are generated by exporting BIM models to OBJ format and importing them into CloudCompare, where virtual scanner positions are placed to mimic realistic acquisition effects such as occlusion and view-dependent incompleteness. The training and testing scenes are shown in **Figure 3**. To assess the impact of synthetic data, two training regimes are examined under identical hyperparameter settings: (i) a model trained solely on real TLS scans, and (ii) a model trained on the combined set of real TLS scans and BIM-synthetic point clouds as shown in **Figure 4**. This experimental design allows for a controlled evaluation of how synthetic augmentation contributes to segmentation performance across different M&E component types.

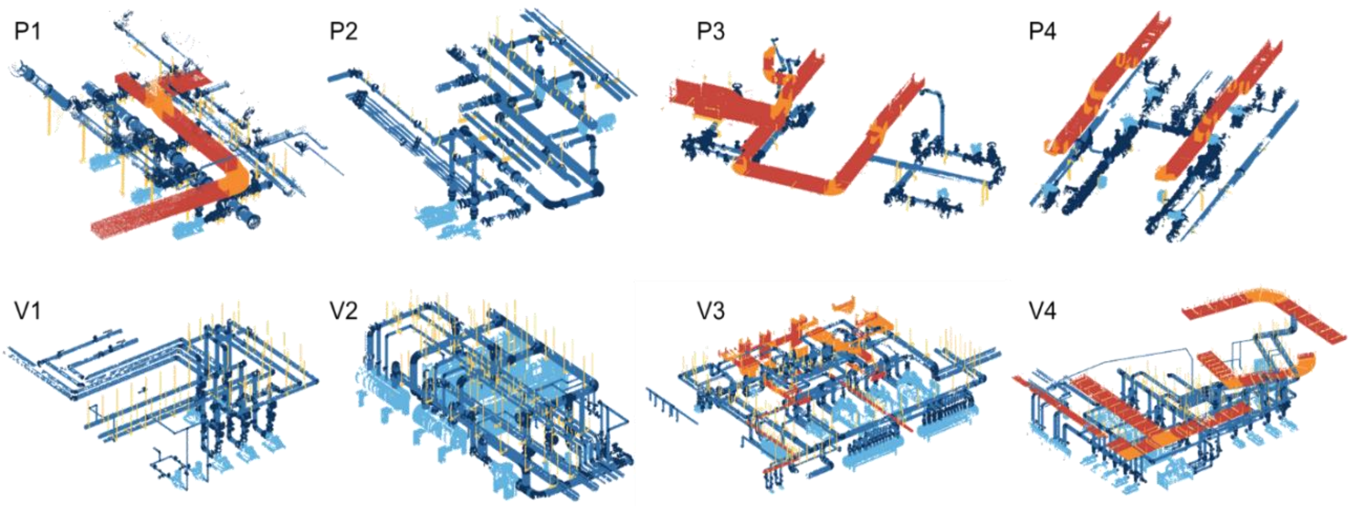

Figure 3. Dataset visualization: training – P1, P2; testing (real) – P3, P4; testing (hybrid) – P3, P4, V1-V4

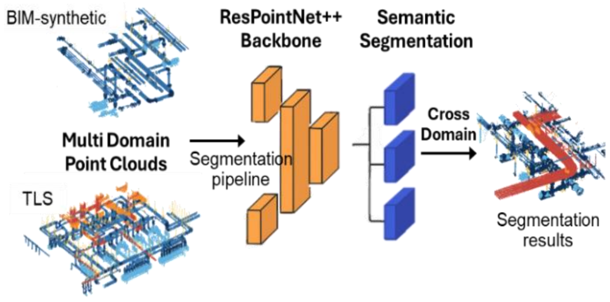

Figure 4. Cross-domain semantic segmentation method

The semantic segmentation output can be refined through a series of post-processing operations, including noise suppression, structural connectivity analysis, and primitive-based geometric fitting. These steps enforce topological consistency, preserve object boundaries, and improve robustness in dense or cluttered environments.

4.2 Evaluation of Segmentation Performance

Evaluation of point cloud segmentation quantitatively assesses the agreement between predicted labels and ground truth. Class-wise performance is typically measured using Intersection over Union (IoU), which evaluates the overlap between the predicted point set and the ground-truth point set of each semantic category. For a given class c, IoU is defined as:

$$\mathrm{IoU}_c = \frac{|\mathcal{P}_c \cap \mathcal{G}_c|}{|\mathcal{P}_c \cup \mathcal{G}_c|} \quad (1)$$

where $\mathcal{P}_c$ and $\mathcal{G}_c$ denote the sets of points predicted and annotated as class c, respectively. The mean Intersection over Union (mIoU), obtained by averaging IoU over all C classes, provides a class-balanced indicator:

$$\mathrm{mIoU} = \frac{1}{C}\sum_{c=1}^{C} \mathrm{IoU}_c \quad (2)$$

Additional metrics include Overall Accuracy (OA):

$$\mathrm{OA} = \frac{\sum_{c=1}^{C} |\mathcal{P}_c \cap \mathcal{G}_c|}{N} \quad (3)$$

where N is the total number of points. OA measures the proportion of correctly classified points across all categories. Furthermore, per-class accuracy is reported to reveal the model's behavior on individual component types, which is particularly important under class imbalance or when different M&E components exhibit varying geometric complexity.

5 Results and Discussion

Table 3 presents the segmentation performance. The segmentation results are also visualized in **Figure 5**. Incorporating synthetic scenes into the training set yields substantial performance gains across nearly all M&E categories. The model trained solely on real scans exhibits limited generalization, particularly for components that are sparse or absent in the real dataset. This leads to extremely low IoU values for Duct fittings (8.93%), Hangers (32.70%), and Machinery (0.17%), while even common classes such as Ducts and Pipes remain suboptimal. In contrast, the hybrid training regime significantly boosts per-class performance: Machinery (+34.27%), Hangers (+27.32%), and Duct fittings (+24.37%) show the largest increases, and even major categories such as Ducts and Pipes improve by 15.00% and 8.27%, respectively. Overall, mIoU increases from 35.37% to 54.86% (+19.49%), and OA improves by more than 11%.

These results highlight the value of supplementing TLS real scans with BIM-synthetic point clouds. The BIM-synthetic data broadens the geometric and structural variability of component instances, enabling the network to better recognize rare, small, or partially occluded elements in real evaluations. The experimental results further demonstrate the potential of the NUS3D hybrid dataset to support the training and improvement of 3D

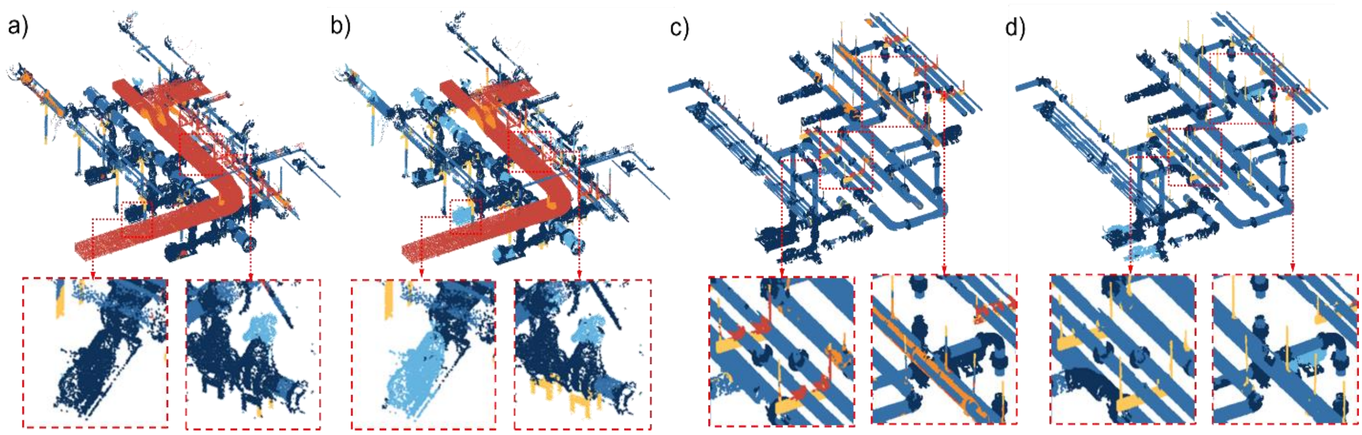

Figure 5. Semantic segmentation results: a) Real – P1; b) Hybrid – P1; c) Real – P2; d) Hybrid – P2

representation learning models. By jointly leveraging BIM-synthetic and TLS data, the dataset improves generalization across design and as-built domains.

Table 3. Segmentation accuracy

Category	TLS	TLS + BIM	Improvement
Duct	56.38	71.37	+15.00
Duct fitting	8.93	33.31	+24.37
Hanger	32.70	60.02	+27.32
Machinery	0.17	34.44	+34.27
Pipe	69.21	77.48	+8.27
Pipe fitting	44.85	52.55	+7.70
mIoU	35.37	54.86	+19.49
OA	66.84	77.96	+11.12

6 Conclusions

This study presents ongoing efforts for developing a hybrid dataset that integrates BIM-synthetic, mobile-scan and TLS point clouds to capture the diverse geometric characteristics of the built environment. NUS3D bridges the gap between design-intent and as-built representations, providing a unified foundation for 3D representation learning and scene understanding for scan-to-BIM and robotic perception. The dataset enables evaluation of algorithms for semantic segmentation, geometric reconstruction, and cross-domain transfer between synthetic and real, static and dynamic, as well as design and as-built domains. Demonstrations on BIM-synthetic and TLS-acquired M&E scenes further illustrate its capability.

Looking ahead, our research will be extended to incorporate mobile-scan point clouds, further enriching the dataset diversity. A cross-domain representation learning pipeline focused on mobile 3D scanning with BIM and TLS point clouds will be investigated with applications to robotic perception, construction progress monitoring and scan-to-BIM and BIM-to-scan verification. These capabilities will support research on point cloud representation learning and generalizable 3D scene understanding.

Acknowledgements

This research is supported by Singapore MOE AcRF Tier-1 Grant (Ref: A-8004304-00-00). Any opinions, findings and conclusions or recommendations expressed in the material are those of the author(s) and do not reflect the views of the grantors.

The dataset presented in this paper integrates point clouds collected by the research group since 2022. The authors would like to express their appreciation to researchers affiliated with the lab and Centre for Digital Building Technology. Users are encouraged to cite this paper as the overarching reference, and, where applicable, cite the corresponding subset publications when using the data.

References

[1] E.B. Anil, P. Tang, B. Akinci, D. Huber, Deviation analysis method for the assessment of the quality of the as-is Building Information Models generated from point cloud data, Automation in Construction 35 (2013) 507-516.

[2] I. Armeni, O. Sener, A. R. Zamir, H. Jiang, I. Brilakis, M. Fischer, S. Savarese,, 3D Semantic Parsing of Large-Scale Indoor Spaces, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016) 1534–1543.

[3] F. Bosché, M. Ahmed, Y. Turkan, C.T. Haas, R. Haas, The value of integrating Scan-to-BIM and Scan-vs-BIM techniques for construction monitoring using laser scanning and BIM: The case of cylindrical MEP components, Automation in Construction 49 (2015) 201-213.

[4] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, O. Beijbom,, nuScenes: A multimodal dataset for autonomous driving, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2020) 11621–11631.

[5] Z. Chen, H. Wang, K. Chen, C. Song, X. Zhang, B. Wang, J.C.P. Cheng, Improved coverage path planning for indoor robots based on BIM and robotic configurations, Automation in Construction 158 (2024) 105160.

[6] A. Dai, A. Chang, M. Savva, M. Halber, T. Funkhouser, M. Nießner,, ScanNet: Richly-annotated 3D reconstructions of indoor scenes, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017) 2432–2443.

[7] D. Eigen, C. Puhrsch, R. Fergus,, Depth Map Prediction from a Single Image using a Multi-Scale Deep Network, Proceedings of the 28th International Conference on Neural Information Processing Systems (NIPS) (2014) 2366 - 2374.

[8] S. Erişen, M. Mehranfar, A. Borrmann, Single Image to Semantic BIM: Domain-Adapted 3D Reconstruction and Annotations via Multi-Task Deep Learning, Remote Sensing 17 (16) (2025) 2910.

[9] M.E. Esfahani, C. Rausch, M.M. Sharif, Q. Chen, C. Haas, B.T. Adey, Quantitative investigation on the accuracy and precision of Scan-to-BIM under different modelling scenarios, Automation in Construction 126 (2021) 103686.

[10] V.J. Gan, D. Hu, Y. Wang, R. Zhai, Automated indoor 3D scene reconstruction with decoupled mapping using quadruped robot and LiDAR sensor, Computer‐Aided Civil and Infrastructure Engineering 40 (15) (2025) 2209-2226.

[11] A. Geiger, P. Lenz, C. Stiller, R. Urtasun,, Vision meets robotics: The KITTI dataset, International Journal of Robotics Research (2013) 1231–1237.

[12] T. Hackel, N. Savinov, L. Ladicky, J. D. Wegner, K. Schindler, M. Pollefeys,, Semantic3D.net: A new large-scale point cloud classification benchmark, ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences (2017) 91–98.

[13] D. Hu, V.J. Gan, C. Yin, Robot-assisted mobile scanning for automated 3D reconstruction and point cloud semantic segmentation of building interiors, Automation in Construction 152 (2023) 104949.

[14] D. Hu, V.J.L. Gan, T. Wang, L. Ma, Multi-agent robotic system (MARS) for UAV-UGV path planning and automatic sensory data collection in cluttered environments, Building and Environment 221 (2022) 109349.

[15] D. Hu, V.J.L. Gan, R. Zhai, Automated BIM-to-scan point cloud semantic segmentation using a domain adaptation network with hybrid attention and whitening (DawNet), Automation in Construction 164 (2024) 105473.

[16] Z. Hu, I. Brilakis, Matching design-intent planar, curved, and linear structural instances in point clouds, Automation in Construction 158 (2024) 105219.

[17] A. Ibrahim, M. Golparvar‐Fard, K. El‐Rayes, Metrics and methods for evaluating model‐driven reality capture plans, Computer‐Aided Civil and Infrastructure Engineering 37 (1) (2022) 55-72.

[18] J. Jung, E. Che, M.J. Olsen, C.E. Parrish, Y. Turkan, S. Yoo, Instance-Based Clustering of Road Markings with Wear and Occlusion from Mobile Lidar Data, Journal of Computing in Civil Engineering 38 (4) (2024) 04024021.

[19] M. Li, V.J.L. Gan, B. Wang, Integrating hierarchical segmentation and vision-language reasoning for spatially complex and occluded MEP point clouds, Automation in Construction 179 (2025) 106455.

[20] Y. Liu, H. Huang, G. Gao, Z. Ke, S. Li, M. Gu, Dataset and benchmark for as-built BIM reconstruction from real-world point cloud, Automation in Construction 173 (2025) 106096.

[21] F. Noichl, Y. Pan, M.S. Mafipour, A. Braun, I. Brilakis, A. Borrmann, From enriched point cloud to structural and MEP models: An automated approach to create semantic-geometric models for industrial facilities, Computing in Civil Engineering 2023, 2023, pp. 92-99.

[22] Y. Perez-Perez, M. Golparvar-Fard, K. El-Rayes, Segmentation of point clouds via joint semantic and geometric features for 3D modeling of the built environment, Automation in Construction 125 (2021) 103584.

[23] J. Sturm, N. Engelhard, F. Endres, W. Burgard, D. Cremers,, A benchmark for the evaluation of RGB-D SLAM systems, Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (2012) 573–580.

[24] T. Wang, V.J.L. Gan, Automated joint 3D reconstruction and visual inspection for buildings using computer vision and transfer learning, Automation in Construction 149 (2023) 104810.

[25] T. Wang, V.J.L. Gan, Enhancing 3D reconstruction of textureless indoor scenes with IndoReal multi-view stereo (MVS), Automation in Construction 166 (2024) 105600.

[26] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, J. Xiao,, 3D ShapeNets: A deep representation for volumetric shapes, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2015) 1912–1920.

[27] C. Yin, B. Wang, V.J.L. Gan, M. Wang, J.C.P. Cheng, Automated semantic segmentation of industrial point clouds using ResPointNet++, Automation in Construction 130 (2021) 103874.

[28] R. Zhai, J. Zou, V.J. Gan, X. Han, Y. Wang, Y. Zhao, Semantic enrichment of BIM with IndoorGML for quadruped robot navigation and automated 3D scanning, Automation in Construction 166 (2024) 105605.