Large-Scale Point Clouds Registration Using Equivariant Graph Neural Network with Building Features

Jinyi Shi[1] and Yujie Lu[1,2]

[1]Shanghai Research Institute for Intelligent Autonomous Systems, Tongji University, China.
[2]College of Civil Engineering, Tongji University, China.
2411928@tongji.edu.cn, lu6@tongji.edu.cn

Abstract –

Registration between as-designed BIM and as-built point clouds is a critical technology in digital construction. However, challenges such as partial overlap and substantial noise limit the effectiveness of traditional methods, particularly in large-scale building scenarios. To address these limitations, this study proposes a building structure matching method based on a geometric-topological joint Equivariant Graph Neural Network (Equivariant GNN) for large-scale full-scene alignment. First, an Equivariant GNN is employed to model the geometric features of individual buildings and the topological relationships among them, while a contrastive learning strategy is introduced to refine the feature space for accurate building matching. Additionally, a score-driven optimization is incorporated as an auxiliary step to achieve high-precision alignment, leveraging the matching results as initialization. Experiments on real construction sites demonstrate that the proposed method consistently outperforms all baselines, achieving a precision of 96.7% with a Mean Absolute Error (MAE) of 0.16 m on Project A and 0.81 m on Project B. Compared to the best-performing baselines, this represents a reduction in MAE of 10.8% and 7.4%, respectively. This study provides an efficient and robust solution for point cloud registration, establishing a foundation for automated construction monitoring.

Keywords –

Point Cloud Regis\4ration; Building Information Modeling (BIM); Graph Neural Network (GNN); Automated Construction Management

1 Introduction

Continuous and systematic monitoring of construction quality and progress is a fundamental aspect of project control. As a primary asset originating during the design phase, BIM is indispensable for automated inspection in construction management. The alignment of as-designed BIM models with as-built point clouds serves as a core enabling technology for digital construction, supporting critical tasks such as progress monitoring, quality inspection, as-built verification, and renovation planning [1]. The widespread adoption of Unmanned Aerial Vehicle (UAV) photogrammetry has made obtaining high-density point cloud coverage a standard practice [2]. These advanced techniques provide a robust geometric foundation for the large-scale alignment between as-designed and as-built point clouds.

However, automated global registration for large-scale scenes remains a significant technical challenge. First, large-scale scenes entail high computational complexity and inherent matching ambiguity [3]. Typical projects involve dozens of buildings and millions of points. Point-to-point registration methods, most notably ICP and its variants, suffer from high computational complexity and susceptibility to local optima [4]. Additionally, BIM and point cloud data exhibit fundamental structural differences [5]. Factors such as partial construction, severe occlusions, and structural similarities on real construction sites exacerbate the inherent ambiguity in establishing accurate correspondences.

In recent years, deep learning methods have made significant advances in point cloud registration tasks. Approaches such as PointNet [6] and PREDATOR [7] employ end-to-end learning strategies to directly learn global transformation matrices from pairs of input point clouds, demonstrating excellent performance on standard datasets. However, these methods exhibit significant limitations when applied to large-scale building complexes [8]. Graph Neural Networks (GNNs), with their ability to explicitly model graph-structured data, can capture both topological relationships and geometric constraints between architectural elements. This capability offers a novel approach to addressing point cloud matching challenges in complex architectural scenarios.

Motivated by these issues, this paper proposes a building structure matching method based on a geometric-topological joint equivariant graph neural network (Equivariant GNN) for large-scale alignment. First, an Equivariant GNN is employed to model the geometric features of individual buildings and their inter-building topological relationships, while a contrastive learning strategy is used to refine the feature space, enabling accurate building matching. Subsequently, a score-based optimization module is implemented to achieve high-precision alignment. The core idea is to reformulate the geometric alignment problem of large-scale scene point clouds as a building-level structural matching problem within a shared equivariant feature space. The proposed geometry-topology joint Equivariant GNN aligns heterogeneous BIM models and as-built point clouds within a unified feature space using equivariant layers and cross-graph attention mechanisms.

The remainder of this paper is structured as follows: Section 2 reviews related work. Section 3 details the key modules of the proposed method. Section 4 describes the experimental setup and presents a comprehensive analysis of the results. Section 5 concludes the paper and outlines future research directions.

2 Literature Review

Point cloud registration [9] is defined as the process of aligning two point clouds by accurately estimating a spatial transformation that establishes precise correspondences between matched points. Typically, one point cloud serves as the reference, while the other is transformed to achieve optimal overlap and spatial consistency. This process generally follows a coarse-to-fine strategy and is regarded as a fundamental problem in computer vision and robotics [10].

Existing registration methods include manual calibration and automated geometric techniques [11]. Manual registration achieved through least-squares optimization of control points is accurate but suffers from poor scalability and low automation in large-scale projects [12]. Automated geometric methods typically align point clouds by matching local descriptors. The ICP algorithm iteratively minimizes distances between point pairs but is sensitive to initial poses and susceptible to local optima [4]. FPFH encodes local geometric information to generate feature descriptors and is often combined with RANSAC for robust matching [13]. However, in architectural environments, these methods often suffer from high failure rates due to sparse features and structural repetition [4]. Moreover, their computational complexity increases quadratically with scene scale, making them challenging to apply to large-scale building cluster registration tasks.

To overcome the limitations of traditional methods, deep learning approaches have achieved significant progress in point cloud registration in recent years. PointNet [6] pioneered the application of deep learning to point cloud processing by utilizing symmetric functions to achieve permutation invariance. Subsequently, a series of variants based on PointNet have been proposed to enhance its representational capacity and applicability. For instance, PointNet++ [14] incorporates hierarchical sampling and local feature aggregation mechanisms, enabling multi-scale modeling of intrinsic geometric structures in point clouds. Despite their success on standard datasets, these methods often struggle with large-scale building clusters, which hinders the effective handling of structural discrepancies and local correspondences between heterogeneous BIM and point clouds.

GNNs have gained attention for their ability to explicitly model the graph structure of data [15] and have demonstrated strong capabilities in point cloud processing tasks. While some studies have explored the potential of GNNs in registration tasks [16], their focus remains primarily on small-scale scenes or homogeneous registration. Concurrently, recent research has emphasized the importance of geometric equivariance. E(n) Equivariant GNNs [17] maintain equivariance to rotations and translations, enabling the learning of geometrically invariant feature representations. However, applying Equivariant GNNs to large-scale heterogeneous BIM and point cloud registration, particularly in combination with topological constraints for building-level matching, remains an open problem.

In summary, despite the significant progress in point cloud registration, existing methods still encounter notable limitations in heterogeneous large-scale scenarios. Traditional methods are constrained by high computational complexity and insufficient feature representation. Deep learning approaches often overlook the spatial topological structure and vertical hierarchical characteristics of building clusters, while current GNN methods have yet to fully integrate equivariance with topological modeling for architectural structures. This study represents a step toward integrating architectural semantics, bidirectional alignment, and topological constraints into equivariant feature learning for large-scale registration between heterogeneous BIM model and point cloud data.

3 Proposed Method

3.1 Method Overview

The study proposes a multi-stage automatic registration method for achieving global alignment between as-designed BIM models and as-built point clouds in large-scale construction environments involving building clusters. As illustrated in Figure 1, the proposed framework consists of three stages:

(1) **Asymmetric Multi-scale Graph Modeling**: The BIM model is hierarchically represented at three levels: region, building, and component. For the on-site point

cloud, building-level point clouds are extracted through downsampling and clustering, accommodating the structural differences between the two types of data.

(2) **Building Cluster Feature Matching based on Equivariant GNN**: A multi-layer building graph is constructed, and a geometry-topology-aware Equivariant GNN is developed to encode features of individual buildings and the relationships between them. The feature space is further optimized through learning.

(3) **Score-Driven Fine Alignment based on Feature Matching**: Fine registration is achieved by aligning the feature space and performing 6-degree-of-freedom (6-DoF) optimization based on building matching relationships.

The input consists of as-designed BIM models and as-built point clouds, while the output consists of precise transformation matrices.

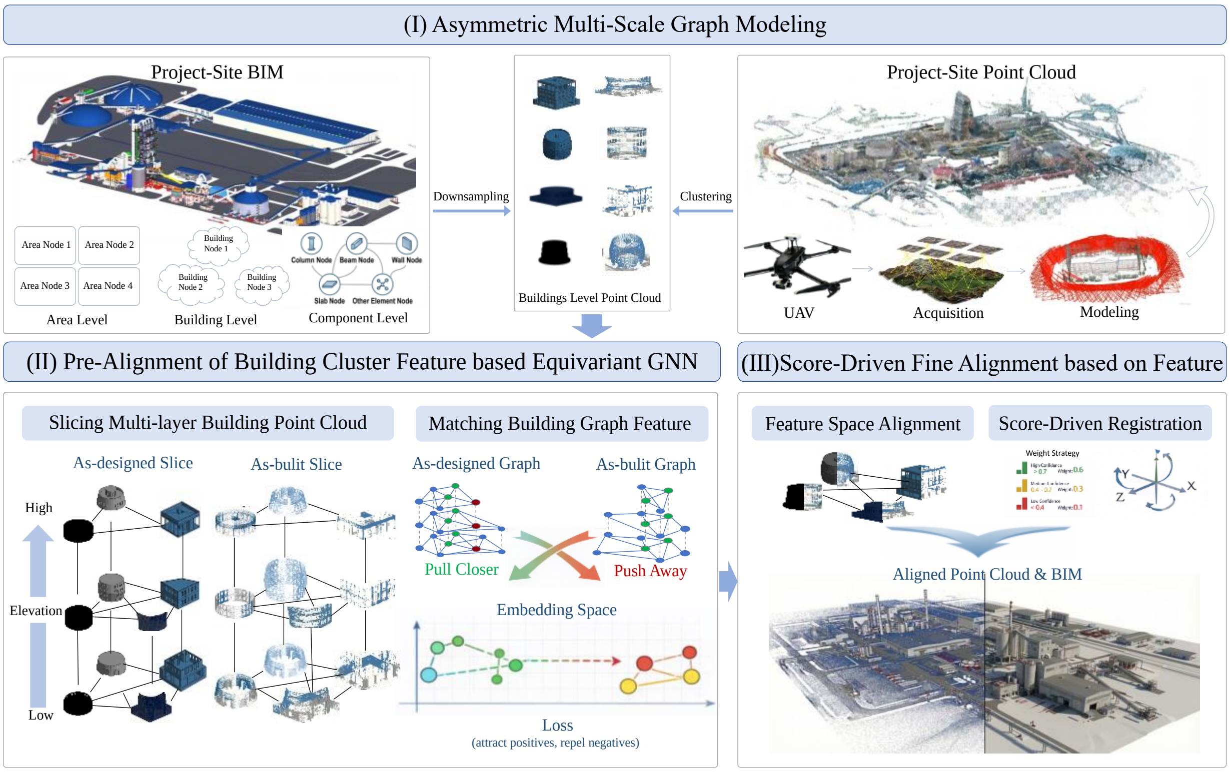

Figure 1. Registration method based on Hierarchical Graph Network

3.2 Multi-scale Graph Modeling

To efficiently process large-scale point cloud data and fully leverage the prior structural knowledge embedded in BIM models, we propose an asymmetric multi-scale graph modeling approach. To this end, a three-level hierarchical graph is constructed for the BIM model, while a single-layer graph is built for the as-built point cloud. By enabling heterogeneous feature learning at the building level, this design facilitates efficient coarse registration in large-scale construction scenarios.

For LOD 300 BIM models, a three-tier hierarchical graph is constructed to capture architectural information ranging from detailed to coarse, as shown in Figure 2. At the component level (the finest granularity), independent elements such as columns, beams, walls, and slabs are extracted from IFC files and mapped as graph nodes. At the building level, component nodes belonging to the same building are grouped, and their features are aggregated to form building-level nodes. This level serves as a pivotal abstraction layer, bridging fine-grained geometric components with broad spatial partitions and providing semantic anchors for cross-modal alignment between BIM and point cloud data in subsequent stages. The regional level divides large construction sites into spatial partitions for coarsest abstraction. Each regional node is generated by aggregating features of the buildings within its area, forming a high-level spatial representation that supports full-site alignment. The BIM graph incorporates two distinct edge types: hierarchical edges capture containment relationships across different levels (e.g., components belonging to buildings, buildings assigned to regions), while adjacency edges encode spatial neighborhood relationships within the same level.

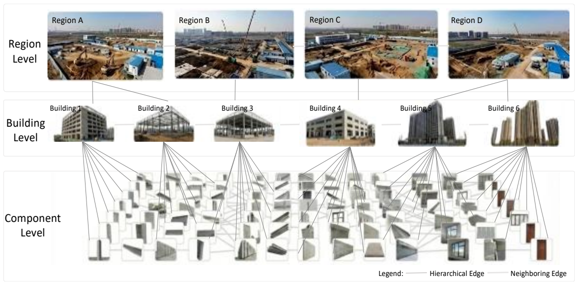

Figure 2. Three-Level hierarchical structure of the as-designed BIM Graph

For point cloud data, a single-layer graph structure is constructed exclusively at the building level to mitigate challenges posed by incompleteness, noise, and the absence of semantic labels. The DBSCAN algorithm is employed on elevation-based projections of the point cloud to segment individual buildings as graph nodes.

3.3 Building Cluster Feature Matching based on Equivariant GNN

3.3.1 Multi-scale Slicing of Building Point Clouds

A multi-level vertical slicing strategy based on elevation is employed to capture structural architectural features across various altitudinal levels. Multiple vertical slices are extracted using height percentiles to segment distinct vertical building layers.

For each height slice, a building-level graph is constructed, where nodes represent individual buildings and edges encode spatial proximity. Building nodes are extracted through efficient grid-based clustering on elevation planes. Points within the slice are projected onto the horizontal plane and grouped using grid partitioning. Connected grid cells form building clusters, and outlier clusters with insufficient point density are pruned.

For each node v_i, an 11-dimensional feature vector f_i is computed to encode geometric and spatial attributes:

$$f_i = [p_{rel}, d_i, n_i, z_i, \rho_i] \tag{1}$$

Where $p_{rel} \in R^3$ represents the relative position to the centroid of the building, $d_i \in R$ denotes the distance to the centroid, $n_i \in R^3$ is the normalized directional vector, $z_i \in R$ is the absolute elevation, and $\rho_i \in R^3$ captures local density features.

3.3.2 Equivariant GNN Modeling

After constructing the multi-scale hierarchical graph, we employ an Equivariant GNN to extract descriptive embeddings for each building while maintaining rotation and translation equivariance.

The network consists of stacked equivariant layers. Each layer processes node features and coordinates separately to maintain equivariance. For any node i in the l-th layer, the update rule is defined as follows:

$$h_i^{(l+1)} = \sigma\left(h_i^{(l)} + \sum_{j \in N(i)} \phi_e\left(h_i^{(l)}, h_j^{(l)}, \|x_i^{(l)} - x_j^{(l)}\|^2\right)\right) \tag{2}$$

$$x_i^{(l+1)} = x_i^{(l)} + \sum_{j \in N(i)} \phi_c\left(m_{ij}^{(l)}\right) \cdot \frac{x_i^{(l)} - x_j^{(l)}}{\|x_i^{(l)} - x_j^{(l)}\|} \tag{3}$$

where $h_i^{(l)}$ represents the feature of node i at the l-th layer, $x_i^{(l)}$ denotes the coordinate, $N(i)$ denotes the set of adjacent neighbors, ϕ_e denotes the edge message function (implemented as an MLP) that processes scalar invariants, and ϕ_c is the coordinate update MLP that generates radial coefficients.

3.3.3 Building Graph Feature Matching

After encoding as-designed and as-built building features using the Equivariant GNN, global representations are derived through global average pooling of node embeddings:

$$f_{BIM} = \frac{1}{N} \sum_{i=1}^{N} e_i^{BIM} \tag{4}$$

$$f_{scene} = \frac{1}{M} \sum_{j=1}^{M} e_j^{scene} \tag{5}$$

Where N and M are the numbers of nodes in as-designed and as-built graphs respectively. This global pooling encodes the global structural topology of each building.

The matching probability is computed by a Multi-Layer Perceptron (MLP) classifier, which takes the concatenated global features from both buildings as input. The network is trained to discriminate between positive and negative building correspondences by integrating geometric and structural information encoded in the Equivariant GNN embeddings. During training, the classifier is optimized using binary cross-entropy loss. During the inference phase, candidate matches are validated only if the predicted probability exceeds a decision threshold. This simple and interpretable approach ensures only high-confidence matches are accepted, enhancing the reliability of the matching results.

3.4 Score-Driven Fine Alignment

3.4.1 Feature Space Alignment

For each matched building pair obtained, a coarse rigid transformation is estimated using the corresponding point clusters. If only one building pair is matched, their centroids are first aligned via translational shift, followed by rotational alignment around the common center of mass. If multiple building pairs are matched, an in-plane

Kabsch algorithm is applied in the plane, with optimal rotations computed via Singular Value Decomposition (SVD) to provide initial alignment results.

3.4.2 Score-Driven Registration

To refine the alignment, a 6-DoF iterative optimization strategy driven by scoring feedback is employed. In contrast to the point-to-point correspondence objective in traditional ICP, this method achieves robust fine registration through stepwise adjustments and scoring verification for each DoF.

The optimization process sequentially fine-tunes the 6-DoF (translation t_x, t_y, t_z and rotation r_x, r_y, r_z). For DoF, an incremental perturbation Δ is applied to the extrinsic parameters, and an alignment score is computed:

$$score = f_{eval}(T(\Delta) \cdot T_{current}) \quad (6)$$

where f_{eval} represents an alignment fidelity metric that considers both point cloud overlap and geometric consistency. To prevent entrapment in local optima, a stochastic escape mechanism is implemented: if the alignment score stagnates, a random perturbation is triggered to explore the neighbor search space. A convergence-based termination criterion is enforced: should the score fail to improve across multiple consecutive iterations over all DoFs, the optimization terminates. The final optimal transformation T^* is obtained by accumulating all valid perturbations.

4 Experiments and Results

4.1 Experiment Setup

The study employs data from six real-world construction projects for training and case studies. Point cloud data for all projects were automatically collected using UAV-based oblique photogrammetry and reconstructed into 3D point cloud models with the assistance of RTK. The dataset comprises 815 sample pairs across four projects, while the remaining two projects are reserved for evaluating model performance during the validation phase. To reduce annotation costs, an automated pipeline for training sample synthesis is introduced to extract positive and negative sample pairs from as-designed and as-built point data.

First, as-designed and as-built point clouds from the training projects were manually aligned to establish a ground truth. These aligned clouds were then sliced along the Z-axis (elevation direction) using multi-level vertical stratification based on absolute coordinates. Slicing configurations were automatically generated based on the overlapping regions, with a fixed slice thickness of 10 m. For each slice, the DBSCAN algorithm was applied to extract independent building clusters. Negative correspondences were synthesized through a stochastic pairing strategy, where as-designed and as-built clusters were randomly paired. To ensure a balanced dataset, the number of negative samples matched the positive samples, and the resulting dataset was split into training and validation sets in an approximately 6:1 ratio. Notably, the test scenarios were kept strictly disjoint from the training set.

Second, to ensure sample accuracy, manual quality assurance was conducted using CloudCompare. Each sample pair was individually inspected to verify spatial correspondences in positive samples and confirm mismatches in negative samples. Furthermore, to enhance rotational robustness, data augmentation was applied by independently performing random 3D rotations on both as-designed and as-built clusters.

Training was conducted on a single NVIDIA RTX 3090 GPU for 100 epochs. To further improve robustness, each sample was augmented with three additional random rotations during training, enabling the model to learn rotation-invariant feature representations.

4.2 Results and Analysis

4.2.1 Ablation Study and Baseline Comparisons

Training metrics and comparative experimental results are summarized in Table 1. To demonstrate the core components of the proposed method, an ablation analysis was conducted to evaluate the Equivariant GNN against the standard GNN and PointNet methods. All models were trained on the same dataset with identical hyperparameters. Experimental results demonstrate that the incorporation of equivariance constraints substantially enhances model performance: compared to the standard GNN, yielding a 19.8% improvement in validation accuracy (rising from 0.814 to 0.975), and the F_1 score increased from 0.833 to 0.975. Although PointNet achieved a slightly higher precision than standard GNN (0.776 vs. 0.753), it exhibited a high miss rate, as evidenced by a recall of only 0.605, resulting in the poorest overall performance (F_1= 0.680).

Table 1. Results Comparison of Baseline Methods

Method	Acc	Prec	Recall	F1
Equivariant GNN	0.975	0.967	0.983	0.975
Standard GNN	0.814	0.753	0.932	0.833
PointNet	0.720	0.776	0.605	0.680

To further evaluate the performance of the proposed method, we conducted comparative experiments with several state-of-the-art (SOTA) robust global registration baselines, such as Super4PCS[18] (a robust global registration method based on 4-point congruent sets), FGR[19] (a fast global registration method), TEASER++[20] (a robust registration framework leveraging Truncated Least Squares), and PREDATOR[7] (a learning-based point cloud registration method designed for low-overlap scenarios). The alignment results of the overall method against these methods are presented in the following section.

4.2.2 Project Evaluation Validation

This study validates the effectiveness of the proposed method on two real-world construction project datasets. Project A is a residential development consisting of multiple buildings that exhibit high geometric homology across individual structures. The structural resemblance between buildings poses significant challenges for inter-structure discrimination based solely on geometric descriptors. Project B comprises several office buildings characterized by highly modular structural layouts. The uniform column-beam spacing and repetitive architectural patterns produce point cloud data with strong regularity, necessitating robust discriminative features to resolve spatial ambiguities.

Figure 3 demonstrates the multi-scale building matching results of the proposed method. By partitioning as-designed and as-built point clouds into low, middle, and high elevation levels, the method reliably establishes building-level correspondences. Matching outcomes are visualized via color-coding: corresponding pairs share identical colors between the as-designed and as-built sides, while gray denotes unmatched units. Project A demonstrates the method's effectiveness in intricate residential environments; despite numerous structures, the method successfully associated building units with diverse geometries and scales. Project B validates the method's robustness in handling repetitive structures. Even with highly similar building shapes, the method achieved accurate matching across all height levels, effectively resolving perceptual aliasing between visually similar yet spatially distinct structures.

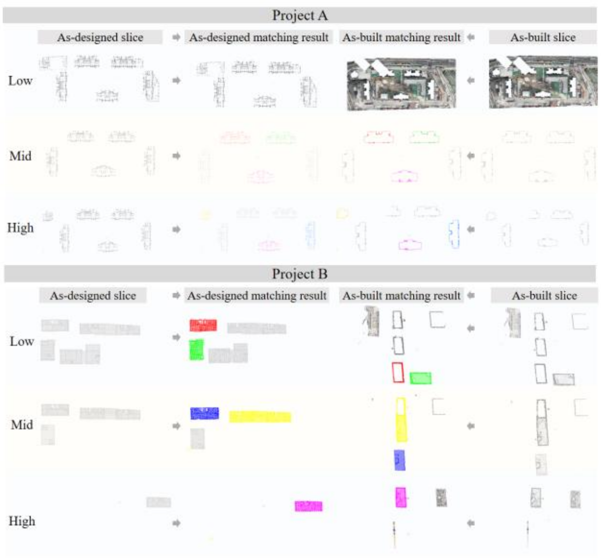

Figure 3. Matching Results of Buildings

Additionally, Table 2 presents a quantitative comparison between the holistic method and traditional approaches using two real-world building project datasets. Given partial overlap between as-designed and as-built point clouds, this study adopts an evaluation protocol specifically localized for partial point cloud registration. For registered point cloud pairs, the Euclidean distance to the nearest neighbor for each point is first computed. Subsequently, the closest point pairs are selected as valid correspondences, with the selection ratio determined through empirical evaluation. Based on manual measurements, ratios of 35% for Project A and 40% for Project B were adopted.

Among the traditional methods, TEASER++ achieved the best performance on Project A (35.7% overlap) with a Mean Absolute Error (MAE) of 0.1837 m and a standard deviation (Matched_Std) of 0.6019 m. In contrast, Super4PCS and FGR exhibited substantially larger registration errors. On Project B (41.7% overlap), FGR yielded an MAE of 3.8461 m, indicating insufficient robustness to scene variations and noise. The learning-based PREDATOR also underperformed, with an MAE of 2.8374 m on Project A, significantly higher than our method.

In contrast, our method achieved the highest accuracy across all metrics. Specifically, on Project A, our method achieved an MAE of 0.1638 m and a Matched_Std of 0.4761 m, surpassing TEASER++ (the second-best) in both precision and consistency. Similarly, on Project B, our MAE of 0.8119 m and Matched_Std of 0.4072 m outperformed Super4PCS. These results demonstrate that our approach not only improves accuracy but also ensures greater registration robustness against outliers—a decisive advantage for practical construction where data quality is inconsistent.

The degradation of conventional methods stems from the violation of their underlying geometric assumptions in construction scenarios. Super4PCS and FGR rely on distinctive geometric primitives and complete structures, which rarely exist in occluded and cluttered as-built scans. While TEASER++ offers resilience to outliers, it remains sensitive to initial correspondence quality, especially when overlapping regions are sparse. For PREDATOR, generalization is constrained by training data distribution; the structural heterogeneity and low overlap of BIM to point cloud tasks fall outside the typical distribution of standard datasets.

Our method is specifically designed for the challenges of multi-building, structurally incomplete, and heavily occluded construction environments—a setting that departs fundamentally from the assumptions of existing algorithms. While conventional methods like Super4PCS, FGR, and TEASER++ presuppose well-structured, high-overlap scenarios for reliable feature extraction, and learning-based approaches such as

PREDATOR require strict distribution alignment between training and test data, our framework bridges this gap by integrating domain-specific architectural priors. By leveraging the hierarchical organization of building structures and inherent topological consistency constraints, the proposed approach achieves robust performance under the adverse conditions typical of real-world construction sites.

Table 2. Quantitative comparison on two datasets (unit: meters)

	Project A (35%)			Project B (40%)		
Method	Matched_Std	MAE	CD	Matched_Std	MAE	CD
Go-ICP	4.9460	2.9804	14.7469	0.6302	0.8876	5.6605
FPFH+RANSAC	5.3926	3.4729	17.0916	0.7450	2.1302	5.8829
Super4PCS	0.9847	0.3923	3.7934	0.4416	0.8771	5.5658
FGR	1.1362	0.6876	5.3538	3.2230	3.8461	18.7076
TEASER++	0.6019	0.1837	2.5361	0.5571	0.9605	10.7627
PREDATOR	3.7148	2.8374	13.7594	2.0719	2.4879	12.7164
Ours	0.4761	0.1638	2.3085	0.4072	0.8119	4.9769

As illustrated in the qualitative results (Figure 4), the proposed method achieves precise alignment for both Project A and Project B. Across both 3D and profile views, our approach consistently establishes accurate structural correspondences within large-scale scenes, characterized by sharp boundaries and high localization fidelity. Conversely, baseline methods such as Go-ICP and FPFH+RANSAC exhibit pronounced misalignment, most notably in the form of vertical tilting and lateral shifts. While Go-ICP shows marginal improvements in Project B, it remains susceptible to local distortions and partial mismatches. These visual assessments underscore the robustness of our multi-stage registration framework when navigating complex architectural geometries and adverse data conditions.

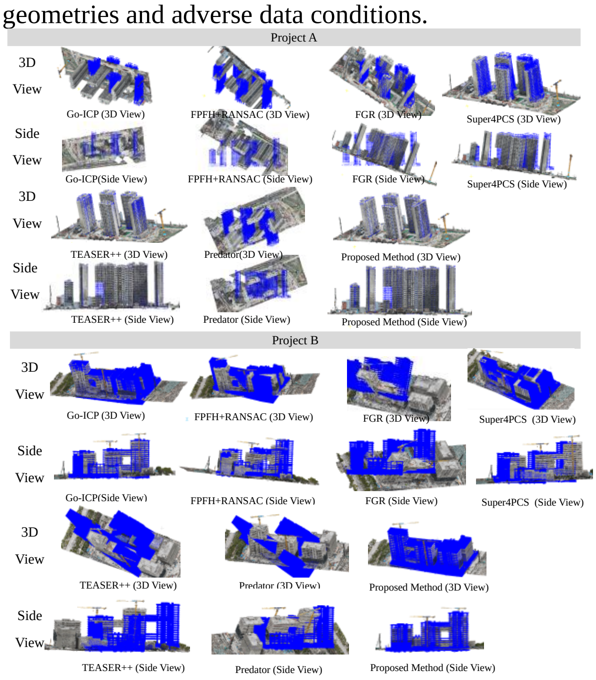

Figure 4. Qualitative Comparison

5 Conclusion

To address the structural discrepancies and matching ambiguities between as-designed BIM and as-built point clouds in large-scale scenarios, this study proposes a building structure matching method based on a geometry-topology joint Equivariant Graph Neural Network (Equivariant GNN). By concurrently encoding individual building geometries and their spatial topological relationships, the method aligns heterogeneous architectural structures within a unified equivariant feature space. This approach effectively mitigates challenges inherent in large-scale construction, such as repetitive structural patterns and expansive search spaces, which typically baffle traditional point-to-point registration.

Ablation studies confirm the superiority of the Equivariant GNN over standard GNN and PointNet architectures under identical experimental configurations. The Equivariant GNN achieved a validation accuracy of 0.975, surpassing the standard GNN (0.814) by 19.8% and PointNet (0.720) by 35.4%. Notably, the model yielded a precision of 0.967 and an F_1 score of 0.975, signifying a negligible rate of false matches—a pivotal requirement for architectural alignment. This performance gain is attributed to the model's rotation-invariant embeddings, which ensure robustness amidst complex geometric and topological configurations.

In quantitative trials across two real-world projects, the proposed method consistently outperformed traditional baselines. For Project A (a residential development), our approach achieved a Mean Absolute Error (MAE) of 0.1638 m, markedly lower than the best-performing baseline, TEASER++ (MAE: 0.1837 m).

Similarly, for Project B (a multi-story office building), our method attained an MAE of 0.8119 m, demonstrating superior resilience compared to FGR (MAE: 3.8461 m). Qualitative assessments further reveal that the method accurately identifies correspondences across vertical layers, maintaining robustness despite repetitive structural patterns and industrial clutter.

Despite these results, certain limitations persist. Our approach is tailored for advanced construction stages; early-stage occlusions and sparse features can compromise topological accuracy. Additionally, the method's reliance on elevation slicing makes structures under four stories (<15 m) harder to process. Finally, the assumption of rigid $SE(3)$ transformation limits its use in vision-based scenarios like SfM or SLAM. Future work will focus on extending this framework to similarity $Sim(3)$ transformations to account for scale drift.

References

[1] N. Abreu, A. Pinto, A. Matos, and M. Pires. Procedural point cloud modelling in scan-to-BIM and scan-vs-BIM applications: A review. *ISPRS International Journal of Geo-Information*, 12(7):260, 2023.

[2] Y. Li and C. Liu. Applications of multirotor drone technologies in construction management. *International Journal of Construction Management*, 19(5):401–412, 2019.

[3] N. Brightman, L. Fan, and Y. Zhao. Point cloud registration: a mini-review of current state, challenging issues and future directions. *AIMS Geosciences*, 9(1):68–85, 2023.

[4] Y. Zhang, L. Lu, X. Luo, and J. Pan. Global BIM-point cloud registration and association for construction progress monitoring. *Automation in Construction*, 168:105796, 2024.

[5] T. Qu and W. Sun. Usage of 3D point cloud data in BIM (building information modeling): Current applications and challenges. *Journal of Civil Engineering and Architecture*, 9(11):1269–1278, 2015.

[6] C. R. Qi, H. Su, K. Mo, and L. J. Guibas. PointNet: Deep learning on point sets for 3D classification and segmentation. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 652–660, Honolulu, USA, 2017.

[7] S. Huang, Z. Gojcic, M. Usvyatsov, A. Wieser, and K. Schindler. PREDATOR: Registration of 3D point clouds with low overlap. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4267–4276, Virtual, 2021.

[8] Y. Zhao, J. Zhang, S. Xu, and J. Ma. Deep learning-based low overlap point cloud registration for complex scenario: The review. *Information Fusion*, 107:102305, 2024.

[9] P. J. Besl and N. D. McKay. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, 1992.

[10] M. Lyu, J. Yang, Z. Qi, R. Xu, and J. Liu. Rigid pairwise 3D point cloud registration: A survey. *Pattern Recognition*, 151:110408, 2024.

[11] J. Marek and P. Chmelař. Survey of point cloud registration methods and new statistical approach. *Mathematics*, 11(16):3564, 2023.

[12] B. Mahmood, S. Han, and D.-E. Lee. BIM-based registration and localization of 3D point clouds of indoor scenes using geometric features for augmented reality. *Remote Sensing*, 12(14):2302, 2020.

[13] R. B. Rusu, N. Blodow, and M. Beetz. Fast point feature histograms (FPFH) for 3D registration. *In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 3212–3217, Kobe, Japan, 2009.

[14] C. R. Qi, L. Yi, H. Su, and L. J. Guibas. PointNet++: Deep hierarchical feature learning on point sets in a metric space. *In Advances in Neural Information Processing Systems*, pages 5099–5108, Long Beach, USA, 2017.

[15] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini. The graph neural network model. *IEEE Transactions on Neural Networks*, 20(1):61–80, 2009.

[16] X. Zhang, J. Li, W. Zhang, Y. Xu, and F. Li. Two-stage point cloud registration framework based on graph neural network and attention. *Electronics*, 13(3):654, 2024.

[17] V. G. Satorras, E. Hoogeboom, and M. Welling. E(n) equivariant graph neural networks. *In Proceedings of the 38th International Conference on Machine Learning (ICML)*, pages 9323–9332, Virtual, 2021.

[18] N. Mellado, D. Aiger, and N. J. Mitra. Super 4PCS fast global pointcloud registration via smart indexing. *Computer Graphics Forum*, 33(5):205–215, 2014.

[19] Q. Y. Zhou, J. Park, and V. Koltun. Fast global registration. *In European conference on computer vision (ECCV)*, pages 766–782, Amsterdam, Netherlands, 2016.

[20] H. Yang, J. Shi, and L. Carlone. TEASER: Fast and certifiable point cloud registration. *IEEE Transactions on Robotics*, 37(2):314–333, 2021.