PEAL: Prior-embedded Explicit Attention Learning for Low-overlap Point Cloud Registration

Junle Yu\textsuperscript{1}  Luwei Ren\textsuperscript{1}  Wenhui Zhou\textsuperscript{1}\textdagger  Yu Zhang\textsuperscript{2}\textdagger  Lili Lin\textsuperscript{3}  Guojun Dai\textsuperscript{1}
\textsuperscript{1}Hangzhou Dianzi University  \textsuperscript{2}Shanghai Jiaotong University  \textsuperscript{3}Zhejiang Gongshang University

Abstract

Learning distinctive point-wise features is critical for low-overlap point cloud registration. Recently, it has achieved huge success in incorporating Transformer into point cloud representation, which usually adopts a self-attention module to learn intra-point-cloud features first, then utilizes a cross-attention module to perform feature exchange between input point clouds. The advantage of Transformer models mainly benefits from the use of self-attention to capture the global correlations in feature space. However, these global correlations may involve ambiguity for point cloud registration task, especially in indoor low-overlap scenarios, because the correlations with an extensive range of non-overlapping points may degrade the feature distinctiveness. To address this issue, we present PEAL, a Prior-embedded Explicit Attention Learning model. By incorporating prior knowledge into the learning process, the points are divided into two parts. One includes points lying in the putative overlapping region and the other includes points located in the putative non-overlapping region. Then PEAL explicitly learns one-way attention with the putative overlapping points. This simplistic design attains surprising performance, significantly relieving the aforementioned feature ambiguity. Our method improves the Registration Recall by 6+\% on the challenging 3DLoMatch benchmark and achieves state-of-the-art performance on Feature Matching Recall, Inlier Ratio, and Registration Recall on both 3DMatch and 3DLoMatch.

1. Introduction

Rigid point cloud registration has always been a foundational yet challenging task in 3D vision and robotics [2,3,10,25], which aims to estimate an optimal rigid transformation to align two point clouds.

Benefiting from the superior feature representation of deep networks, keypoints-based point cloud registration methods have become dominant in recent years [4,9,12,34,37]. The core idea is to learn to match the learned keypoints across different point clouds. Recently, the keypoint-free methods [25,36] demonstrate promising performance following the coarse-to-fine fashion. They seek correspondences between downsampled point clouds (superpoints), which are then propagated to individual points to yield dense correspondences. Thus, the accuracy of superpoint matching is crucial to the overall performance of point cloud registration. GeoTransformer [25] proposes a geometric self-attention module that encodes the distance of point pairs and the angle of triplet to extract transformation-invariant features. This approach significantly improves the accuracy of superpoint matching.

However, GeoTransformer may still suffer from ambiguous matching in certain scenarios with numerous similar structure or low geometric discriminative patches (superpoints) [25]. Moreover, the self-attention mechanism may exacerbate matching ambiguity, especially for low-overlap registration tasks. Prior works [3,25] advocate that modeling geometric consistent correlations among overlap-
pointing superpoints/points is the key to the success of superpoints/points matching, while the global correlations learning via the geometric self-attention is inevitably interfered by numerous superpoints in the non-overlapping region. In other words, the correlations with non-overlapping superpoints may disrupt the inter-frame geometric consistency learning and degrade the feature distinctiveness for registration, which makes the resultant learned superpoint features ambiguous and leads to numerous outlier matches (Fig. 1 (a)).

To address the aforementioned issues, we design a Prior-embedded Explicit Attention Learning model (PEAL). It first leverages an overlap prior to divide the superpoints into anchor ones (the superpoints lying in putative overlapping region) and non-anchor ones (the superpoints located in putative non-overlapping region). Then it alleviates the interference of non-anchor superpoints by introducing an one-way attention mechanism, which solely models the correlations from non-anchor superpoints to anchor ones. Benefiting from the promising overlap ratio in the anchor region, anchor superpoints can be reckoned as simultaneously existing in both two frames, thus the one-way attention is capable of acquiring the essential local geometric consistent correlation from the anchor region, which helps the non-anchor superpoints encoding the inter-frame local geometric consistency and relieves the global feature ambiguity (Fig. 1 (c)). Furthermore, the embedding prior design involved in PEAL makes refining transformation possible in an iterative fashion.

In this paper, we introduce two models depending on the methods of obtaining prior, with extensive experiments on indoor benchmarks demonstrating the superiority of PEAL. Compared to state-of-the-art methods, both of the two models achieve significant improvements on Registration Recall on the challenging 3DLoMatch benchmark. In summary, our contributions are summarized as follows:

- To the best of our knowledge, we are the first to explicitly inject overlap prior into Transformer to facilitate low-overlap point cloud registration, and various overlap priors can be integrated into this framework, such as 3D overlap prior, 2D overlap prior, and self-overlap-prior.
- An explicit one-way attention module, which can significantly relieves the feature ambiguity generated by self-attention. It can be plugged into other transformer-based point cloud registration networks.
- A novel iterative pose refined fashion for low-overlap point cloud registration.

2. Related Work

Correspondence-based Point Cloud Registration. Correspondence-based methods [7, 9, 12] firstly establish correspondences between learned keypoints and then recover the transformation with a robust pose estimator such as RANSAC or other RANSAC-free estimator [3, 5, 6, 22, 25]. Many works focus on learning keypoint detectors [4, 17] and feature descriptors [1, 7, 32]. Recently, detector-free methods [25, 36] have become prevalent in a coarse-to-fine fashion. Our method inherits the detector-free methods and focuses on improving the matching accuracy of the coarse level.


Image matching. Image matching [21, 26, 28, 30] is a basic and important technology in computer vision. SIFT [21] and ORB [27] are typical handcrafted local features which are widely adopted in many 3D computer vision tasks. Learning-based methods [28, 30] can significantly improve the performance of local features under large viewpoints and illumination changes. We further explore how to leverage image signals to facilitate 3D point cloud registration.

Transformers. The Transformer models [31] utilize a novel attention mechanism, employing multiple layers of self and cross multi-head attention to facilitate information exchange between input and output, demonstrating the superior performance of feature representation for NLP and vision tasks. Focal self-attention [33] is to incorporate fine-grained local and coarse-grained global interactions to capture both short and long-range visual dependencies. Deformable DETR [39] uses deformable attention, which solely attends to a small set of key sampling points around a reference to aggregate multi-scale image features. GeoTransformer [25] utilizes geometric self-attention to extract transformation-invariant geometric features by encoding geometric structure. In this paper, we adopt an explicit attention learning fashion to learn distinctive features.

3. Method

We define point clouds \( P = \{ p_{x_i} \in \mathbb{R}^3 \mid i = 1, \ldots, N \} \) and \( Q = \{ q_{y_i} \in \mathbb{R}^3 \mid i = 1, \ldots, M \} \), as source and target, respectively. The purpose of point cloud registration is to recover the unknown rigid transformation \( SE(3) \), which consists of a rotation \( R \in SO(3) \) and a translation \( T \in \mathbb{R}^3 \) for aligning \( P \) and \( Q \).

The pipeline of our method is illustrated in Fig. 2. We
We follow GeoTransformer [25] for superpoint matching in order to extract the superpoint correspondences, which are then propagated to the point matching module to obtain dense point correspondences for calculating the final transformation. At last, the iterative update module (Sec. 3.3) is introduced, which can iteratively refine the transformation by using the predicted overlap prior (self-overlap-prior) generated by the currently estimated transformation.

3.1. Overlap Prior Prediction

Existing point cloud registration methods still face challenges in accurately registering two point clouds, especially in low-overlap scenarios. This often results in partially aligned or completely erroneous final transformations. Despite this, these partially aligned scenarios still show potential for achieving higher overlap. And the estimated overlapping regions in these cases can serve as a 3D overlap prior. Our goal is to improve these partially aligned scenarios by using the 3D overlap prior, transforming them into successfully registered ones.

On the other hand, inspired by multi-modal learning methods, it is worth considering the use of 2D images in RGB-D datasets. Image matching approaches could be utilized to estimate 2D correspondences, which can then be projected onto point clouds through a projection module. By doing so, we can obtain a preliminary estimation of the overlap between images, which serves as a 2D overlap prior.

3D overlap prior We use GeoTransformer [24] to generate 3D overlap prior, which is currently the state-of-the-art method. The predicted overlapping region is generated by the nearest neighbor search within a threshold in the Euclidean space according to the transformation matrix estimated by the pretrained GeoTransformer. The overlapping region is regarded as 3D overlap prior, namely 3dprior. The

Figure 2. KPConv-FPN downsamples the input point clouds and learns features in multiple resolution levels. The geometric self-attention learns a hybrid geometric feature by encoding the geometric structure first, followed by intra-frame attention between anchor (selected according to the given overlap prior, painted in red or green) and non-anchor superpoints (painted in blue or yellow). This enables the learning of distinctive geometric features, which are then exchanged between two point clouds using a feature-based cross-attention. The resulting correspondences are then sent to the point matching module to calculate the transformation. During testing, the self-overlap-prior is computed based on the current transformation and iteratively transferred to the network.

Figure 3. The pipeline of acquiring 3D overlap prior. We show the correct estimated overlapping point cloud in green and the incorrect one in red.

first introduce how to acquire overlap prior (Sec. 3.1). Given the overlap prior, we then adopt the coarse-to-fine technique to extract correspondences. We build our method upon GeoTransformer [25], which utilizes KPConv-FPN to downsample the input point clouds and extract point-wise features simultaneously. The first and the last (coarsest) level of downsampled points correspond to the dense points and the superpoints. The superpoints are denoted by $\hat{P}$, and $\hat{Q}$, with the associated learned features denoted as $\hat{F}_P$ and $\hat{F}_Q$.

The superpoints with local patch containing overlap prior points are regarded as anchor superpoints, denoted as $\hat{P}_A$, $\hat{Q}_A$ for $\hat{P}$, $\hat{Q}$, respectively, while the other superpoints are viewed as non-anchor superpoints denoted as $\hat{P}_{N,A}$, $\hat{Q}_{N,A}$. And a superpoint matching module (Sec. 3.2) is used to extract superpoint correspondences. During superpoint matching, we first compute intra-frame geometric self-attention like GeoTransformer [25], then an explicit one-way attention module is proposed to encode the inter-frame local geometric consistency, which captures the one-way correlation from non-anchor superpoints to anchor ones. And finally, the inter-frame feature-based cross attention is computed to perform feature exchange between two input point clouds.
pipeline of acquiring 3dprior is illustrated in Fig. 3.

**2D overlap prior** For the RGB-D dataset, we follow PCR-CG [38] which utilizes the 2D image matching method to generate 2D correspondences and convey them to point clouds via a 2D-3D projection module. SuperGlue [28] is utilized to extract image correspondences first. However, 2D correspondences are sparse, combined with the presence of invalid projections, making it difficult to obtain sufficient overlapping points. To tackle this issue, we follow PCR-CG [38] by creating a box for each matched pixel in every pair of image matches. Lastly, a projection module is utilized to lift 2D overlapping pixels to 3D overlapping points that serves as the overlap prior, called 2dprior. The pipeline of acquiring 2dprior is illustrated in Fig. 4.

### 3.2. Superpoint Matching

The registration performance of existing coarse to fine methods [25, 36] heavily relies on accurate superpoint matching, and unreliable superpoint matching causes the failure of the registration. Some recent methods, such as GeoTransformer [25], propose encoding global geometric structure to learn transformation-invariant features.

However, utilizing self-attention for learning intra-frame point-wise features may introduce ambiguity because self-attention learns the global correlations, in which the correlations with a large range of non-overlapping superpoints introduce ambiguous ones. While the correlations with the overlapping superpoints are the key to encoding inter-frame geometric consistency for the point cloud registration task.

We propose to model the correlation with overlapping superpoints to acquire inter-frame geometric consistent correlations. Specifically, we suggest incorporating overlap prior knowledge into the computation of point-wise features by modeling non-global correlations with the anchor superpoints. Compared to the original global overlap ratio, this approach allows for a higher overlap ratio in the anchor region and enables the network to avoid introducing ambiguous geometric correlations.

To this end, we propose an one-way attention module to explicitly learn one-way correlation with anchor superpoints. We achieve superpoint matching by utilizing three attention modules, including a geometric self-attention [25] module to learn intra-frame point-wise geometric features, an explicit one-way attention module to encode inter-frame local geometric consistency, and a feature-based cross-attention module to perform feature exchange between source and target point clouds. These three attention modules are interleaved for $N_t$ times to extract hybrid features $H^P$ and $H^Q$ for reliable superpoint matching (see Fig. 2).

**Geometric self-attention.** During superpoint matching, given $\mathcal{P}$ and $\mathcal{Q}$, in order to learn point-wise attention features, we follow GeoTransformer [25] to compute the intra-frame self-attention $X^P$ and $X^Q$. Given the input feature matrix $X \in \mathbb{R}^{P \times d_t}$, the output feature matrix $Z \in \mathbb{R}^{P \times d_t}$ is obtained by performing a weighted sum of all projected input features:

$$z_i = \sum_{j=1}^{\left|\mathcal{P}\right|} a_{i,j} \left(x_j W^V\right)$$

(1)

the weight coefficient $a_{i,j}$ is obtained through a row-wise softmax function applied to the attention score $e_{i,j}$, and the computation of $e_{i,j}$ is shown as followed:

$$e_{i,j} = \frac{(x_i W^Q) \left(x_j W^K + r_{i,j} W^R\right)^T}{\sqrt{d_t}}$$

(2)

Here, $r_{i,j} \in \mathbb{R}^{d_t}$ is a geometric structure embedding which consists of a pair-wise distance embedding and a triplet-wise angular embedding [25]. $W^Q, W^K, W^V, W^R \in \mathbb{R}^{d_t \times d_t}$ are projection matrices for queries, keys, values, and geometric structure embeddings, respectively.

Then, we get superpoints attention feature matrices $X^\mathcal{P}$ and $X^\mathcal{Q}$, anchor superpoints attention feature matrices $X^{\mathcal{P},A}$ and $X^{\mathcal{Q},A}$. $X^{\mathcal{P},A}$ represents the anchor superpoints attention matrices of $\mathcal{P}$, which is indexed from the self-attention matrices $X^\mathcal{P}$ according to divided putative overlapping superpoints, and the same goes for $X^{\mathcal{Q},A}$. Similarly, $X^{\mathcal{P},NA}$ and $X^{\mathcal{Q},NA}$ represent the non-anchor superpoints attention feature matrices for $\mathcal{P}$ and $\mathcal{Q}$, respectively.

**Explicit one-way attention.** We propose an one-way attention module to explicitly learn the intra-frame correlation with the anchor superpoints, which is key to encoding inter-frame local geometric consistency. Given anchor superpoints attention feature matrices $X^{\mathcal{P},A}$ and non-anchor superpoints attention feature matrices $X^{\mathcal{P},NA}$, the non-anchor superpoints attention feature matrices $Z^{\mathcal{P},NA}$ can be com-
computed with the features of anchor superpoint features $X^{\hat{P}}_A$.

$$z_{m,n}^{P} = \sum_{n=1}^{p_{m,n}} a_{m,n} x_{m,n} W^{V}$$

(3)

Similar to geometric self-attention, $a_{m,n}$ represents a row-wise softmax on the attention score $e_{m,n}$, which is the feature correlation between the $X^{P,N_A}$ and $X^{P,A}$.

$$e_{m,n} = \left( x_{m,n} W^{P,A} \right)^T \left( x_{n} W^{P,K} \right) \frac{1}{\sqrt{d_{f}}}$$

(4)

The attention features for $X^{Q,N_A}$ are updated same as $X^{P,N_A}$, while $X^{Q,A}$ and $X^{P,A}$ remain unchanged.

Next, we use an inter-frame feature-based cross-attention module to encode global inter-frame geometric consistency between input point clouds [25]. The resultant superpoint features are then sent to the superpoint matching module for reasoning precise superpoint correspondences. Finally, we perform point matching [25] to extract the dense point correspondences and local-to-global registration to obtain the estimated transformation.

3.3. Iterative Update

Our superpoint matching module incorporates an explicit one-way attention module that is sensitive to anchor superpoints, allowing for better-estimated transformations by improved overlap ratio in the anchor region. This leads to more distinctive features for superpoint matching and reliable correspondences, resulting in higher accuracy and more reliable final transformation. In other words, it can be designed in an iterative fashion to refine the transformation. The iterative update module starts from the currently estimated transformation, and the self-overlap-prior is computed according to the current transformation and iteratively transferred into the network. At each iteration, it produces an updated transformation $T_k$, which is then applied to estimate $T_{k+1} = f(T_k)$.

4. Experimental Results

In this section, we evaluate our method on indoor 3DMatch [37] and 3DLoMatch [17] benchmarks.

4.1. 3DMatch & 3DLoMatch

Dataset. 3DMatch [37] is composed of 62 scenes among which 46 are used for training, 8 for validation, and 8 for testing. Each scene has its corresponding RGB-D data. We evaluate our approach using preprocessed training point clouds provided by [17] and test it on both the 3DMatch and 3DLoMatch protocols. The point cloud pairs in 3DMatch have an overlap of more than 0.3, while those in 3DLoMatch have a lower overlap of 0.1 ~ 0.3. We collect the associated RGB-D data [38] of these two benchmarks, where each point cloud is fused by 50 consecutive depth frames.

<table>
<thead>
<tr>
<th>Samples</th>
<th>3DMatch &amp; 3DLoMatch</th>
<th>5000</th>
<th>2500</th>
<th>1000</th>
<th>500</th>
<th>250</th>
</tr>
</thead>
<tbody>
<tr>
<td>PerfectMatch [12]</td>
<td>93.0</td>
<td>94.3</td>
<td>92.9</td>
<td>90.1</td>
<td>82.9</td>
<td>83.6</td>
</tr>
<tr>
<td>FCDF [7]</td>
<td>97.4</td>
<td>97.3</td>
<td>97.0</td>
<td>96.7</td>
<td>96.6</td>
<td>76.6</td>
</tr>
<tr>
<td>D3Feat [4]</td>
<td>93.6</td>
<td>95.4</td>
<td>94.1</td>
<td>94.1</td>
<td>93.1</td>
<td>67.3</td>
</tr>
<tr>
<td>SpinNet [1]</td>
<td>97.6</td>
<td>97.2</td>
<td>96.8</td>
<td>95.5</td>
<td>94.5</td>
<td>75.3</td>
</tr>
<tr>
<td>Predator [17]</td>
<td>96.6</td>
<td>96.6</td>
<td>96.5</td>
<td>96.3</td>
<td>96.3</td>
<td>78.4</td>
</tr>
<tr>
<td>YOHO [32]</td>
<td>98.2</td>
<td>97.6</td>
<td>97.5</td>
<td>97.7</td>
<td>96.0</td>
<td>79.4</td>
</tr>
<tr>
<td>CoFiNet [36]</td>
<td>98.1</td>
<td>98.3</td>
<td>98.1</td>
<td>98.2</td>
<td>98.2</td>
<td>83.1</td>
</tr>
<tr>
<td>PCR-CG [38]</td>
<td>97.4</td>
<td>97.5</td>
<td>97.7</td>
<td>97.3</td>
<td>97.6</td>
<td>80.4</td>
</tr>
<tr>
<td>GeoTransformer [25]</td>
<td>97.9</td>
<td>97.9</td>
<td>97.9</td>
<td>97.6</td>
<td>97.6</td>
<td>88.3</td>
</tr>
</tbody>
</table>

| Inlier Ratio (%) | PEAL(ours) | 99.0 | 99.0 | 99.1 | 99.1 | 98.6 | 98.6 | 98.6 | 98.6 | 99.2 |
| Registration Error (m) | PEAL(ours) | 74.2 | 64.4 | 60.7 | 55.7 | 48.7 | 42.5 | 34.1 | 27.6 | 21.5 |

Table 1. Evaluation results on 3DMatch and 3DLoMatch.

The RGB images and depth images are in pairs. Therefore, each point cloud is also associated with 50 RGB frames.

Experiments Setup. When acquiring the 3D overlap prior, the predicted overlap region is generated by the nearest neighbor search under the estimated transformation within a threshold of 0.0375m. To obtain 2D overlap prior, considering the computation cost of inferring 2D correspondences, we use three frames of images for the source and the target point clouds and extract correspondences from nine pairs of images using SuperGlue [28], with an image resolution of [640, 480]. We project the XYZ coordinates of each point cloud onto its associated image plane to obtain pixel-point corresponding relation with a depth threshold of 0.2m. During training, we use an Adam optimizer with a learning rate of 1e-4 and a batch size of 1 on a single GPU (RTX3090) for a total of 20 epochs. For testing, we always use PEAL-3dprior for iterative updates and perform six iterations to refine the transformation. Other experimental setup goes the same with GeoTransformer [25].
REGTR [35], and Lepard [18].

**RANSAC estimator results.** Following [4, 17], we report the results with different numbers of correspondences using the RANSAC estimator in Tab. 1. Our method achieves the highest Feature Matching Recall on all the sampled correspondences on 3DMatch and 3DLoMatch. Our method improves by no less than 3.4% on 3DLoMatch and 1.1% on 3DMatch compared to the baseline GeoTransformer [25]. For Inlier Ratio, PEAL improves by 1.5% ~ 11.2% on 3DLoMatch, and improves by 0.5% ~ 8.1% on 3DMatch. The improvement is more prominent with fewer sampled correspondences, for instance, we achieve 60%+ Inlier Ratio on 500 and 250 sampled correspondences on 3DLoMatch. For Registration Recall, PEAL outperforms the previous best (Lepard [18] in 3DMatch Tab. 2 (top), GeoTransformer [25] in 3DLoMatch Tab. 1 (bottom)) by 1.1% on 3DMatch and 6.7% on 3DLoMatch.

**RANSAC-free estimator.** We compare the registration results of the RANSAC-free estimator in Tab. 2 (bottom). We separate our models into PEAL-2dprior and PEAL-3dprior according to the method of acquiring overlap prior. Both of them achieve state-of-the-art performance on 3DMatch 3DLoMatch. The improvement is more prominent than using the RANSAC estimator. PEAL-2dprior improves previous best (REGTR [35] in 3DMatch, GeoTransformer [25] in 3DLoMatch) by 2.3% on 3DMatch and 7.2% on 3DLoMatch. In addition, PEAL-3dprior outperforms the baseline by 2.6% on 3DMatch and 4.8% on 3DLoMatch, showing significant improvement compared to baseline method.

**RRE and RTE.** We then compare the RRE and RTE with the recent state of the arts: [1, 4, 7, 17, 25, 35, 36] in Tab. 3. As shown in this table, we achieve the second best in RRE, RTE on 3DMatch, and RTE on 3DLoMatch, and the best in RRE on 3DLoMatch.

**Ablation studies.** In this section, we ablate the different setups for the proposed modules in our method. In practice, we notice local-to-global registration (LGR) [25] estimator is much more stable and faster than the RANSAC estimator while achieving a comparable performance (see Tab. 2). Thus, we conduct our ablation experiments based on the LGR estimator.

To explore the effectiveness of explicit one-way attention module, we conduct various one-way attention strategies between anchor and non-anchor superpoints, reporting the FMR, IR, and RR on 3DMatch and 3DLoMatch. We compared three strategies: anchor to non-anchor, non-anchor to anchor, and bidirectional (combine both non-anchor to anchor and anchor to non-anchor). Our findings indicate that the local geometry correlation from anchor points to non-anchor points is inter-frame geometrically inconsistent when performing anchor to non-anchor attention. When updating the features of anchor points, they will be severely contaminated by the non-anchor region, leading to a significant drop in registration recall compared to GeoTransformer, as shown in Tab. 4. Overall, our study confirms that explicit one-way attention from non-anchor superpoints to anchor superpoints can effectively encode inter-frame local geometric consistency.

To evaluate prior-embedded explicit attention learning, we compare it with a prior-embedded implicit attention learning approach that involves making anchor points salient. Specifically, prior to computing the self-attention, the anchor points are initialized with the pretrained KPConv
Table 5. Ablation experiments about the implicit attention learning and explicit attention learning with embedded prior.

<table>
<thead>
<tr>
<th>Model</th>
<th>anchor superpoints selection</th>
<th>RR(%)</th>
<th>3DM</th>
<th>3DLM</th>
</tr>
</thead>
<tbody>
<tr>
<td>PEAL-3dprior</td>
<td>PEAL-3dprior</td>
<td>random choose 40%</td>
<td>93.5</td>
<td>78.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>random choose 80%</td>
<td>94.1</td>
<td>78.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>base region</td>
<td>94.1</td>
<td>78.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+ 4 nearest superpoints</td>
<td>93.9</td>
<td><strong>78.9</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>+ 12 nearest superpoints</td>
<td>93.5</td>
<td>77.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+ 16 nearest superpoints</td>
<td>93.2</td>
<td>77.2</td>
</tr>
</tbody>
</table>

Table 6. Ablation experiments on the anchor region. We take the overlapping region estimated by GeoTransformer [25] as the base region.

<table>
<thead>
<tr>
<th>Model</th>
<th>Views</th>
<th>RR(%)</th>
<th>3DM</th>
<th>3DLM</th>
</tr>
</thead>
<tbody>
<tr>
<td>PEAL-2dprior</td>
<td>1</td>
<td>92.5</td>
<td>78.8</td>
<td></td>
</tr>
<tr>
<td>PEAL-2dprior</td>
<td>2</td>
<td>93.7</td>
<td>80.7</td>
<td></td>
</tr>
<tr>
<td>PEAL-2dprior</td>
<td>3</td>
<td><strong>94.3</strong></td>
<td><strong>81.2</strong></td>
<td></td>
</tr>
<tr>
<td>PEAL-2dprior</td>
<td>4</td>
<td>94.3</td>
<td>81.2</td>
<td></td>
</tr>
</tbody>
</table>

Table 7. Ablation experiments on the number of images.

The performance is close to baseline, implying that our method does not perform well with randomly initialized overlap prior. However, randomly expanding/shrinking the base anchor region by 20% or choosing the 16 nearest nodes only results in a slight drop in performance, suggesting that changes to the base anchor region have limited effect on the overlap ratio and consequently do not significantly affect the final performance.

Tab. 7 demonstrates that registration recall can be improved with more views involved, because more overlapping point clouds can be found with the number of views growing, especially in low-overlap scenarios with large viewpoints and illumination changes.

Next, we ablate the iterative update module in Tab. 8 and Fig. 5. We evaluate the number of iterations on registration recall. We use the pretrained PEAL to generate the initial pose for ICP and PEAL-iter. The registration recall continues to improve with increasing iterations, as depicted in Fig. 5. The improvement slows down after 8 iterations, so we set the iteration number to 6 in our experiments tak-
We present PEAL for rigid point cloud registration, which explicitly incorporates overlap prior to attention learning. It yields distinctive features for superpoint and dense point matching, and various overlap priors can be integrated into this framework, such as 3D overlap prior, 2D overlap prior, and self-overlap-prior. PEAL learns to extract the discriminative geometric features through explicit one-way attention, significantly relieving the feature ambiguity generated by self-attention.

**Acknowledgments** This work is supported in part by the National Key R&D Program of China (2017YFE0118200), the Zhejiang Provincial Natural Science Foundation of China (LTY22F020001, LY21F010007) and the National Natural Science Foundation of China (62076083).

<table>
<thead>
<tr>
<th># Sampled Points</th>
<th>Registration Recall (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predator [17]</td>
<td>59.8  61.2  62.4  60.8  58.1</td>
</tr>
<tr>
<td>Predator+PEAL</td>
<td><strong>70.2</strong>  <strong>70.2</strong>  <strong>69.4</strong>  <strong>68.6</strong>  <strong>63.3</strong></td>
</tr>
</tbody>
</table>

Table 9. Compared with Predator baseline on 3DLoMatch.


