# MCI-Net: A Robust Multi-Domain Context Integration Network for Point Cloud Registration

**Shuyuan Lin[1], Wenwu Peng[1], Junjie Huang[1], Qiang Qi[2], Miaohui Wang[3], Jian Weng[1]***

[1]College of Cyber Security, Jinan University, Guangzhou, China
[2]School of Data Science, Qingdao University of Science and Technology, Qingdao, China
[3]College of Computer Science  Software Engineering, Shenzhen University, Shenzhen, China
swin.shuyuan.lin@gmail.com

## Abstract

Robust and discriminative feature learning is critical for high-quality point cloud registration. However, existing deep learning–based methods typically rely on Euclidean neighborhood-based strategies for feature extraction, which struggle to effectively capture the implicit semantics and structural consistency in point clouds. To address these issues, we propose a multi-domain context integration network (MCI-Net) that improves feature representation and registration performance by aggregating contextual cues from diverse domains. Specifically, we propose a graph neighborhood aggregation module, which constructs a global graph to capture the overall structural relationships within point clouds. We then propose a progressive context interaction module to enhance feature discriminability by performing intra-domain feature decoupling and inter-domain context interaction. Finally, we design a dynamic inlier selection method that optimizes inlier weights using residual information from multiple iterations of pose estimation, thereby improving the accuracy and robustness of registration. Extensive experiments on indoor RGB-D and outdoor LiDAR datasets show that the proposed MCI-Net significantly outperforms existing state-of-the-art methods, achieving the highest registration recall of 96.4% on 3DMatch.

**Code** — http://www.linshuyuan.com

## Introduction

Point cloud registration seeks to estimate the optimal transformation between a pair of partially overlapping point clouds and serves as a fundamental task in various 3D vision applications (Hu et al. 2024; Zhang et al. 2024; Lin et al. 2022), such as reconstruction, tracking, and robotic localization. Classical methods typically adopt a two-stage paradigm: first, encoding geometric structures into feature descriptors; then, establishing correspondences by matching the most similar descriptors between the two frames.

Establishing robust and distinctive feature correspondences is essential for high-quality point cloud registration. Recently, deep neural network–based methods (Ao et al. 2021; Yu et al. 2023a) have been widely used for local feature extraction. However, most of these approaches rely on Euclidean geometry to define local neighborhoods—typically using k-nearest neighbors or fixed-radius search—which inherently limits their ability to capture high-level semantic and structural consistency. Moreover, relying solely on geometric neighborhoods to construct local structures in point clouds presents further limitations. These limitations hinder accurate feature extraction and the establishment of reliable correspondences, ultimately reducing the robustness and accuracy of registration. To address these challenges, recent studies have employed hyperbolic feature embedding (Xie et al. 2024; Ermolov et al. 2022) to enhance the modeling of implicit semantic relationships in point clouds. Other works have introduced local attention mechanisms (Lai et al. 2022; Fan et al. 2022; Lai et al. 2022; Lin et al. 2024) to assign varying levels of importance to each point's neighbors, thereby suppressing noise and emphasizing key features. Neighborhood learning methods, such as (Lu et al. 2021), use networks to learn the weights of neighbors within geometric clusters and aggregate them accordingly to improve representation quality. While these

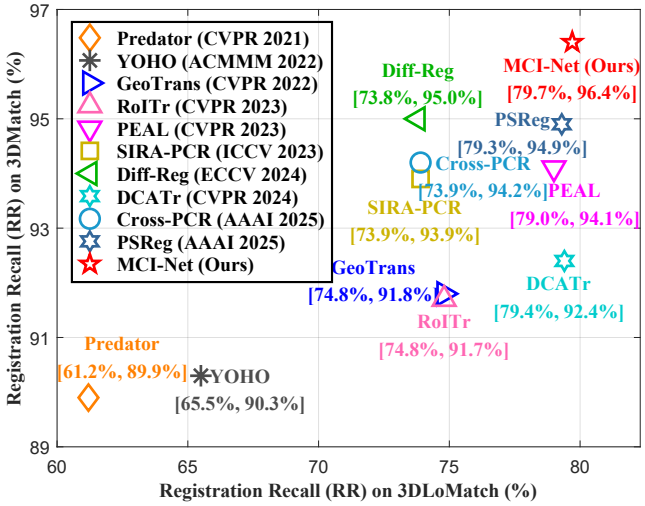


Figure 1: Registration recall (RR) on 3DLoMatch and 3DMatch with 2,500 points. The proposed method significantly outperforms all the state-of-the-art approaches.

*Corresponding Author.

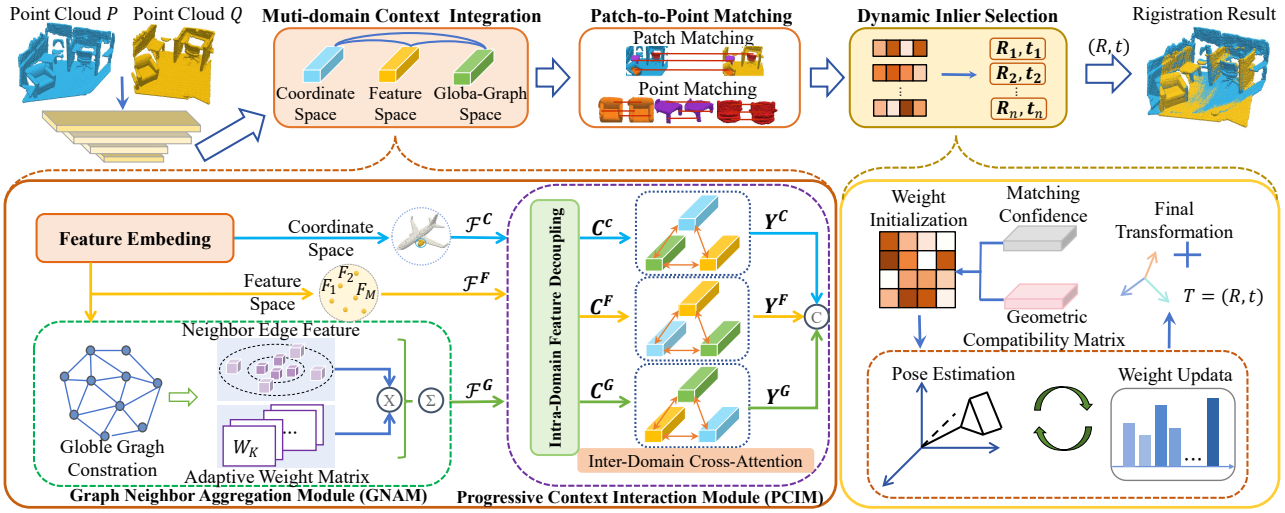


Figure 2: Architecture of the proposed MCI-Net. Given point clouds $P$ and $Q$, patch-level and point-level features are first extracted using the PAConv-based feature pyramid network (Yao et al. 2024). The extracted patch features are embedded into coordinate and feature spaces, and then processed by the multi-domain context integration module, where GNAM captures global structural relationships, and PCIM decouples and fuses contextual information across domains to enhance feature discriminability. The patch-to-point matching strategy is used to obtain point correspondences. Finally, the dynamic inlier selection method iteratively updates correspondence weights to progressively refine the inlier set, yielding the final pose estimation $(R, t)$.

methods have improved the robustness and adaptability of local features, they remain constrained by geometric space modeling and fail to overcome the semantic expressiveness limitations of Euclidean neighborhoods.

Consequently, there remains a critical need for neighborhood construction strategies that can capture global structural patterns and implicit semantic relationships beyond local geometric constraints. Inspired by recent advances in image matching (Dai et al. 2022; Zhao et al. 2021; Liu et al. 2024; Luanyuan et al. 2024), we propose the **multi-domain context integration network (MCI-Net)** for robust point cloud registration. First, we propose the **graph neighborhood aggregation module (GNAM)** that effectively captures structural relationships within point clouds by constructing a global graph. Second, we design the **progressive context interaction module (PCIM)** that combines intra-domain feature decoupling and inter-domain context interaction to enhance feature representation across spatial domains. Finally, we propose the **dynamic inlier selection method (DISM)** that continuously optimizes the correspondence weights by integrating historical residuals from pose estimation, thereby adaptively highlighting reliable correspondences and suppressing outliers. As shown in Figure 1, MCI-Net achieves the best RR compared to other competing methods in different overlapping scenes. An overview of the MCI-Net architecture is shown in Figure 2, and the main contributions of this paper are as follows:

- We propose MCI-Net, integrating multi-domain contextual information to effectively improve feature representation and registration performance.
- We propose GNAM that constructs a global graph to capture holistic structural relationships within point clouds.
- We propose PCIM that leverages intra-domain feature decoupling and intra-domain context interaction to enhance the representational and discriminative power of features through contextual information.
- We propose DISM that dynamically refines matching weights by incorporating historical residual information from multiple rounds of pose estimation, significantly improving registration accuracy and robustness.

## Related Work

### 3D Feature Descriptors

Feature-based matching establishes stable point correspondences by comparing feature similarities between point clouds, without requiring prior pose information. In recent years, learning-based 3D descriptors have significantly outperformed traditional handcrafted methods. PPFNet (Deng, Birdal, and Ilic 2018b) and PPF-FoldNet (Deng, Birdal, and Ilic 2018a) utilize point pair features (PPF) to represent local structures and learn high-dimensional features through neural networks; yet these approaches still rely on handcrafted geometric relationships, limiting their adaptability to complex transformations such as rotations. To improve robustness under rotational variations, SpinNet (Ao et al. 2021) maps local surfaces to cylindrical coordinates and extracts rotation-invariant local features using a 3D cylindrical convolutional network. Nevertheless, the expressive capability of cylindrical parameterization is limited, making it difficult to adapt to diverse geometric structures. YOHO (Wang et al. 2022) and RoReg (Wang et al. 2023) leverage group convolution based on the icosahedral group to construct rotation-equivariant and rotation-invariant feature descriptors. Due to the requirement of extracting features under 60 different ro-

tations in the group structure, the overall computational cost is extremely high, and the method is less adaptable to non-uniformly distributed point clouds. Furthermore, RoITR (Yu et al. 2023a) builds pose-invariant local geometric features and aggregates them via multiple attention layers to achieve rotation-invariant representations, but the use of multiple attention layers leads to high computational and memory overhead. PARE-Net (Yao et al. 2024) employs position-aware rotation-equivariant convolution, explicitly incorporating spatial geometric information and rotation equivariance to enhance feature distinctiveness and robustness; however, it primarily focuses on modeling local neighborhood relationships, with limited capability for capturing global structural information across the entire point cloud.

Although existing methods have achieved notable progress in feature representation and robustness, most remain confined to Euclidean neighborhood modeling and fail to capture global structural relationships in complex scenarios. To overcome this limitation, this paper focuses on collaborative modeling across multiple spatial domains. By integrating geometric, feature, and global structural information, the proposed approach enhances structural perception and global consistency in point cloud matching, thereby significantly improving registration robustness.

### 3D Transformation Estimators

During the matching process between point pairs, the estimated correspondences often contain noise and outliers. As a result, a robust 3D transformation estimator is essential. The classical RANSAC algorithm (Fischler and Bolles 1981) is widely used due to its simplicity, but it suffers from limitations such as low accuracy and the need for many iterations to converge, which hinder its effectiveness in high-precision, real-time scenarios. DCP (Wang and Solomon 2019) addresses this by using differentiable singular value decomposition (SVD) to estimate transformations based on soft correspondences derived from feature similarity. Its one-shot assignment strategy remains sensitive to ambiguous or noisy matches. More recently, coarse-to-fine correspondence learning methods have gained attention for their ability to refine matches in stages and improve robustness. CoFiNet (Yu et al. 2021) introduces both self-attention and cross-attention layers in the coarse matching stage to learn descriptor correspondences, followed by optimal transport for fine matching. The matching process primarily relies on feature similarity and lacks geometric structural awareness. As a result, in scenarios with highly repetitive or symmetric structures, it is prone to significant matching ambiguity and a large number of outlier correspondences. GeoTrans (Qin et al. 2022) enhances local-to-global registration and pose estimation by embedding geometric structure into the self-attention mechanism to improve feature representation. However, when structural distinctiveness is weak or the data is noisy, the method struggles to filter out incorrect correspondences, which compromises registration robustness.

However, the aforementioned methods typically rely on correspondences generated in a single pass for pose estimation, making it difficult to effectively eliminate noise and outliers. In contrast, the strategy proposed in this paper adopts an iterative optimization approach. By incorporating a history-aware weight update mechanism, the method progressively refines the inlier set, leading to more robust and accurate 3D transformation estimation.

## Methodology

Given two partially overlapping point clouds $P = \{p_i \in \mathbb{R}^3 \mid i = 1, \ldots, N\}$ and $Q = \{q_j \in \mathbb{R}^3 \mid j = 1, \ldots, M\}$, the point cloud registration problem seeks the optimal rigid transformation that minimizes the weighted sum of point-to-point errors of a predicted correspondence set $C$, where each correspondence $(p_k, q_k)$ is associated with a confidence weight $w_k$. The transformation can be computed as follows:

$$\min_{R,t} \sum_{(p_k,q_k)\in C} w_k \|Rp_k + t - q_k\|_2^2, \tag{1}$$

where $\|\cdot\|_2$ denotes the Euclidean norm, and $R \in \mathrm{SO}(3)$ and $t \in \mathbb{R}^3$ represent the rotation and translation between $P$ and $Q$, respectively.

### Graph Neighborhood Aggregation

To address the limitations of geometric neighborhood construction in capturing structural and semantic consistency in point clouds, we propose GNAM, which constructs a global graph using optimized graph convolutions and further performs adaptive neighborhood aggregation based on the constructed global graph.

**Global Graph Construction.** Specifically, Let $\mathcal{F}^F$ denote the embedded features of each point generated by the feature embedding layer. These features are then fed into a prediction layer for probabilistic modeling, resulting in a local probability set $\mathcal{W}_l$. We then define the global graph $G_g = \{V_g, E_g\}$, where $\mathcal{W}_l$ serves as the node set $V_g$ of the global graph, and the edge set $E_g$ consists of directed connections between each node and its spatial neighbors. Based on these node features, we construct a weighted adjacency matrix $\tilde{A}$ via an outer product operation—i.e., using the correlation of $\mathcal{W}_l$ to measure relationships between nodes in the graph—and incorporate an identity matrix $\mathbf{I}$ to preserve self-connections. These operations can be expressed as:

$$\mathcal{W}_l = \mathrm{ReLU}(\tanh(\mathrm{MLP}(\mathcal{F}^F))), \tag{2}$$

$$\tilde{A} = \mathcal{W}_l \mathcal{W}_l^\top + \mathbf{I}, \tag{3}$$

where MLP$(\cdot)$ is a linear layer used to reduce the channel dimension to 1. The activation functions $\tanh(\cdot)$ and ReLU$(\cdot)$ are used to compute the weights; a higher weight indicates a stronger correlation between connected points. Finally, we adopt a spectral graph convolutional layer (Zhao et al. 2021; Kipf and Welling 2016) as the global graph relationship modeler, yielding the global graph:

$$F_g = \sigma(L\mathcal{F}^F), \tag{4}$$

where $L = \tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}$ is the graph Laplacian used to project the original features into the spectral domain. $\tilde{D} = \mathrm{diag}\left(\sum_j \tilde{A}_{ij}\right)$ denotes the degree matrix of $\tilde{A}$, and $\sigma$ represents the ReLU$(\cdot)$ activation function.

**Adaptive Neighborhood Aggregation.** After constructing the global graph, it is necessary to effectively mine consensus information within neighborhood. Specifically, for each node $p_i$, we construct a local neighborhood cluster $\mathcal{G}_i = \{p_j\}_{j=1}^{K}$ in global graph based on the similarity principle of K-nearest neighbors. Standard pooling operators are typically used to aggregate features within the neighborhood cluster. However, such operations may discard important underlying relationships between graph nodes. To address this, we aggregate neighborhood information in the global graph using adaptive neighborhood weights:

$$\alpha_{ij} = \text{Softmax}(\text{MLP}(Concat[F_i, F_i - F_{ij}])), \quad (5)$$

$$\mathcal{F}_i^G = \sum_{p_j \in \mathcal{G}_i} \alpha_{ij} F_{ij}, \quad (6)$$

where $F_i$ and $F_{ij}$ are the feature representations of the central point $p_i$ and its $j$-th spatial neighbor in the neighborhood cluster, respectively. The edge feature $Concat[F_i, F_i - F_{ij}]$ of the neighborhood cluster is first passed through an MLP, followed by a Softmax function to compute the neighborhood attention weights $\alpha_{ij}$. The neighborhood features $F_{ij}$ are then weighted and summed using these attention weights to generate the computed aggregated feature $\mathcal{F}_i^G$ for the central point $i$. The final feature representation of the global graph is composed of the aggregated features $\mathcal{F}_i^G$ for all points, denoted as $\mathcal{F}^G$. This mechanism enables each node to effectively perceive features of surrounding points that are structurally consistent with it by adaptively aggregating neighborhood information in the global graph. As a result, it enhances both the structural consistency and discriminative power of the feature representation.

## Progressive Context Interaction

To better exploit local-global contextual dependencies across multiple domains, we propose PCIM that enables contextual interaction at both intra-domain and inter-domain levels. At the intra-domain level, features are decomposed into components aligned with global features and residual components, allowing for the decoupling and joint modeling of local and global information. At the inter-domain level, we employ an inter-domain cross-attention to fuse contextual information from different domains, thereby further enhancing the representational capacity of features.

**Intra-Domain Feature Decoupling.** Let $\{\mathcal{F}^C, \mathcal{F}^F, \mathcal{F}^G\}$ denote the three feature sets from the coordinate space, feature space, and global graph, respectively, and let $\mathbf{C}_i$ represent a feature vector in a specific space domain. To construct global contextual information, we first perform global average pooling over all point features to obtain a global feature representation $\mathbf{C}_g$. To incorporate global context while preserving local structural information, we explicitly decompose each point's feature into a projection component aligned with the global feature and a residual component, thereby enabling structural decoupling and collaborative modeling of local and global information:

$$\mathbf{C}_i^{\text{proj}} = \frac{\mathbf{C}_i \cdot \mathbf{C}_g}{\|\mathbf{C}_g\|^2} \mathbf{C}_g, \quad (7)$$

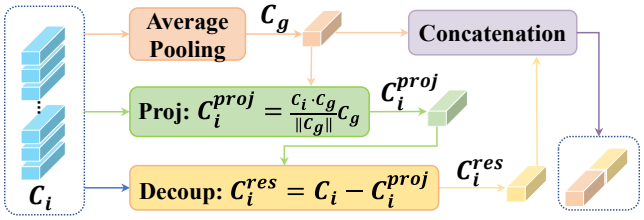


Figure 3: The proposed intra-domain feature decoupling. Features are decomposed into global-aligned and residual parts, and the residual is then concatenated with the global feature to capture both local details and global context.

$$\mathbf{C}_i^{\text{res}} = \mathbf{C}_i - \mathbf{C}_i^{\text{proj}}. \quad (8)$$

Here, $\mathbf{C}_i^{\text{proj}}$ denotes the projection component of the local feature onto the global feature, while $\mathbf{C}_i^{\text{res}}$ represents the residual component that captures local information independent of the global context. We then concatenate the residual component $\mathbf{C}_i^{\text{res}}$ with the global feature $\mathbf{C}_g$ to form the decoupled feature for the given space domain, constructing a unified representation that integrates both local structural integrity and global contextual awareness. The overall structure is illustrated in Figure 3. Finally, we adopt a parallel architecture to independently model the features of each space domain, with the resulting features represented as $\{\mathbf{C}^C, \mathbf{C}^F, \mathbf{C}^G\} \in \mathbb{R}^{N \times d}$.

**Inter-Domain Context Interaction.** Since features from different domains capture diverse structural relationships and contextual information within the point cloud, we adopt a grouped approach to enable information complementarity and collaborative fusion across multi-domain features. As shown in Figure 4, we construct three parallel inter-domain cross-attention branches, each using one domain feature as the core while leveraging the other two to compute cross-attention. This enables selective information enhancement for the dominant feature. Taking the first branch as an example, its update process is formulated as:

$$\Psi = \text{Softmax}\left(\phi_Q(\mathbf{C}^F) \cdot \phi_K(\mathbf{C}^G)^\top\right), \quad (9)$$

$$\mathbf{Y}^C = \mathbf{C}^C + \lambda \cdot \text{MLP}\left(\Psi \cdot \phi_V(\mathbf{C}^C)\right), \quad (10)$$

where $\Psi$ denotes the attention weight matrix computed from the other two feature mappings, and $\phi_Q$, $\phi_K$, $\phi_V$ represent feature mapping modules consisting of linear layers, normalization, and activation functions. $\lambda$ is a learnable parameter initialized to 0. Similarly, the second and third branches produce the interaction features $\mathbf{Y}^F$ and $\mathbf{Y}^G$, respectively. Finally, we concatenate these three features along the channel dimension to form the overall output, achieving unified modeling of multi-neighborhood contextual information.

## Patch-to-Point Matching

To construct the matching pipeline, we adopt a coarse-to-fine strategy (Yu et al. 2021; Qin et al. 2022), which divides the point cloud matching task into two stages. First, patch-level features are used for coarse filtering of the point clouds

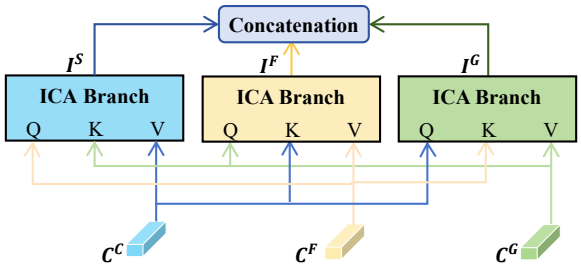


Figure 4: The proposed inter-domain context interaction. It contains three parallel inter-domain cross-attention (ICA) branches for contextual information fusion across domains.

to identify potential overlapping regions, significantly reducing the search space for matching. Then, within the filtered candidate regions, fine-grained matching relationships are further estimated based on point-level features.

**Patch Matching.** For the source point cloud $P$ and the target point cloud $Q$, initial patch-level features $\hat{\mathbf{X}}^{(P)}$ and $\hat{\mathbf{X}}^{(Q)}$ are first extracted using the feature pyramid backbone. These features are then further enhanced by the multi-domain context integration module to obtain $\hat{\mathbf{F}}^{(P)}$ and $\hat{\mathbf{F}}^{(Q)}$, respectively. We use the GeoTrans method (Qin et al. 2022) to process these two sets of features, along with positional information, by passing them through alternating self-attention and cross-attention layers, thereby generating the mixed features $\hat{\mathbf{H}}^{(P)}$ and $\hat{\mathbf{H}}^{(Q)}$. A Gaussian correlation matrix $S_{m,n}$ is then computed for the mixed features. After normalization and top-$k$ selection, we obtain the patch correspondence set $\hat{C} = \left\{(\hat{\mathbf{p}}_i, \hat{\mathbf{q}}_j) \mid (i,j) \in \text{Topk}_{m,n}\left(S_{m,n}\right)\right\}$.

**Point Matching.** After obtaining the patch correspondences $\hat{C} = (\hat{\mathbf{p}}_i, \hat{\mathbf{q}}_j)$, we construct a similarity matrix based on the point-level features $\mathbf{F}^{\hat{P}_i}$ and $\mathbf{F}^{\hat{Q}_j}$ within the corresponding patch regions $(G_{\hat{P}_i}, G_{\hat{Q}_j})$:

$$M_{i,j} = \frac{\mathbf{F}^{\hat{P}_i} \cdot \mathbf{F}^{\hat{Q}_j}}{\sqrt{\sum_{\hat{k}=1}^{\hat{d}} \left(\mathbf{F}_{\hat{k}}^{\hat{P}_i}\right)^2} \cdot \sqrt{\sum_{\hat{k}=1}^{\hat{d}} \left(\mathbf{F}_{\hat{k}}^{\hat{Q}_j}\right)^2}}, \tag{11}$$

where $\hat{d}$ denotes the feature dimension. Then, the similarity matrix $M_{i,j}$ is normalized using the Sinkhorn algorithm (Sinkhorn and Knopp 1967) to obtain a soft assignment matrix $Z_i$. To ensure bidirectional reliability between point pairs, we retain only those point pairs that rank within the top-$k$ scores in both the rows and columns of $Z_i$, forming the point matching set $\tilde{C}_i$ within the current patch. Finally, the local results of all $N_c$ patch pairs are aggregated into the global point-level matching set $\tilde{C} = \bigcup_{i=1}^{N_c} \tilde{C}_i$.

**Loss Function.** Following (Yao et al. 2024), we employs three loss functions to jointly optimize network training: the patch matching loss $\mathcal{L}_{\text{patch}}$, the point matching loss $\mathcal{L}_{\text{point}}$, and the contrastive rotation loss $\mathcal{L}_{\text{rot}}$. The overall loss is defined as $\mathcal{L} = \mathcal{L}_{\text{patch}} + \mathcal{L}_{\text{point}} + \mathcal{L}_{\text{rot}}$.

## Dynamic Inlier Selection

Accurate pose estimation requires identifying reliable correspondences under ambiguous structures. To this end, we propose DISM to initialize weights jointly based on matching confidence and geometric consistency, and iteratively refine them using residual feedback.

**Weight Initialization.** Due to geometric similarities between local regions, multiple candidate matches with similar features often emerge, leading to matching ambiguities. To mitigate this, we introduce geometric compatibility as a simple yet effective consistency measure, based on the principle that the distance between point pairs should remain consistent before and after transformation (Bai et al. 2021; Lee et al. 2021). For each point-wise correspondence $(\tilde{\mathbf{p}}_i, \tilde{\mathbf{q}}_j)$ established in the point matching set $\tilde{C}$, the initial weight $w_{ij}^{(0)}$ is determined jointly by the matching confidence and the geometric compatibility matrix:

$$w_{ij}^{(0)} = \beta_{ij} \cdot \gamma_{ij}, \beta_{ij} = \max\left(0, 1 - \frac{d_{ij}^2}{\varepsilon^2}\right), \tag{12}$$

where $\gamma_{ij}$ denotes the confidence of correspondence. The geometric compatibility matrix $\beta_{ij}$ measures the geometric consistency between matched correspondences. $d_{ij}$ represents the distance difference between the source and target correspondence, and $\varepsilon$ is a threshold parameter that controls the tolerance for length differences.

**History-Aware Iterative Pose Estimation.** To improve the robustness of pose estimation, we design an iterative optimization strategy incorporating a history-aware mechanism. In each iteration, we perform weighted SVD based on the current weighted consensus set to update the rigid transformation parameters. The weights of the correspondences are then adjusted according to the current rotation residuals, thereby reducing the influence of potential mismatches. Specifically, let $\hat{R}^{(n)}$ denote the rotation estimate obtained at the $n$-th iteration. The rotation residual between two consecutive estimates is expressed as:

$$\theta^{(n)} = \text{Angle}\left(\hat{R}^{(n-1)}, \hat{R}^{(n)}\right), \tag{13}$$

where $\text{Angle}(\cdot,\cdot)$ measures the angular difference between two rotation matrices, indicating the degree of change in the estimation. Since early iterations may produce unstable estimates, we introduce an iteration-dependent historical weighting mechanism to control the influence of each residual on the final weight update. The weight at the $n$-th iteration is updated as:

$$w_{ij}^{(n)} = w_{ij}^{(0)} \cdot \prod_{k=1}^{n} \exp\left(-E(k) \cdot \theta^{(k)}\right), \tag{14}$$

where $E(k) = 2k/(m(m+1))$ is a normalized historical weight function that determines the influence of each residual, with $m$ being the total number of iterations. The final transformation is represented by $T = \{R, t\}$, based on the pose estimate from the last iteration. This strategy incorporates multi-round residual information and mitigates the impact of unstable early estimates, thereby improving the accuracy and robustness of pose estimation.

| Samples | 3DMatch 5000 | 2500 | 1000 | 500 | 250 | 3DLoMatch 5000 | 2500 | 1000 | 500 | 250 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Registration Recall (RR, %)** | | | | | | | | | | |
| FCGF | 85.1 | 84.7 | 83.3 | 81.6 | 71.4 | 40.1 | 41.7 | 38.2 | 35.4 | 26.8 |
| Predator | 89.0 | 89.9 | 90.6 | 88.5 | 86.6 | 59.8 | 61.2 | 62.4 | 60.8 | 58.1 |
| YOHO | 90.8 | 90.3 | 89.1 | 88.6 | 84.5 | 65.2 | 65.5 | 63.2 | 56.5 | 48.0 |
| GeoTrans | 92.0 | 91.8 | 91.8 | 91.4 | 91.2 | 75.0 | 74.8 | 74.2 | 74.1 | 73.5 |
| RoITr | 91.9 | 91.7 | 91.8 | 91.4 | 91.0 | 74.7 | 74.8 | 74.8 | 74.2 | 73.6 |
| PEAL | 94.4 | 94.1 | 94.1 | 93.9 | 93.4 | 79.2 | 79.0 | 78.8 | 78.5 | 77.9 |
| SIRA-PCR | 93.6 | 93.9 | 93.9 | 92.7 | 92.4 | 73.5 | 73.9 | 73.0 | 73.4 | 71.1 |
| Diff-Reg | | | 95.0 | | | | | 73.8 | | |
| DCATr | 92.6 | 92.4 | 92.2 | 91.9 | 91.6 | 76.8 | 76.4 | 75.7 | 75.1 | 73.7 |
| Cross-PCR | 94.5 | 94.2 | 94.2 | 94.3 | 94.0 | 73.7 | 73.9 | 74.1 | 74.2 | 74.1 |
| PSReg | 95.7 | 94.9 | 95.1 | 95.0 | 95.2 | 79.3 | 79.3 | 78.7 | 78.7 | 78.4 |
| MCI-Net | **96.4** | **96.4** | **96.6** | **96.5** | **96.6** | **79.7** | **79.7** | **79.9** | **79.6** | **79.5** |
| **Inlier Ratio (IR, %)** | | | | | | | | | | |
| FCGF | 56.8 | 54.1 | 48.7 | 42.5 | 34.1 | 21.4 | 20.0 | 17.2 | 14.8 | 11.6 |
| Predator | 58.0 | 58.4 | 57.1 | 54.1 | 49.3 | 26.7 | 28.1 | 28.3 | 27.5 | 25.8 |
| YOHO | 64.4 | 60.7 | 55.7 | 46.4 | 41.2 | 25.9 | 23.3 | 22.6 | 18.2 | 15.0 |
| GeoTrans | 71.9 | 75.2 | 76.0 | 82.2 | 85.1 | 43.5 | 45.3 | 46.2 | 52.9 | 57.7 |
| RoITr | 82.6 | 82.8 | 83.0 | 83.0 | 83.0 | 54.3 | 54.6 | 55.1 | 55.2 | 55.3 |
| PEAL | 74.8 | 81.3 | 86.0 | 87.9 | 89.2 | 49.1 | 54.1 | 60.5 | 63.6 | 65.0 |
| SIRA-PCR | 70.8 | 78.3 | 83.7 | 85.9 | 87.4 | 43.3 | 49.0 | 55.9 | 58.8 | 60.7 |
| Diff-Reg | | | 30.9 | | | | | 9.6 | | |
| DCATr | 87.6 | 86.5 | 84.7 | 81.0 | 76.5 | 62.1 | 60.3 | 57.9 | 53.3 | 48.4 |
| Cross-PCR | 88.7 | 88.7 | 88.7 | 88.7 | 88.7 | 65.9 | 65.9 | 65.9 | 65.9 | 65.9 |
| PSReg | 75.8 | 82.4 | 87.1 | 88.9 | 90.0 | 49.9 | 55.5 | 61.9 | 64.5 | 66.3 |
| MCI-Net | **88.8** | **88.9** | **89.8** | **90.6** | **91.1** | **66.0** | **66.1** | **67.2** | **68.0** | **68.7** |
| **Feature Matching Recall (FMR, %)** | | | | | | | | | | |
| FCGF | 97.4 | 97.3 | 97.0 | 96.7 | 96.6 | 76.6 | 75.4 | 74.2 | 71.7 | 67.3 |
| Predator | 96.6 | 96.6 | 96.5 | 96.3 | 96.5 | 78.6 | 77.4 | 76.3 | 75.7 | 75.3 |
| YOHO | 98.2 | 97.6 | 97.5 | 97.7 | 96.0 | 79.4 | 78.1 | 76.3 | 73.8 | 69.1 |
| GeoTrans | 97.9 | 97.9 | 97.9 | 97.9 | 97.6 | 88.3 | 88.6 | 88.8 | 88.6 | 88.3 |
| RoITr | 98.0 | 98.0 | 97.9 | 98.0 | 97.9 | **89.6** | **89.6** | **89.5** | **89.4** | **89.3** |
| PEAL | 98.4 | 98.4 | 98.4 | 98.4 | 98.4 | 87.7 | 87.8 | 87.8 | 88.0 | 87.4 |
| SIRA-PCR | 98.2 | 98.4 | 98.4 | 98.5 | 98.5 | 88.8 | 89.0 | 88.9 | 88.6 | 87.7 |
| Diff-Reg | | | 96.2 | | | | | 69.6 | | |
| DCATr | 98.2 | 98.3 | 98.4 | 98.0 | 98.1 | 87.7 | 87.5 | 87.7 | 87.2 | 87.4 |
| Cross-PCR | 97.9 | 97.7 | 97.9 | 97.7 | 97.7 | 83.5 | 83.3 | 83.1 | 83.1 | 83.4 |
| PSReg | **98.6** | **98.6** | **98.6** | **98.6** | **98.6** | 86.4 | 86.6 | 87.1 | 87.4 | 87.1 |
| MCI-Net | 98.4 | 98.4 | 98.4 | 98.4 | 98.5 | 85.1 | 85.1 | 85.1 | 85.1 | 85.0 |

Table 1: Evaluation results on 3DMatch and 3DLoMatch. The best and second-to-best results of baseline methods are respectively marked in bold and underlined.

| Methods | RRE (°) | RTE (cm) | RR (%) | Time (s) |
|---|---|---|---|---|
| FCGF | 0.30 | 9.5 | 96.6 | **0.18** |
| D3Feat | 0.30 | 7.2 | **99.8** | - |
| Predator | 0.27 | 6.8 | **99.8** | 0.69 |
| SpinNet | 0.47 | 9.9 | 99.1 | 14.57 |
| CoFiNet | 0.41 | 8.2 | **99.8** | 0.65 |
| BUFFER | 0.26 | 7.1 | **99.8** | 0.30 |
| GeoTrans | **0.23** | 6.2 | **99.8** | 0.31 |
| MCI-Net | **0.23** | **5.0** | **99.8** | 0.23 |

Table 2: Evaluation results on KITTI Odometry. The best and second-to-best results of baseline methods are respectively marked in bold and underlined.

## Experiments

In this section, we evaluate the proposed method against various advanced methods on multiple benchmark datasets, including indoor RGB-D point cloud datasets 3DMatch (Zeng et al. 2017) and 3DLoMatch (Huang et al. 2021), and outdoor LiDAR point cloud dataset KITTI Odometry (Geiger, Lenz, and Urtasun 2012). In addition, we have performed comprehensive ablation studies. All experiments are conducted on Ubuntu 18.04 with a single Nvidia RTX3090.

### Indoor Scenarios: 3DMatch and 3DLoMatch

**Dataset.** The proposed MCI-Net is evaluated on the indoor point cloud registration benchmark datasets 3DMatch (Zeng et al. 2017) and 3DLoMatch. According to (Huang et al. 2021), 3DMatch consists of point cloud pairs with overlap ratios above 30%, while 3DLoMatch contains pairs with overlap ratios ranging from 10% to 30%.

**Metrics.** Following (Bai et al. 2020; Qin et al. 2022), we report inlier ratio (IR), feature matching recall (FMR), and registration recall (RR) as evaluation metrics.

**Registration Results.** In Table 1, we compare the correspondence results of MCI-Net with those of other recent methods, such as FCGF (Choy, Park, and Koltun 2019), Predator (Huang et al. 2021), YOHO (Wang et al. 2022), GeoTrans (Qin et al. 2022), RoITr (Yu et al. 2023a), PEAL (Yu et al. 2023b), SIRA-PCR (Chen et al. 2023), Diff-Reg (Wu et al. 2024), DCATr (Chen et al. 2024), Cross-PCR (Zhao et al. 2025), and PSReg (Huang et al. 2025), under varying numbers of correspondences (i.e., 5000, 2500, 1000, 500 and 250). Compared to these methods, MCI-Net demonstrates significant improvements in RR and IR on both the 3DMatch and 3DLoMatch datasets. In terms of RR, this metric directly reflects the registration success rate and is often considered the most critical metric. MCI-Net consistently achieves the highest RR under varying sampling densities, owing to its ability to learn contextual information from multiple domain spaces, effectively addressing the limitations of methods that rely solely on geometric positions for feature extraction. In terms of IR, MCI-Net outperforms all existing methods across all sampling density settings. This improvement is largely attributed to our DISM, which dynamically increases the weights of high-confidence inliers during the iterative process, preserving true correspondences while suppressing noise and mismatches. In terms of FMR, MCI-Net performs comparable to PSReg and PEAL on the 3DMatch dataset. It is important to note that PSReg and PEAL require additional prior information about overlap, whereas MCI-Net operates without any prior information. However, in extremely low-overlap scenarios (e.g., 3DLoMatch), the sparse nature of the data makes it challenging to extract sufficient structural cues and reliable inliers, which results in a decrease in the FMR metric for MCI-Net.

**Qualitative Results.** We present qualitative results in Figure 5, where our results are compared with GeoTrans (Qin et al. 2022). The results show that MCI-Net achieves higher inlier ratios (IR) for correspondences on both the 3DMatch and low-overlap 3DLoMatch datasets, and demonstrates more accurate registration results.

### Outdoor Scenarios: KITTI Odometry

**Dataset.** KITTI Odometry (Geiger, Lenz, and Urtasun 2012) is a widely used outdoor benchmark dataset captured by LiDAR sensors for autonomous driving applications. Following (Bai et al. 2020; Qin et al. 2022), we use sequences 0–5 for training, 6–7 for validation, and 8–10 for testing.

**Metrics.** Following prior work (Huang et al. 2021), we employ three metrics to evaluate: relative rotation error (RRE),

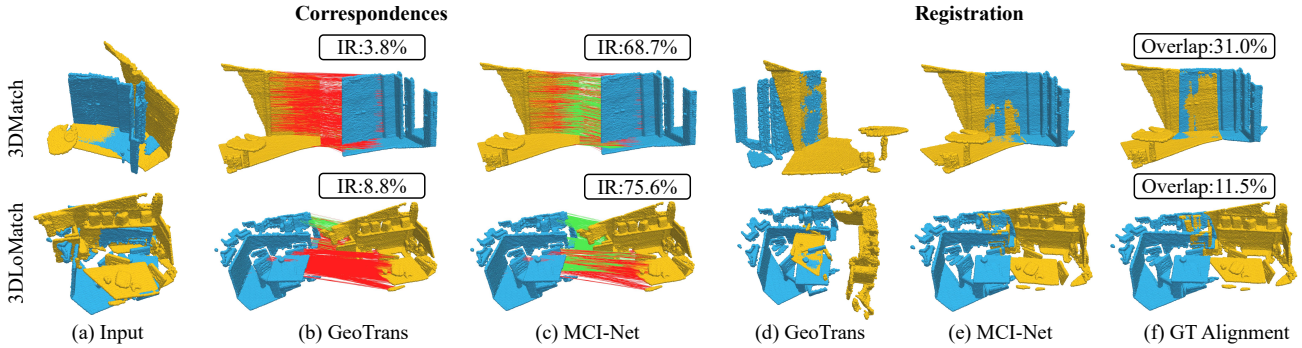


Figure 5: Qualitative results on 3DMatch and 3DLoMatch datasets. Columns (b) and (c) show the correspondences, while columns (d) and (e) demonstrate the registration results. Green/red lines indicate inliers/outliers.

| No. | Methods | 3DMatch | | | 3DLoMatch | | |
|---|---|---|---|---|---|---|---|
| | | RR | IR | FMR | RR | IR | FMR |
| (1) | Full Model (Ours) | **96.4** | **88.8** | **98.4** | **79.7** | **65.9** | **85.1** |
| (2) | w/o GNAM | 95.2 | 87.9 | 98.0 | 78.9 | 65.8 | 85.0 |
| (3) | w/o IFD | 95.3 | 88.1 | 98.1 | 79.3 | 65.8 | **85.1** |
| (4) | w/o ICI | 95.7 | 88.3 | 98.3 | 79.5 | 65.7 | 85.0 |

Table 3: Ablation studies of MCI-Net.

| Pose Estimator | 3DMatch | | | 3DLoMatch | | |
|---|---|---|---|---|---|---|
| | RR | RRE | RTE | RR | RRE | RTE |
| RANSAC | 94.6 | 2.29 | 0.072 | 78.8 | 3.51 | 0.103 |
| LGR | 94.9 | 1.88 | 0.060 | 79.1 | 2.88 | 0.086 |
| DISM (Ours) | **96.4** | **1.72** | **0.054** | **79.8** | **2.63** | **0.075** |

Table 4: Ablation studies of pose estimator.

relative translation error (RTE) and registration recall (RR).

**Registration Results.** We evaluate our performance by comparing experimental results with recent methods, including FCGF (Choy, Park, and Koltun 2019), D3Feat (Bai et al. 2020), Predator (Huang et al. 2021), SpinNet (Ao et al. 2021), CoFiNet (Yu et al. 2021), BUFFER (Ao et al. 2023), and GeoTrans (Qin et al. 2022). The quantitative results are shown in Table 2 and demonstrate that MCI-Net achieves leading performance on the KITTI Odometry dataset. Specifically, MCI-Net achieves the highest RR, as well as the lowest RRE and RTE among all compared methods. Furthermore, MCI-Net delivers outstanding runtime performance. Although FCGF achieves slightly faster runtime than MCI-Net, its RR is 3.2% lower than MCI-Net. The experimental results demonstrate that MCI-Net generalizes well to outdoor datasets.

## Ablation Studies

**Graph Neighborhood Aggregation Module (GNAM):** As shown in Table 3 (2), we remove GNAM results in a noticeable drop in performance. This validates the effectiveness of our approach in enhancing global structural awareness and improving registration robustness.

**Intra-Domain Feature Decoupling (IFD):** As shown in Table 3 (3), we remove the intra-domain feature decoupling module, which limits the ability of the model to fully exploit the synergy between local structure and global context. This results in limited feature representation and a substantial decline in registration accuracy.

**Inter-Domain Context Interaction (ICI):** We replace the intra-domain cross-attention with independent self-attention in each domain, thereby removing explicit fusion among features from different domains. As shown by the performance drop in Table 3 (4), this ablation demonstrates that employing intra-domain cross-attention to fuse contextual information from different domains further enhances feature representation and registration robustness.

**Pose Estimator:** To demonstrate the superiority of our proposed pose estimator, we compare our DISM with two classical baselines: RANSAC (Fischler and Bolles 1981) and LGR (Qin et al. 2022). As shown in Table 4, MCI-Net achieves consistently competitive performance across all evaluation metrics on both the 3DMatch and 3DLoMatch datasets. These results validate that, compared to traditional pose estimation methods, the proposed method offers significantly better robustness and accuracy.

## Conclusion

In this paper, we propose MCI-Net, the multi-domain context integration network designed to enhance feature representation and registration performance. To capture the overall structural relationships of point clouds, we propose GNAM within the constructed global graph. To enhance the contextual representation and discriminative capability of features, we propose PCIM, which decouples intra-domain features and fuses inter-domain contexts across spatial scales. Moreover, we propose DISM that iteratively refines inlier weights by incorporating historical residuals from pose estimation, effectively enhancing registration reliability and robustness. Extensive experiments on indoor and outdoor benchmarks show that MCI-Net significantly outperforms recent competing methods. However, in extremely low-overlap scenarios, MCI-Net still faces certain limitations in registration performance due to the scarcity of structural cues and reliable inliers. In future work, we plan to further explore its potential and scalability in large-scale point cloud scenarios and more diverse processing tasks.

## Acknowledgments

This work was supported in part by National Natural Science Foundation of China (Nos. U22A2095, 62476112, 62332007, U22B2028, 62501343); in part by Guangdong Basic and Applied Basic Research Foundation (Nos. 2024A1515011740, 2025A1515010181); in part by Natural Science Foundation of Shandong Province (No. ZR2024QF294); in part by Fundamental Research Funds for the Central Universities (Nos. 21624404, 23JNSYS01); in part by Science and Technology Major Project of Tibetan Autonomous Region of China (No. XZ202201ZD0006G), National Joint Engineering Research Center of Network Security Detection and Protection Technology, Guangdong Key Laboratory of Data Security and Privacy Preserving, Guangdong Hong Kong Joint Laboratory for Data Security and Privacy Protection, Engineering Research Center of Trustworthy AI, Ministry of Education, and Guangdong Key Laboratory of Intelligent Information Processing, Shenzhen University.

## References

Ao, S.; Hu, Q.; Wang, H.; Xu, K.; and Guo, Y. 2023. Buffer: Balancing accuracy, efficiency, and generalizability in point cloud registration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1255–1264.

Ao, S.; Hu, Q.; Yang, B.; Markham, A.; and Guo, Y. 2021. Spinnet: Learning a general surface descriptor for 3d point cloud registration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 11753–11762.

Bai, X.; Luo, Z.; Zhou, L.; Chen, H.; Li, L.; Hu, Z.; Fu, H.; and Tai, C.-L. 2021. Pointdsc: Robust point cloud registration using deep spatial consistency. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 15859–15869.

Bai, X.; Luo, Z.; Zhou, L.; Fu, H.; Quan, L.; and Tai, C.-L. 2020. D3feat: Joint learning of dense detection and description of 3d local features. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 6359–6367.

Chen, H.; Yan, P.; Xiang, S.; and Tan, Y. 2024. Dynamic cues-assisted transformer for robust point cloud registration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 21698–21707.

Chen, S.; Xu, H.; Li, R.; Liu, G.; Fu, C.-W.; and Liu, S. 2023. Sira-pcr: Sim-to-real adaptation for 3d point cloud registration. In *Proceedings of the IEEE International Conference on Computer Vision*, 14394–14405.

Choy, C.; Park, J.; and Koltun, V. 2019. Fully convolutional geometric features. In *Proceedings of the IEEE International Conference on Computer Vision*, 8958–8966.

Dai, L.; Liu, Y.; Ma, J.; Wei, L.; Lai, T.; Yang, C.; and Chen, R. 2022. MS2DG-Net: Progressive correspondence learning via multiple sparse semantics dynamic graph. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 8973–8982.

Deng, H.; Birdal, T.; and Ilic, S. 2018a. Ppf-foldnet: Unsupervised learning of rotation invariant 3d local descriptors. In *Proceedings of the European Conference on Computer Vision*, 602–618.

Deng, H.; Birdal, T.; and Ilic, S. 2018b. Ppfnet: Global context aware local features for robust 3d point matching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 195–205.

Ermolov, A.; Mirvakhabova, L.; Khrulkov, V.; Sebe, N.; and Oseledets, I. 2022. Hyperbolic vision transformers: Combining improvements in metric learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7409–7419.

Fan, Z.; Song, Z.; Liu, H.; Lu, Z.; He, J.; and Du, X. 2022. Svt-net: Super light-weight sparse voxel transformer for large scale place recognition. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, 551–560.

Fischler, M. A.; and Bolles, R. C. 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6): 381–395.

Geiger, A.; Lenz, P.; and Urtasun, R. 2012. Are we ready for autonomous driving? the kitti vision benchmark suite. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3354–3361.

Hu, Y.; Ye, S.; Zhao, W.; Lin, M.; He, Y.; Wen, Y.-H.; He, Y.; and Liu, Y.-J. 2024. Oˆ 2-recon: Completing 3d reconstruction of occluded objects in the scene with a pre-trained 2D diffusion model. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 2285–2293.

Huang, S.; Gojcic, Z.; Usvyatsov, M.; Wieser, A.; and Schindler, K. 2021. Predator: Registration of 3d point clouds with low overlap. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4267–4276.

Huang, X.; Huang, Z.; Zuo, Y.; Gong, Y.; Zhang, C.; Liu, D.; and Fang, Y. 2025. PSReg: Prior-guided Sparse Mixture of Experts for Point Cloud Registration. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 3788–3796.

Kipf, T. N.; and Welling, M. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.

Lai, X.; Liu, J.; Jiang, L.; Wang, L.; Zhao, H.; Liu, S.; Qi, X.; and Jia, J. 2022. Stratified transformer for 3d point cloud segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 8500–8509.

Lee, J.; Kim, S.; Cho, M.; and Park, J. 2021. Deep hough voting for robust global registration. In *Proceedings of the IEEE International Conference on Computer Vision*, 15994–16003.

Lin, S.; Chen, X.; Xiao, G.; Wang, H.; Huang, F.; and Weng, J. 2024. Multi-stage network with geometric semantic attention for two-view correspondence learning. *IEEE Transactions on Image Processing*, 33: 3031–3046.

Lin, S.; Luo, H.; Yan, Y.; Xiao, G.; and Wang, H. 2022. Co-clustering on bipartite graphs for robust model fitting. *IEEE Transactions on Image Processing*, 31: 6605–6620.

Liu, X.; Qin, R.; Yan, J.; and Yang, J. 2024. NCMNet: Neighbor consistency mining network for two-view correspondence pruning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Lu, F.; Chen, G.; Liu, Y.; Zhang, L.; Qu, S.; Liu, S.; and Gu, R. 2021. Hregnet: A hierarchical network for large-scale outdoor lidar point cloud registration. In *Proceedings of the IEEE International Conference on Computer Vision*, 16014–16023.

Luanyuan, D.; Du, X.; Zhang, H.; and Tang, J. 2024. Mgnet: Learning correspondences via multiple graphs. In *Proceedings of the AAAI conference on Artificial Intelligence*, volume 38, 3945–3953.

Qin, Z.; Yu, H.; Wang, C.; Guo, Y.; Peng, Y.; and Xu, K. 2022. Geometric transformer for fast and robust point cloud registration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 11143–11152.

Sinkhorn, R.; and Knopp, P. 1967. Concerning nonnegative matrices and doubly stochastic matrices. *Pacific Journal of Mathematics*, 21(2): 343–348.

Wang, H.; Liu, Y.; Dong, Z.; and Wang, W. 2022. You only hypothesize once: Point cloud registration with rotation-equivariant descriptors. In *Proceedings of the ACM International Conference on Multimedia*, 1630–1641.

Wang, H.; Liu, Y.; Hu, Q.; Wang, B.; Chen, J.; Dong, Z.; Guo, Y.; Wang, W.; and Yang, B. 2023. RoReg: Pairwise point cloud registration with oriented descriptors and local rotations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(8): 10376–10393.

Wang, Y.; and Solomon, J. M. 2019. Deep closest point: Learning representations for point cloud registration. In *Proceedings of the IEEE International Conference on Computer Vision*, 3523–3532.

Wu, Q.; Jiang, H.; Luo, L.; Li, J.; Ding, Y.; Xie, J.; and Yang, J. 2024. Diff-reg: Diffusion model in doubly stochastic matrix space for registration problem. In *Proceedings of the European Conference on Computer Vision*, 160–178.

Xie, Y.; Zhu, J.; Li, S.; Hu, N.; and Shi, P. 2024. HECPG: Hyperbolic embedding and confident patch-guided network for point cloud matching. *IEEE Transactions on Geoscience and Remote Sensing*.

Yao, R.; Du, S.; Cui, W.; Tang, C.; and Yang, C. 2024. PARE-Net: Position-aware rotation-equivariant networks for robust point cloud registration. In *Proceedings of the European Conference on Computer Vision*, 287–303.

Yu, H.; Li, F.; Saleh, M.; Busam, B.; and Ilic, S. 2021. Cofinet: Reliable coarse-to-fine correspondences for robust pointcloud registration. *Proceedings of the Advances in Neural Information Processing Systems*, 34: 23872–23884.

Yu, H.; Qin, Z.; Hou, J.; Saleh, M.; Li, D.; Busam, B.; and Ilic, S. 2023a. Rotation-invariant transformer for point cloud matching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 5384–5393.

Yu, J.; Ren, L.; Zhang, Y.; Zhou, W.; Lin, L.; and Dai, G. 2023b. PEAL: Prior-embedded explicit attention learning for low-overlap point cloud registration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 17702–17711.

Zeng, A.; Song, S.; Nießner, M.; Fisher, M.; Xiao, J.; and Funkhouser, T. 2017. 3Dmatch: Learning local geometric descriptors from rgb-d reconstructions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1802–1811.

Zhang, J.; Zhou, Z.; Lu, G.; Tian, J.; and Pei, W. 2024. Robust 3d tracking with quality-aware shape completion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 7160–7168.

Zhao, C.; Ge, Y.; Zhu, F.; Zhao, R.; Li, H.; and Salzmann, M. 2021. Progressive correspondence pruning by consensus learning. In *Proceedings of the IEEE International Conference on Computer Vision*, 6464–6473.

Zhao, G.; Guo, Z.; Du, Z.; and Ma, H. 2025. Cross-PCR: A robust cross-source point cloud registration framework. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 10403–10411.