

Coarse-to-fine Point Cloud Registration Based on Superpoint Overlap Prediction

Mengchong Sun¹, Jinyu Tan¹, Yutao Zhang¹, Juntao Yang^{1,*}, Xue Zhang², Yuan Liu³, Jianzhong Chen³

¹College of Geodesy and Geomatics, Shandong University of Science and Technology, Qingdao 266590, China -
smc20010713@163.com; tjy2424904994@163.com; zhangyutao060116@outlook.com; jtyang@sdust.edu.cn

²College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, China -
xuezhong@sdust.edu.cn

³Shandong Provincial Institute of Land Surveying and Mapping, Jinan 250013, China- 1562407087@163.com; chenjz@shandong.cn

Keywords: Point cloud registration, Coarse-to-fine correspondences, Overlap prediction, Superpoint matching.

Abstract:

Point cloud registration plays an important role in 3D reconstruction and other point cloud-related tasks. The establishment of reliable and high-quality point correspondences is essential for accurately recovering the transformation matrix between point clouds. In recent years, the coarse-to-fine strategy have gained widespread attention to construct reliable point correspondences. However, in the coarse-scale superpoint matching stage, superpoints in non-overlapping regions can degrade the matching quality, thereby limiting the reliability of the refined point correspondences. To address this issue, this paper proposes a coarse-to-fine point cloud registration method based on superpoint overlap prediction, which focuses on optimizing the construction of superpoint correspondences at the coarse scale and effectively improving registration accuracy. Firstly, we employ a position-aware attention mechanism to enhance superpoint features under geometric constraints. Then, the superpoint overlap prediction module generates overlap masks based on the enhanced features, effectively filtering out superpoints over non-overlapping regions. This ensures that only the available superpoints over overlapping regions participate in the matching process, leading to more accurate superpoint correspondences and improved registration accuracy and robustness. Experimental results on indoor 3DMatch and 3DLoMatch datasets, as well as the outdoor KITTI dataset, demonstrate that our proposed method achieves the superior registration performance.

1. Introduction

With the advancement of 3D laser scanning technology, point clouds have become a critical data source in applications such as 3D reconstruction, making the processing of point cloud data essential. As a crucial step in point cloud processing, point cloud registration aims to compute a transformation matrix that aligns local point clouds from different coordinate systems into a unified coordinate system, thereby reconstructing a complete 3D scene (Yang et al., 2024). Point cloud registration has now been widely applied in various fields, including 3D reconstruction (Yang et al., 2022) and virtual reality (Nguyen et al., 2022). Therefore, the topic of point cloud registration holds significant theoretical value and has profound implications for practical applications.

In traditional point cloud registration methods, the Iterative Closest Point (ICP) algorithm (Besl and McKay, 1992) establishes point correspondences through nearest-neighbor search and uses iterative optimization to minimize the geometric error between two sets of point clouds. While the ICP algorithm is highly versatile, it suffers from issues such as sensitivity to the initial position and slow convergence. To solve the above issues, some studies have introduced branch-and-bound methods (Yang et al., 2013) or constraint strategies based on covariance matrices (Segal et al., 2009) to accelerate convergence and improve registration accuracy. However, these methods still face efficiency issues when dealing with low-overlap or dense point clouds. Unlike the traditional point cloud registration methods, researchers have devoted to classic 3D feature descriptors such as SHOT (Salti et al., 2014) and FPFH (Rusu et al., 2009) and established point correspondences by exploiting the similarity between descriptors. These 3D feature descriptor-based methods perform well in specific scenarios, effectively extracting point cloud features and performing matching. However, traditional descriptors

primarily rely on geometric information, making it difficult to capture higher-level semantic features. As a result, point correspondences constructed in complex scenes with substantial noise or repetitive structures are less reliable, leading to further degradation in registration accuracy.

In recent years, deep learning has attracted widespread attention in the point cloud registration field due to its powerful feature learning and representation capabilities. Many studies (Bai et al., 2020, Huang et al., 2020) have begun to leverage neural networks to learn semantic features of point clouds, replacing traditional hand-crafted 3D feature descriptors to construct more reliable point correspondences. Deep learning-based point cloud registration approaches can generally be classified into two categories. The first category includes keypoint detection-based methods (Zaman et al., 2023, Huang et al., 2021), which primarily rely on constructing high-discriminative feature descriptors to improve the reliability of point correspondences. For example, RoReg (Wang et al., 2023) improves the accuracy of keypoint detection and optimizes the feature matching process by introducing orientation descriptors and a local rotation estimation mechanism. This strategy enhances the reliability of point correspondences and effectively improves the accuracy of point cloud registration. Instead of keypoint detection (Yu et al., 2021, Qin et al., 2023), the second type of methods adopts a coarse-to-fine strategy to progressively establish point correspondences. These methods construct superpoint correspondences at the coarse scale, which are then refined at the fine scale, thereby avoiding the keypoint detection process and enhancing the reliability of the constructed point correspondences.

Despite the promising results achieved by lots of existing methods, the point cloud registration task still faces significant challenges for several reasons. Firstly, the introduction of attention mechanism helps superpoints to aggregate more context

* Corresponding author: jtyang@sdust.edu.cn

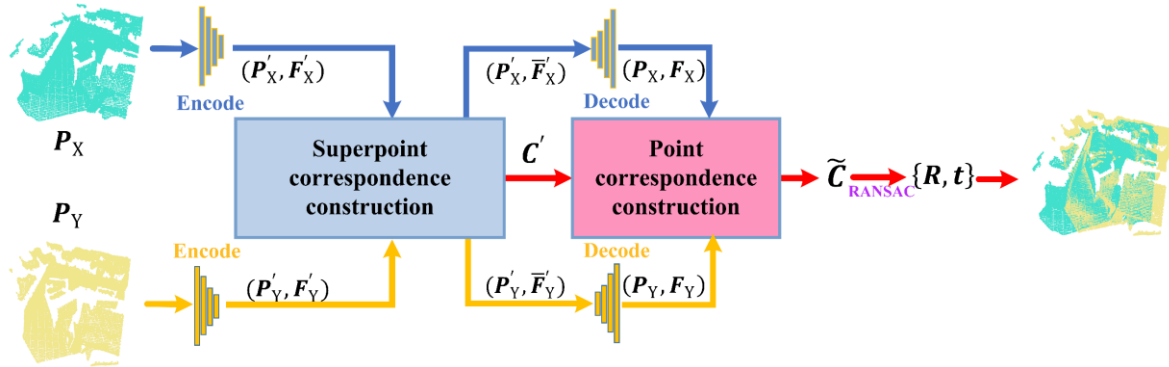


Figure 1. The pipeline of the proposed method.

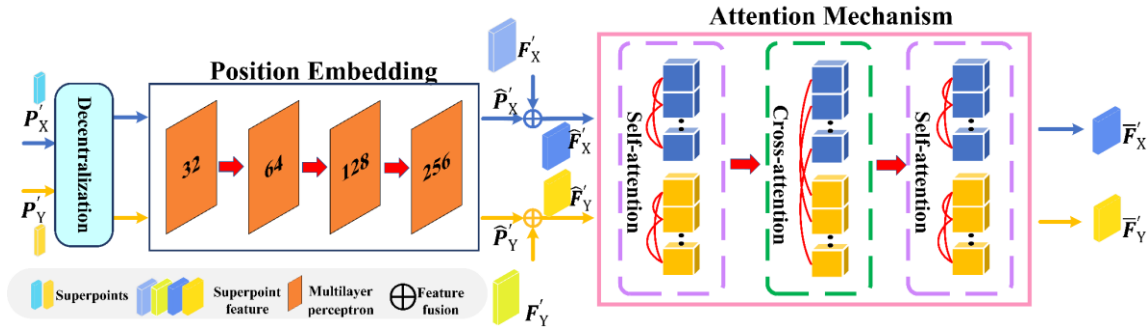


Figure 2. The pipeline of the position-aware attention.

information and enhance the expressiveness of point cloud features. However, the attention mechanism often ignores the geometric structure of the point cloud and leads to insufficient expression of the geometric information of the point cloud features, thereby increasing the risk of false correspondence. Secondly, previous researches (Yu et al., 2021) have demonstrated that a coarse-to-fine matching strategy offers significant advantages in improving the accuracy of point correspondences. However, during the coarse-scale superpoint matching phase, the process is susceptible to noise interference and non-overlapping superpoints, which weakens the quality of superpoint matching and reduces the reliability of subsequent point correspondences.

To address the aforementioned challenges, we propose a coarse-to-fine point cloud registration network based on superpoint overlap prediction. Our contributions of this paper can be summarized as follows:

- (1) We design a position-aware attention mechanism that integrates positional information into superpoint features. By using self-attention and cross-attention under geometric constraints, it effectively fuses local superpoint features with global contextual information.
- (2) In the coarse-scale superpoint matching phase, we introduce a superpoint overlap prediction (SOP) module, which uses an overlap mask to filter out non-overlapping superpoints and improve matching robustness.
- (3) Evaluation results on large-scale indoor and outdoor datasets demonstrate that our proposed method achieves superior registration performance, with a registration recall of 90.2% on indoor scenes and 99.8% on outdoor scenes.

2. Methodology

Let $P_X, P_Y \in \mathbb{R}^3$ be two point clouds with partial overlap. Our goal is to estimate a rigid transformation parameter to align P_X and P_Y into the same coordinate system, where $R \in$

$SO(3)$ and $t \in \mathbb{R}^3$. We first establish superpoint correspondences at the coarse scale (Section 2.1), then refine these correspondences during the point correspondence construction phase. Based on the refined correspondences, we estimate the rigid transformation parameters to achieve point cloud registration (Section 2.2). The overall process of the proposed method is illustrated in Figure 1.

2.1 Superpoint correspondence construction

2.1.1 Position-aware attention: To obtain superpoint features for matching, we first encode the input raw point cloud and enhance the features using position-aware attention. Specifically, we use the encoding layers of KPConv to extract features and perform downsampling on the input raw point cloud data, reducing computational complexity while preserving key information. Through this process, we obtain the superpoint features $F'_X \in \mathbb{R}^{n' \times b}$ and $F'_Y \in \mathbb{R}^{m' \times b}$, as well as the uniformly distributed superpoints $P'_X \in \mathbb{R}^{n' \times 3}$ and $P'_Y \in \mathbb{R}^{m' \times 3}$. The superpoints and their features are then fed into the position-aware attention module. By injecting positional information into the superpoint features, the network enhances its ability to capture contextual information while improving its perception of the geometric structure of the point clouds. This process is illustrated in Figure 2.

Taking the superpoints P'_X from the source point cloud as an example, the process is as follows: First, we center the superpoints to remove the effects of global spatial shifts on feature learning. Next, a position encoding network maps the point cloud's spatial information into a high-dimensional feature space, which is fused with the original features to enhance the superpoint representation. The specific calculation of the position encoding is given by Equation (1).

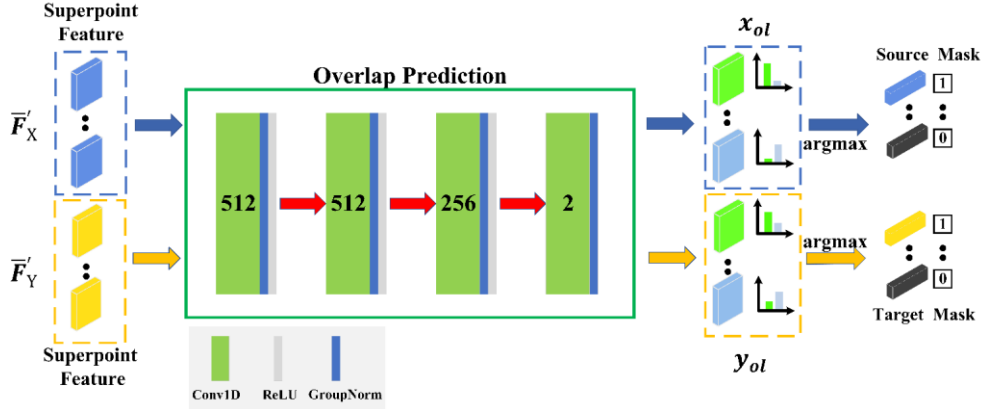


Figure 3. The pipeline of the SOP.

$$\hat{F}'_X = F'_X + \phi\left(P'_X - \frac{1}{n'} \sum_i^{n'} (P'_{x_i})\right) \quad (1)$$

In the equation, $\phi(\cdot)$ denotes the position encoding function. $\hat{F}'_X$ represents the superpoint features with embedded positional information. A similar operation is performed on the superpoints P'_Y of the target point cloud, resulting in the features $\hat{F}'_Y$.

The superpoint features $\hat{F}'_X$ and $\hat{F}'_Y$, now augmented with positional information, are fed into the attention mechanism. This mechanism alternates between self-attention and cross-attention operations to further enhance the feature representations of the superpoints. Subsequently, a multi-layer perceptron (MLP) is used to deeply fuse the processed features with the original features, preserving critical information and improving feature discriminability. To better illustrate the processing steps, we describe the attention mechanism using the superpoint features of the source point cloud as an example. First, self-attention is applied to the superpoint features $\hat{F}'_X$.

$$\bar{F}'_{X_{SA}} = \text{MLP}(\text{Concat}[\hat{F}'_X, \text{MHAttn}(\hat{F}'_X, \hat{F}'_X, \hat{F}'_X)]) \quad (2)$$

Next, cross-attention is applied to $\bar{F}'_{X_{SA}}$, enabling superpoints to extract relevant information from the target point cloud and enhancing feature consistency between corresponding points. Following this, another round of self-attention is performed to further refine the superpoint feature representation. As a result, we obtain the final superpoint features $\bar{F}'_X$, which integrate both positional information and global contextual cues. Similarly, the same processing steps are applied to the target point cloud's superpoint features $\hat{F}'_Y$, yielding the enhanced features $\bar{F}'_Y$.

2.1.2 Superpoint overlap prediction: To enhance the accuracy of superpoint correspondences in the coarse stage, we introduce a SOP module during the superpoint matching phase. This module performs a binary classification task to generate an overlap mask, effectively filtering out non-overlapping superpoints and preventing incorrect correspondences that may disrupt the matching process.

The enhanced superpoint features $\bar{F}'_X$ and $\bar{F}'_Y$ are fed into the SOP module, which consists of 1D convolutional layers with ReLU activation, and employs GroupNorm for normalization to

stabilize training. SOP utilizes these enhanced features to predict the probability of each superpoint belonging to the overlapping region and constructs the overlap mask accordingly. The key workflow of SOP is illustrated in Figure 3. Specifically, we first estimate the probability of each superpoint being in the overlapping region:

$$x_{ol} = h(\bar{F}'_X) \quad (3)$$

$$y_{ol} = h(\bar{F}'_Y) \quad (4)$$

In this process, $h(\cdot)$ denotes the overlap prediction function. We first compute the overlap probabilities x_{ol} and y_{ol} , then apply the argmax operation to generate the overlap masks M_X and M_Y , which identify the overlapping superpoints. Using these masks, we filter out non-overlapping features, retaining only the features within the overlapping regions, denoted as $\bar{F}'_{MX}$ and $\bar{F}'_{MY}$, for further processing. Next, we construct a superpoint feature similarity matrix S' based on $\bar{F}'_{MX}$ and $\bar{F}'_{MY}$ to quantify the similarity between superpoints. To refine this similarity matrix, we employ the Sinkhorn algorithm, which iteratively optimizes S' to satisfy double normalization constraints, ensuring a more reliable matching process. Finally, based on the optimized similarity matrix S' , we derive the superpoint correspondence set C' .

2.2 Point correspondence construction

To refine the superpoint correspondences obtained in the previous stage, we first utilize the decoding layers of the KPConv network. Through nearest-neighbor interpolation, upsampling, and skip connections, we restore the superpoint features to the original resolution of P_X and P_Y , obtaining their corresponding fine-grained feature representations $F_X \in \mathbb{R}^{n \times c}$ and $F_Y \in \mathbb{R}^{m \times c}$. Based on these refined features, we adopt a point-to-node grouping strategy to propagate superpoint correspondences to fine-scale patches, where each patch consists of a set of associated points and their respective feature descriptors. Specifically, for any given superpoint $P'_X(i')$, its associated point set and feature descriptor set can be expressed as:

$$\begin{cases} G_{i'}^P = \{p \in P_X | \|p - P'_X(i')\| \leq \|p - P'_X(j')\|, \forall j' \neq i'\} \\ G_{i'}^F = \{f \in F_X | f \leftrightarrow p \text{ with } p \in G_{i'}^P\} \end{cases} \quad (5)$$

In the equation, $\|\cdot\| = \|\cdot\|_2$ represents the Euclidean distance, and p denotes each point in the point cloud to be assigned. $G_{i'}^P$ and $G_{i'}^F$ represent the generated associated point set and its

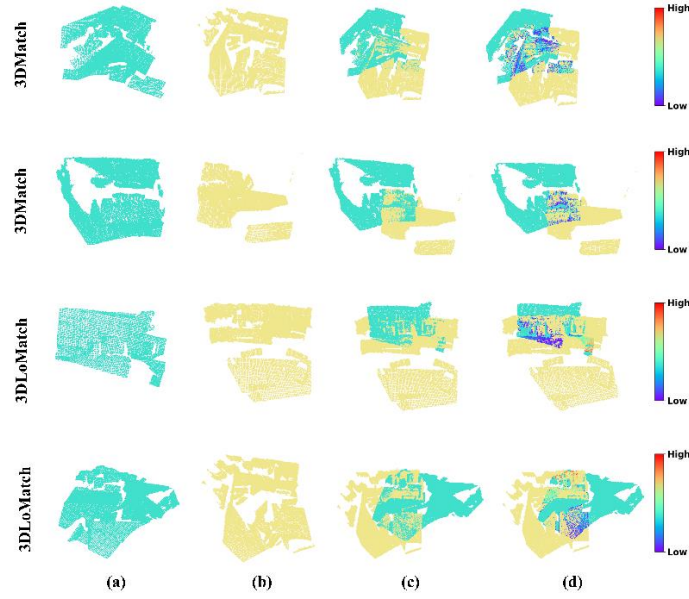


Figure 4. Registration results on 3DMatch and 3DLoMatch. (a) and (b) represent the input point cloud pairs (green represents the source point cloud, and yellow denotes the target point cloud.), (c) shows the ground truth alignment, and (d) presents the estimated registration result.

corresponding feature descriptor, respectively. Through the point-to-node grouping strategy, we extend the superpoint correspondences to their associated patch correspondences, where the correspondences in geometric space are represented as $C_p = \{(G_{i'}^p, G_{j'}^p)\}$ and the correspondences in feature space are represented as $C_F = \{(G_{i'}^F, G_{j'}^F)\}$.

For the obtained patch correspondences, we introduce a density-adaptive mechanism to extract high-quality point correspondences. Specifically, for a given patch correspondence $\{(G_{i'}^p, G_{j'}^F), (G_{j'}^p, G_{i'}^F)\}$, we first utilize an attention mechanism to aggregate local point cloud information while masking out invalid points using "-∞". Then, we perform an optimal transport operation to obtain the similarity matrix $S(i, j)$ and select high-confidence point correspondences to form the point correspondence set $\tilde{C}_i$ for the given patch correspondence. Finally, we merge all point correspondence sets to construct the global point correspondence set $\tilde{C} = \bigcup_{i=1}^{N_{\tilde{C}}} \tilde{C}_i$. Finally, based on the point correspondence set $\tilde{C}$, we solve for the optimal rigid transformation parameters using weighted Singular Value Decomposition (SVD) within the RANSAC framework, thereby achieving the rigid alignment between the source and target point clouds.

2.3 Loss function

To enable the model to learn higher-quality point correspondences and achieve optimal registration performance, we design three loss functions superpoint matching loss $\mathcal{L}_c$, overlap prediction loss $\mathcal{L}_{ov}$, and point matching loss $\mathcal{L}_f$ to provide supervision at different stages of the model.

Superpoint matching loss: For a given pair of superpoints $P_X'(i')$ and $P_Y'(j')$, we use the proportion of true point correspondences in their expanded sets $G_{i'}^p$ and $G_{j'}^p$ to construct their true matching probability. Specifically, the ground truth matching probability $W'(i', j')$ can be defined as:

$$W'(i', j') = \left| \left\{ p \in G_{i'}^p \mid \min_{q \in G_{j'}^p} \|\bar{T}_Y^X(p) - q\|_2 < \tau_p \right\} \right| / |G_{i'}^p| \quad (6)$$

In this equation, $\bar{T}_Y^X$ represents the ground truth rotation and translation matrix, and τ_p denotes the predefined distance threshold. The loss function for the superpoint matching stage can be defined as:

$$\mathcal{L}_c = \frac{-\sum_{i', j'} W'(i', j') \log(S'(i', j'))}{\sum_{i', j'} W'(i', j')} \quad (7)$$

Overlap prediction loss: We define the overlap prediction as a binary classification task, using cross-entropy loss for supervision. The overlap loss of superpoints in the source point cloud can be defined as follows:

$$\mathcal{L}_{ox} = \frac{1}{n'} \sum_{i=1}^{n'} o_{x_i}^* \cdot \log \hat{o}_{x_i} + (1 - o_{x_i}^*) \cdot \log(1 - \hat{o}_{x_i}) \quad (8)$$

In the equation, $o_{x_i}^*$ represents the ground truth overlap labels. Similarly, the $\mathcal{L}_{oy}$ is calculated in this way. The total overlap loss can be calculated by the following equation: $\mathcal{L}_{ov} = \frac{1}{2}(\mathcal{L}_{ox} + \mathcal{L}_{oy})$.

Point matching loss: The matching probability $B \in \mathbb{R}^{(k+1) \times (k+1)}$ for any given patch pair $(G_{i'}^p, G_{j'}^p) \in C_p$ can be defined as:

$$B(i, j) = \begin{cases} 1, & \|\bar{T}_Y^X(G_i^p(i)) - G_j^p(j)\|_2 < \tau_{p'} \\ 0, & \text{otherwise} \end{cases} \quad \forall i, \forall j \in [1, k] \quad (9)$$

We set the rows and columns corresponding to duplicate sampled points in matrix B to 0, and set $B(k+1, k+1) = 0$ to

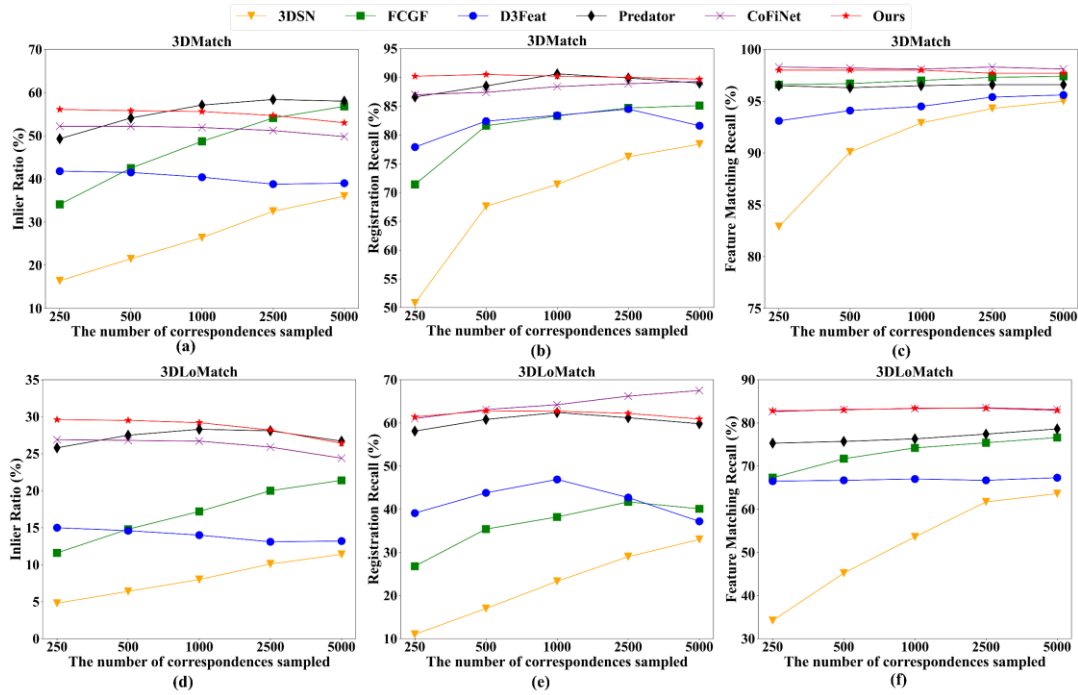


Figure 5. Evaluation results of 3DMatch and 3DLoMatch datasets under different sample sizes ranging from 250 to 5000. (a) to (c) represent the test results on 3DMatch, and (d) to (f) represent the test results on 3DLoMatch.

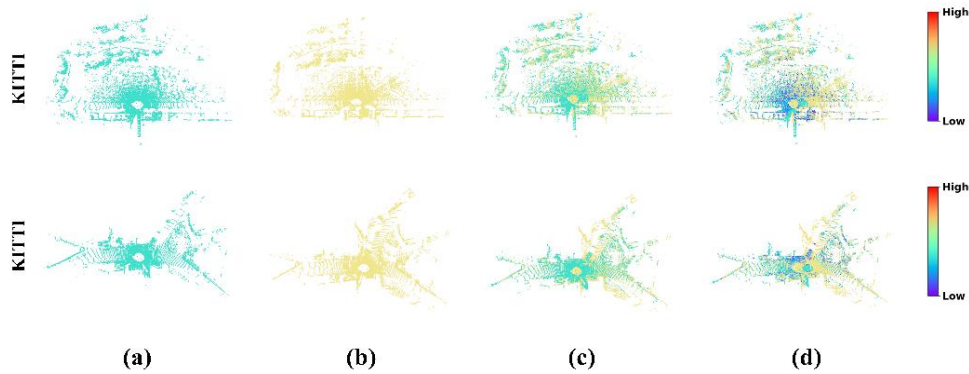


Figure 6. Visualization of registration results on KITTI. (a) and (b) represents the input point cloud pairs (green represents the source point cloud, and yellow denotes the target point cloud.), (c) shows the ground truth alignment, and (d) presents the estimated registration result.

eliminate their side effects during the training process. Therefore, the loss function for the point matching stage is defined as:

$$\mathcal{L}_f = - \sum_{i,j} \mathbf{B}(i,j) \log(\mathcal{S}(i,j)) / \sum_{i,j} \mathbf{B}(i,j) \quad (10)$$

Finally, the total loss can be calculated by the following equation: $\mathcal{L} = \mathcal{L}_c + \mathcal{L}_{ov} + \mathcal{L}_f$.

3. Experiments

To evaluate the performance of the proposed method, we conduct experiments on the indoor datasets 3DMatch (overlap ratio > 30%) and 3DLoMatch (overlap ratio 10%–30%) and compare it with existing methods, including 3DSN(Gojcic et al., 2018), FCGF(Choy et al., 2019), D3Feat(Bai et al., 2020), Predator(Huang et al., 2021), and CoFiNet(Yu et al., 2021). Furthermore, to assess the robustness and generalization ability of the proposed method in large-scale outdoor environments, we also perform experiments on the KITTI dataset.

3.1 Evaluation on 3DMatch and 3DMLoMatch

Evaluation Metrics: Based on prior work, we evaluate the method’s performance using Registration Recall (RR), Inlier Ratio (IR), and Feature Matching Recall (FMR). Specifically, (1) RR quantifies the success rate of point cloud registration, defined as the percentage of point cloud pairs where the rigid transformation estimated by RANSAC has an error below a specified threshold (e.g., $RMSE < 0.2m$). (2) IR measures the reliability of correspondences, representing the fraction of matched points whose geometric residuals are below a predefined threshold (e.g., $\tau_1 = 0.1m$) under the ground truth transformation. (3) FMR defines as the proportion of point cloud pairs where the IR exceeds a given threshold (e.g., $\tau_2 = 5\%$).

Registration Results: To assess the robustness of the proposed method under varying levels of point correspondence sparsity and examine the effect of correspondence quantity on registration accuracy, we follow (Bai et al., 2020) and evaluate the method’s performance across different sampling densities, comparing it with existing approaches. The results are presented in

Method	RTE(m)	RRE(°)	RR(%)
DGR	0.320	<u>0.37</u>	98.7
FMR	0.660	1.49	90.6
SpinNet	0.099	0.47	<u>99.1</u>
CoFiNet	<u>0.085</u>	0.41	99.8
Ours	0.080	0.32	99.8

Table 1 Registration results comparison on the KITTI dataset. The best result is shown in bold and the second best result is underlined.

Figure 4. In terms of IR, the proposed method consistently achieves superior performance across different sampling densities on 3DMatch and 3DLoMatch datasets. Specifically, on 3DMatch, when sampling 250 and 500 correspondences, our method surpasses all competing approaches in IR. On 3DLoMatch, our method achieves the highest matching quality in most cases, except at 5000 samples, where it is slightly outperformed by Predator. For RR, which measures final registration accuracy, our method exhibits competitive performance on 3DMatch, with a slight disadvantage at 1000 samples. In all other cases, it outperforms competing methods, achieving a 0.4%–3.2% improvement in accuracy. On 3DLoMatch, except for a slight gap compared with CoFiNet, all methods achieve better performance than other methods. Further analysis indicates that at 250 samples, both IR and RR reach their peak values across both datasets. The registration results are visualized in Figure 5. Moreover, experimental results indicate that as the number of point correspondences increases, the IR and RR tends to degrade. This decline is primarily attributed to the rising proportion of mismatches, which adversely affects the RANSAC-based transformation estimation. The increase in outlier ratio weakens the support from inliers, thereby reducing both the registration accuracy and the stability of the estimation process.

3.2 Evaluation on KITTI

Evaluation Metrics: For the outdoor dataset KITTI, we assess performance using three key metrics: Relative Translation Error (RTE), Relative Rotation Error (RRE), and Registration Recall (RR). (1) RTE measures the Euclidean distance error between the predicted and ground-truth translation vectors. (2) RRE quantifies the geodesic distance error between the predicted and ground-truth rotation matrices. (3) RR assesses registration success, defined as the proportion of point cloud pairs where RRE is below 5° and RTE is below $2m$ after applying the estimated transformation.

Registration Results: To assess the capability of the proposed method in outdoor point cloud registration, we conduct experiments on the KITTI dataset. We compare our approach with DGR(Choy et al., 2020), FMR(Huang et al., 2020), SpinNet(Ao et al., 2021), and CoFiNet(Yu et al., 2021), with the comparative results summarized in Table 1. Our method demonstrates competitive registration performance against existing approaches in outdoor scenarios, highlighting its strong generalization ability to large-scale environments. Figure 6 illustrates the qualitative registration results on KITTI.

3.3 Ablation study

3.3.1 The reliability of correspondences: To validate the reliability of the point correspondences established by our method, we estimate the rigid transformation parameters directly from the extracted correspondences via Singular Value Decomposition (SVD), without employing the robust pose estimator

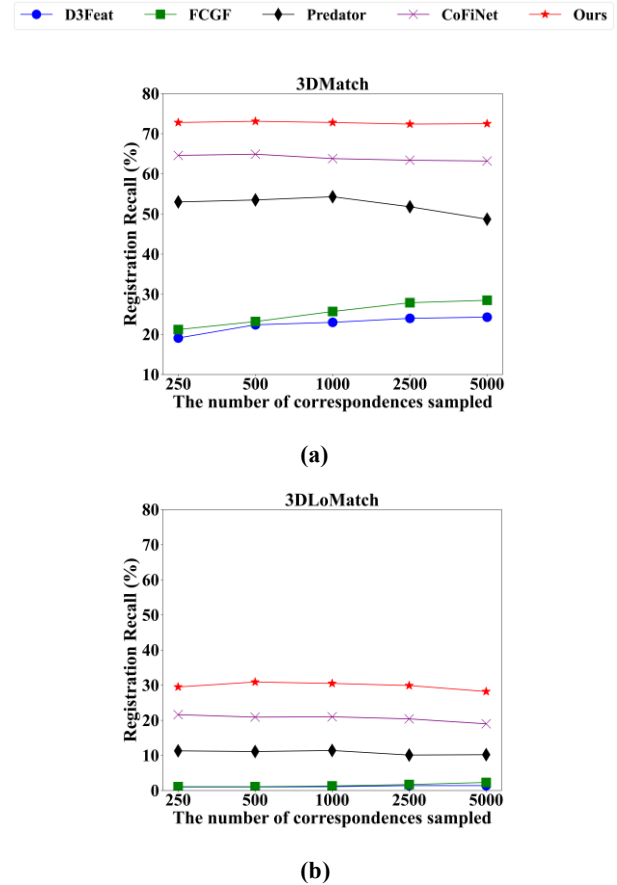


Figure 7. Registration results without RANSAC. (a) represents the registration results on 3DMatch, and (b) represents the registration results on 3DLoMatch.

RANSAC. The experimental results are illustrated in Figure 7.

As shown in the Figure 7, without relying on the RANSAC algorithm, D3Feat, FCGF, and Predator demonstrate strong registration performance on the high-overlap 3DMatch dataset. However, their registration accuracy deteriorates significantly on the low-overlap 3DLoMatch dataset, highlighting a lack of robustness. In contrast, our method consistently delivers stable and superior registration performance across both the 3DMatch and 3DLoMatch datasets. Notably, when compared to the second-best performer, CoFiNet, our approach achieves an average RR improvement of approximately 9%, indicating that the proposed method generates more reliable point correspondences and exhibits strong adaptability in varying overlap scenarios.

3.3.2 The influence of the number of superpoints: Our method employs a coarse-to-fine strategy to gradually establish point correspondences, with the final correspondences refined from the coarse-scale superpoint correspondences. To explore the effect of the number of superpoint correspondences on point matching accuracy and registration performance, we evaluate the IR and RR under varying qualities and quantities of superpoint correspondences. The results are illustrated in Figure 8. The results indicate that increasing the confidence threshold τ_c effectively filters out more reliable superpoint correspondences, thereby improving the IR. However, this also reduces the number of superpoint correspondences available for establishing point correspondences, which consequently leads to a decrease in RR. To address this, we propose a dual-threshold strategy, adding a secondary threshold τ_m to ensure that each point

PE	SOP	3DMatch			3DLoMatch		
		IR(%)	RR(%)	FMR(%)	IR(%)	RR(%)	FMR(%)
×	×	51.2	89.5	98.0	24.3	60.7	80.5
×	√	<u>53.5</u>	89.1	98.1	27.4	61.2	82.0
√	×	54.7	88.6	97.8	<u>28.1</u>	63.0	<u>82.7</u>
√	√	54.7	90.0	97.7	28.2	<u>62.2</u>	83.4

Table 2 The ablation studies of different modules. The best result is shown in bold and the second best result is underlined.

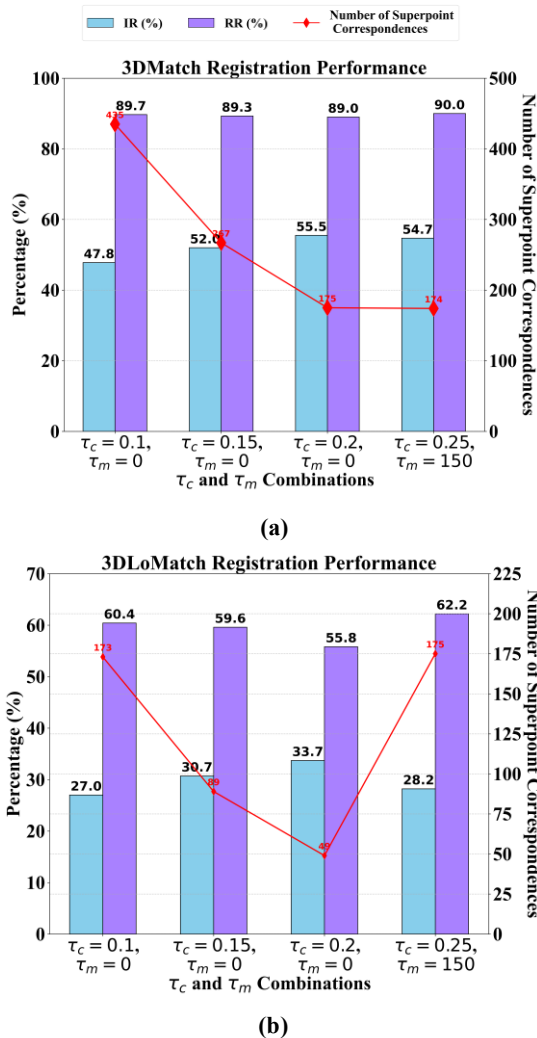


Figure 8. The ablation study on the corresponding number of different superpoints (sample number 2500). (a) and (b) represent the experimental results on 3DMatch and 3DLoMatch, respectively.

cloud pair retains at least τ_m superpoint correspondences, thereby mitigating the issue of insufficient point correspondences caused by an overly high τ_c . Although this strategy slightly decreases the IR, it results in a significant improvement in the RR, thereby enhancing the registration success rate while maintaining a reasonable IR.

3.3.3 Different module ablation experiments: To evaluate the impact of the Position Encoding (PE) and SOP modules on registration performance, we conducted ablation experiments on the 3DMatch and 3DLoMatch datasets, with the results shown in Table 2. The experiments indicate that using SOP alone provides limited improvements in IR and FMR, suggesting that while it helps mitigate interference from non-overlapping regions, its overall contribution to registration performance is not

significant. In contrast, the PE module effectively enhances IR, but its impact on final registration performance remains limited. However, when PE and SOP work together, not only does IR further improve, but RR also reaches its optimal level. The two modules complement each other, jointly enhancing matching quality and overall registration performance.

4. Summary and outlook

With the rapid development of 3D scanning technology, point cloud registration has become increasingly popular. In this work, we propose a coarse-to-fine point cloud registration method based on superpoint overlap perception, focusing on optimizing superpoint matching at the coarse scale. In our approach, we introduce a position-aware attention mechanism to enhance superpoint feature representation and design a superpoint overlap prediction module that generates a mask by predicting overlap probabilities, filtering out invalid superpoints in non-overlapping regions to improve matching accuracy. Experimental results demonstrate that our proposed method achieves high accuracy and robustness on the 3DMatch, 3DLoMatch, and KITTI datasets. However, in low-overlap scenarios, the accuracy of point correspondences still has room for improvement. In the future, we will focus on optimizing the distribution of point correspondences in low-overlap cases to enhance registration accuracy.

Acknowledgements

This work is jointly supported by the National Natural Science Foundation of China (No. 42201486, 42371453, 62302278), the Natural Science Foundation of Shandong Province under Grant ZR2023QF014, the Qingdao Natural Science Foundation under Grant 23-2-1-112-zyyd-jch, the Higher Education Institutions Youth Innovation and Science & Technology Support Program of Shandong Province under Grant 2024KJH066, the Young Talent of Lifting engineering for Science and Technology in Shandong, China [SDAST2024QTA055].

Reference

- Ao, S., Hu, Q., Yang, B., Markham, A. & Guo, Y. Spinn et: Learning a general surface descriptor for 3d point cloud registration. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021. 11753-11762. <https://doi.org/10.1109/CVPR46437.2021.01158>.
- Bai, X., Luo, Z., Zhou, L., Fu, H., Quan, L. & Tai, C.-L. D3feat: Joint learning of dense detection and description of 3d local features. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020. 6359-6367. <https://doi.org/10.48550/arXiv.2003.03164>.
- Besl, P. J. & McKay, N. D. Method for registration of 3-D shapes. Sensor fusion IV: control paradigms and data structures, 1992. Spie, 586-606. <https://doi.org/10.1117/12.57955>.
- Choy, C., Dong, W. & Koltun, V. Deep global registration.

- Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020. 2514-2523.<https://doi.org/10.48550/arXiv.2004.11540>.
- Choy, C., Park, J. & Koltun, V. Fully convolutional geometric features. Proceedings of the IEEE/CVF international conference on computer vision, 2019. 8958-8966
- Gojcic, Z., Zhou, C. & Wieser, A. 2018. Learned compact local feature descriptor for tfs-based geodetic monitoring of natural outdoor scenes. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4, 113-120.<https://doi.org/10.5194/isprs-annals-IV-2-113-2018>.
- Huang, S., Gojcic, Z., Usvyatsov, M., Wieser, A. & Schindler, K. Predator: Registration of 3d point clouds with low overlap. Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition, 2021. 4267-4276.<https://doi.org/10.48550/arXiv.2011.13005>.
- Huang, X., Mei, G. & Zhang, J. Feature-metric registration: A fast semi-supervised approach for robust point cloud registration without correspondences. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020. 11366-11374.<https://doi.org/10.48550/arXiv.2005.01014>.
- Nguyen, A.-D., Choi, S., Kim, W., Kim, J., Oh, H., Kang, J. & Lee, S. 2022. Single-image 3-d reconstruction: Rethinking point cloud deformation. *IEEE Transactions on Neural Networks and Learning Systems*.<https://doi.org/10.1109/TNNLS.2022.3211929>.
- Qin, Z., Yu, H., Wang, C., Guo, Y., Peng, Y., Ilic, S., Hu, D. & Xu, K. 2023. Geotransformer: Fast and robust point cloud registration with geometric transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45, 9806-9821.<https://doi.org/10.1109/TPAMI.2023.3259038>.
- Rusu, R. B., Blodow, N. & Beetz, M. Fast point feature histograms (FPFH) for 3D registration. 2009 IEEE international conference on robotics and automation, 2009. IEEE, 3212-3217.<https://doi.org/10.1109/ROBOT.2009.5152473>.
- Salti, S., Tombari, F. & Di Stefano, L. 2014. SHOT: Unique signatures of histograms for surface and texture description. *Computer vision and image understanding*, 125, 251-264.<https://doi.org/10.1016/j.cviu.2014.04.011>.
- Segal, A., Haehnel, D. & Thrun, S. Generalized-icp. Robotics: science and systems, 2009. Seattle, WA, 435.<https://doi.org/10.15607/rss.2009.v.021>
- Wang, H., Liu, Y., Hu, Q., Wang, B., Chen, J., Dong, Z., Guo, Y., Wang, W. & Yang, B. 2023. RoReg: Pairwise point cloud registration with oriented descriptors and local rotations. *IEEE Transactions on pattern analysis and machine intelligence*, 45, 10376-10393.<https://doi.org/10.1109/TPAMI.2023.3244951>.
- Yang, B., Dong, Z., Liang, F. & Mi, X. 2024. *Ubiquitous Point Cloud: Theory, Model, and Applications*, CRC Press.
- Yang, J., Li, H. & Jia, Y. Go-icp: Solving 3d registration efficiently and globally optimally. Proceedings of the IEEE International Conference on Computer Vision, 2013. 1457-1464.<https://doi.org/10.1109/iccv.2013.184>
- Yang, T., Ye, J., Zhou, S., Xu, A. & Yin, J. 2022. 3D reconstruction method for tree seedlings based on point cloud self-registration. *Computers and Electronics in Agriculture*, 200, 107210.<https://doi.org/10.1016/j.compag.2022.107210>.
- Yu, H., Li, F., Saleh, M., Busam, B. & Ilic, S. 2021. Cofinet: Reliable coarse-to-fine correspondences for robust pointcloud registration. *Advances in Neural Information Processing Systems*, 34, 23872-23884.<https://doi.org/10.48550/arXiv.2110.14076>.
- Zaman, A., Yangyu, F., Ayub, M. S., Irfan, M., Guoyun, L. & Shiya, L. 2023. CMDGAT: Knowledge extraction and retention based continual graph attention network for point cloud registration. *Expert Systems with Applications*, 214, 119098.<https://doi.org/10.1016/j.eswa.2022.119098>.