Abstract

Recent state-of-the-art learning-based approaches to point cloud registration have largely been based on graph neural networks (GNN). However, these prominent GNN backbones suffer from the indistinguishable features problem associated with oversmoothing and structural ambiguity of the high-level features, a crucial bottleneck to point cloud registration that has evaded scrutiny in the recent relevant literature. To address this issue, we propose the Distinctiveness oriented Positional Equilibrium (DoPE) module, a novel positional embedding scheme that significantly improves the distinctiveness of the high-level features within both the source and target point clouds, resulting in superior point matching and hence registration accuracy. Specifically, we use the DoPE module in an iterative registration framework, whereby the two point clouds are gradually registered via rigid transformations that are computed from DoPE’s position-aware features. With every successive iteration, the DoPE module feeds increasingly consistent positional information to would-be corresponding pairs, which in turn enhances the resulting point-to-point correspondence predictions used to estimate the rigid transformation. Within only a few iterations, the network converges to a desired equilibrium, where the positional embeddings given to matching pairs become essentially identical. We validate the effectiveness of DoPE through comprehensive experiments on various registration benchmarks, registration task settings, and prominent backbones, yielding unprecedented performance improvement across all combinations.

1. Introduction

Point cloud registration is a well-known task by which two point clouds are matched via a rigid transformation. For a source point cloud \( X \) and a target point cloud \( Y \), the registration problem is finding a rigid transformation that minimizes the geometric shape differences between \( Y \) and the transformed \( X \). In many applications such as 3D reconstruction and simultaneous localization and mapping (SLAM), the registration process has long relied on traditional, non-learning-based algorithms to predict the optimal rigid transformations.

Recently, deep learning methods have brought remarkable advances in a variety of 3D vision tasks, ranging from classification, segmentation, and point cloud registration. A common theme among many learning-based registration methods [18, 19, 24, 5, 25] is the fact they are comprised of 1) a feature extraction backbone, usually a graph neural network (GNN), which generates per-point feature descriptors via iterative local aggregation, followed by 2) a feature matching step, which computes point-to-point matchability scores, or (soft) correspondences, between the source and target point clouds using their extracted features.

For example, Deep Closest Point (DCP) [18] computes point correspondences from learned features, via attention combined with pointer generation, in order to desensitize the network from initialization and avoid local minima. RPM-Net [24] incorporates Robust Point Matching (RPM) [3] into the feature matching step to be able to also handle outliers and missing correspondences. On the other hand, DeepGMR [25] avoids exhaustive point-to-point correspondences all together by learning correspondences from both point clouds to a common distribution inside a learned latent space.

While these recent methods have made significant improvements to the feature matching step and displayed state-of-the-art performance, they overlook a key design consideration for feature extraction that can critically affect registration accuracy: the distinctiveness of the per-point features within both the source and target point clouds; that is, in order to obtain accurate point-to-point correspondences for estimating the optimal rigid transformation, the desired point features should sufficiently represent the geometric pattern in the neighborhood of any given point while still being distinguishable enough from the local patterns surrounding other points within the same point cloud. However, many of the GNN backbones typically used to embed the input point clouds into the feature space [12, 20] are susceptible to oversmoothing [6, 17, 1] and structural

---

*This work was done while Taewon Min, Chonghyuk Song, and Eunseok Kim were with the Agency for Defense Development (ADD).  
†Corresponding author: iwshim@add.re.kr
ambiguity, resulting in indistinguishable point features. Figure 1a demonstrates this phenomenon, henceforth referred to as the indistinguishable feature problem, which results in an overwhelming number of ambiguous point-to-point correspondences as opposed to the sharp matches desired for accurate registration.

To address these issues, we propose Distinctiveness Oriented Positional Equilibrium (DoPE), a novel, lightweight positional embedding module that significantly improves the intra-set distinctiveness of both the source and target point cloud embeddings, thereby enhancing the resulting point-to-point correspondences. Specifically, DoPE disambiguates the per-point features by augmenting them with global positional information computed with respect to the centroid of the combined source and target point clouds, which act as the origin of a shared coordinate system. We use this DoPE module in an iterative registration framework, where by the two point clouds are gradually aligned by rigid transformations computed from DoPE’s position-aware features. We use DoPE as part of an iterative registration framework, whereby the two point clouds are gradually aligned by rigid transformations computed from DoPE’s position-aware features.

• We identify and analyze the contributing factors to the indistinguishable features problem, a critical bottleneck to point-cloud registration that is prevalent in GNN-based architectures but has evaded scrutiny in the recent registration literature.

• To address this issue, we propose the Distinctiveness Oriented Positional Equilibrium (DoPE) module, a novel positional embedding scheme that disambiguates point features and enhances the resulting rigid-transformation predictions. We use DoPE as part of an iterative registration framework, whereby the two point clouds are gradually aligned by rigid transformations computed from DoPE’s position-aware features.

• We demonstrate the effectiveness of DoPE by incorporating the module into the state-of-the-art registration architectures and performing comprehensive experiments on various registration datasets and task settings, yielding unprecedented performance improvement across all combinations.

2. Related Work

Deep Learning on Point Clouds  Deep Learning on point clouds was pioneered by PointNet [11], which directly con-
Graph neural networks (GNN) provide a natural way to encode the local geometry of point clouds by virtue of local aggregation at each layer. For example, PointNet++ [12] recursively applies PointNet on a locally constructed graph (e.g., ball query or k-NN graphs); dynamic graph convolutional neural networks (DGCNN) [20] constructs a local neighborhood graph in the feature space and applies local feature aggregation on the edges connecting neighboring pairs of points. While GNN-based approaches have made significant improvements to 3D vision tasks such as point cloud classification and segmentation, many GNN backbones are prone to oversmoothing [6, 17, 1] and structural ambiguity, resulting in indistinguishable point features, which are detrimental to the feature matching step in point cloud registration.

**Learning-based Registration** The latest learning-based approaches to registration [18, 19, 24, 5, 25] have largely focused on improving the matching process between the embedded feature descriptors of the input point clouds, which are typically generated by a graph neural network (GNN). Specifically, DCP [18] finds matched correspondences from learned features via attention combined with pointer generation, while RPM-Net [24] incorporates Robust Point Matching [3] into a learning framework to be able to handle missing correspondences. PRNet [19] and IDAM [5] both extract keypoints and then iteratively find keypoint-to-keypoint correspondences. DeepGMR [25] explicitly proposes a probabilistic registration model by using Gaussian Mixture Model (GMM) parameters.

However, these methods remain oblivious to the fact that the GNN backbones typically used to embed the input point clouds are prone to the indistinguishable feature problem, thereby severely lacking the intra-set distinctiveness required to generate accurate point-to-point correspondences from the embedded features. In this paper, we identify and analyze the contributing factors to the indistinguishable features problem and propose a novel positional embedding module to significantly enhance the intra-set distinctiveness of the per-point features.

Concurrent work [8] has also suggested the use of positional encoding to improve the intra-set distinctiveness of the point descriptors. However, the positional encoding scheme in [8] remains local to each point cloud and was not born out of the full awareness of the underlying registration bottleneck inadvertently addressed by the proposed method. On the other hand, DoPE, fully motivated by the indistinguishable features problem, feeds positional embeddings computed with respect to the centroid of the combined source and target point clouds, which act as the origin of a *shared* coordinate system. As a result, the network remains aware of the relative spatial orientation of the point clouds during the iterative registration process, a trait that we empirically show to be indispensable to DoPE’s outstanding performance.

### 3. Preliminaries

#### 3.1. Problem Statement

Point cloud registration is the process of finding a rigid transformation that best aligns two unaligned point clouds. Let \( \mathcal{X}, \mathcal{Y} \) be two finite point sets, which contain \( J \) and \( K \) points, respectively. Assuming that \( \{x_1, x_2, x_3, \ldots, x_N\} \subset \mathcal{X} \) and \( \{y_{x_1}, y_{x_2}, y_{x_3}, \ldots, y_{x_N}\} \subset \mathcal{Y} \) are two sets of corresponding point clouds and \( N \) is the number of corresponding pairs (\( N \leq J \) and \( N \leq K \)), the optimal rotation \( \hat{R} \) and translation \( \hat{t} \) are estimated as follows:

\[
(\hat{R}, \hat{t}) = \arg\min_{R \in SO(3), t \in \mathbb{R}^3} \sum_{i=1}^{N} \| (Rx_i + t) - y_i \|^2 ,
\]

where \( (\hat{R}, \hat{t}) \) comprise the rigid transformation that best aligns the two point clouds.

#### 3.2. Iterative Registration Process

To find the optimal transformation in Eq. (1), many works [19, 24, 5] follow the iterative procedure shown in Figure 2. In each iteration, the source and target point clouds \( \mathcal{X} \) and \( \mathcal{Y} \) are first fed into the feature extraction layer to generate the high-level features. Next, the feature matching layer finds point-to-point correspondence. Finally, the optimal transformation \( \hat{T} = [\hat{R} \; \hat{t}; \; 0 \; 1] \) is estimated using Singular Value Decomposition (SVD) [14]. The whole process is repeated with \( \mathcal{Y} \) and \( \mathcal{X} \) transformed by \( \hat{T} \) computed in the previous iteration, until the estimated transformation converges to the ground-truth. In this work, we propose a lightweight, efficient module called DoPE and use it as part of this iterative registration framework in between the feature extraction and feature matching layers.
ture extraction and matching layers. By doing so, we enhance the intra-set distinctiveness of the intermediate features of every iteration and achieve increasingly more accurate point-to-point correspondences.

3.3. Feature Ambiguity in Learning-based Point Cloud Registration

In this sub-section, we identify and analyze the contributing factors to the above-mentioned indistinguishable feature problem. In doing so, we demonstrate that lack of intra-set distinctiveness has been a huge bottleneck to the latest GNN-based registration architectures and describe the motivations for DoPE’s design.

Indistinguishable features in GNNs are largely manifested in two ways: first, it has been empirically shown that a wide variety of GNN is prone to the oversmoothing problem [6, 17, 1], a phenomenon whereby repeated application of the message propagation step in each GNN layer renders all graph nodes to converge towards similar features across the entire point cloud, as displayed in Figure 1a. This is severely detrimental to the downstream registration procedure as it makes it challenging for the network to find the best match for a given point of point set \( X \) if there is very little difference amongst all of the point features of point set \( Y \), and vice versa. The oversmoothing problem in GNNs can be alleviated to an extent by attention mechanisms [10, 9] thanks to their data-dependent, attention-weighted aggregation scheme, with a similar argument having been made for oversmoothing in CNNs for images [27, 21]. This is corroborated in Figure 1, where the features of the backbone processed by both local (Figure 1b) and non-local (Figure 1c) attention are significantly more distinctive than the raw vanilla features (Figure 1a). More notably, the increase in intra-set distinctiveness is accompanied by a meaningful improvement to the registration error.

However, attention-based feature aggregation fails to address what is a more subtle contributing factor to indistinguishable GNN features: structural ambiguity, a problem induced by the (partial) translation invariance encoded in the message propagation step of prominent GNN backbones [18, 19, 24, 5] that renders points that are separate, but encode locally similar structures of the point cloud to have near-identical features. For example, the point feature enclosed by the red circle in Figure 1b and 1c remain similar to the corresponding points that lie on the other legs of the chair. Such features can potentially hamper the registration process due to spurious matches between locally similar structures in the source and target point clouds. This is where the positional embedding in the proposed DoPE module comes into the equation; it provides the network with additional cues to distinguish between the four legs and to be able to appropriately match the constituent points of each leg to the correct counterpart in the target point cloud. As shown in Figure 1d and 1e, the positional embedding further enhances the intra-set distinctiveness to its upper limit, yielding significant improvements to registration accuracy. This demonstrates that although structural ambiguity isn’t as conspicuous of a phenomenon as is oversmoothing, it has been the biggest bottleneck to the latest GNN-based registration architectures, an observation that has not only motivated our work but one that will hopefully motivate the future design of GNNs for registration.

4. Proposed Method

We now present DoPE, a novel positional embedding unit used as part of an iterative registration framework (Figure 2) that disambiguates the backbone GNN features for more effective point cloud registration. In Section 4.1, we introduce the constituent operations of the DoPE module and its properties; in Section 4.2, we describe how the iterative application of the DoPE module converges the network to the positional equilibrium, a fixed point with high registration accuracy where matching points are given essentially identical positional embeddings; finally, in Section 4.3, we outline the loss function that we use to further encourage feature disambiguation in an end-to-end manner.

4.1. DoPE Module

Joint-origin Update The DoPE module disambiguates the backbone features via positional embeddings followed by a non-local attention operation. In order to feed positional embeddings to both the source and target point clouds \( X \) and \( Y \), we first compute the joint origin, the origin of a shared coordinate system where the positional information is defined. In iteration \( t \) of the forward pass of the registration pipeline, we update the joint-origin \( \tilde{z}^{(t)} \) as the center-of-mass of the union of \( X^{(t)} = \{ x_{1}^{(t)}, x_{2}^{(t)}, \cdots, x_{J}^{(t)} \} \) and \( Y = \{ y_{1}, y_{2}, \cdots, y_{K} \} \):

\[
\tilde{z}^{(t)} = \frac{1}{J + K} \left( \sum_{i=1}^{J} x_{i}^{(t)} + \sum_{i=1}^{K} y_{j} \right).
\]

The joint-origin is an essential aspect of the DoPE module. Computing positional embeddings with respect to a shared coordinate frame allows the network to remain aware of the relative spatial orientation of the point clouds during iterative registration. Furthermore, updating the joint-origin as the centroid of the combined point clouds enforces a special type of translation invariance, whereby the DoPE module feeds the same set of positional embeddings to point cloud pairs with the same relative configuration in 3D space, but located in different absolute positions. As a result, our registration architecture is invariant to such variations that may occur within the dataset itself or even throughout the iterative process, thereby narrowing down the space of regis-
each point as follows: attention [16] can intuitively increase the distinctiveness of
dition to adding the positional embedding into the back-
equilibrium first computes the positional embedding of each
features, the positional equilibrium explicitly combines the
Feature Disambiguation To disambiguate the backbone
features, the positional equilibrium explicitly combines the
positional information with respect to the joint-origin with
scenarios faced by the DoPE module and benefiting
the overall training procedure.

**Feature Disambiguation** To disambiguate the backbone
features, the positional equilibrium explicitly combines the
positional information with respect to the joint-origin with
the backbone high-level features via self-attention. After esti-
ming joint origin \( \vec{z}^{(t)} \) after \( t \)-th iteration, the positional
equilibrium first computes the positional embedding of each
point, and then add to the backbone features as follows:

\[
F_{x_i} \leftarrow F_{x_i} + \mathcal{M}(x_i - \vec{z}^{(t)}), \quad F_{y_j} \leftarrow F_{y_j} + \mathcal{M}(y_j - \vec{z}^{(t)}),
\]

(3)

where \( \mathcal{M} \) is a shared multi-layer perceptron (MLP). In addi-
tion to adding the positional embedding into the back-
bone features, aggregating other contextual cues via self-
attention [16] can intuitively increase the distinctiveness of
each point as follows:

\[
F_{x_i} \leftarrow \sum_{j \in S(X)} \alpha_{x_{ij}} F_{x_j}, \quad F_{y_j} \leftarrow \sum_{i \in S(Y)} \alpha_{y_{ij}} F_{y_i},
\]

(4)

where \( \alpha_{x_{ij}} = \text{Softmax}_i(q_{x_{ij}}^T k_{x_{ij}}) \) is the similarity be-
between \( i \)-th query and \( j \)-th key features of \( x \) by setting
\( q_{x_{ij}} = F_{x_i} \) and \( k_{x_{ij}} = F_{x_j} \). \( \alpha_{y_{ij}} \) is also defined as iden-
tical to \( \alpha_{x_{ij}} \). Because the feature update process in Eq. (4)
has quadratic space-time complexity when the features are
aggregated from all points, we sample the features to be ag-
gregated (i.e. we sample the keys) to model the long-range
dependencies with light-weight memory and computation.
Inspired by [4], we use random sampling to uniformly select
\( |S| \) points from which the features are aggregated. Since
\( N \gg |S| \), the complexity of DoPE is low with the order of
\( O(N \cdot |S|) \).

**4.2. Positional Equilibrium**

In the iterative registration framework outlined in Fig-
ure 2, the joint origin and the correspondence matrix are al-
ternately refined such that, the DoPE module feeds increa-
singly consistent positional information to would-be cor-
responding pairs and enhances the resulting correspondence
predictions. In turn, the rigid transformation estimated from
this correspondence matrix updates the joint-origin used to
provide positional embeddings in the next iteration. For ex-
ample, in Figure 3, we conduct point cloud registration on
source (blue-green) and target (pink) point clouds of the air-
plane. Assume that \( x_j \) and \( y_k \) are the corresponding pair of
the source and target points located in the engine part of
the airplane. We denote \( x_j^0 \) as the initial point of \( x_j \). At the
beginning of the registration process (Figure 3b,) the posi-
tional information of \( x_j^0 \) and \( y_k \) is not so close to each other,
leading to the mismatch in Figure 3c. Because the posi-
tional information of \( x_j^0 \) and \( y_k \) is closer, the network
predicts more correct correspondence matrix (Figure 3d.)
Within only a few iterations, \( x_j^{t} \) converges to \( y_k \), indica-
tive of the positional equilibrium where the positional em-
bleddings given to matching pairs become essentially iden-
tical, as shown in Figure 3e.

**4.3. Loss Function**

To encourage the network to learn distinctive feature de-
scriptors, we adopt a loss function based on the equilibrium-
state correspondence matrix of our architecture. For the ide-
ally distinctive feature descriptors, the feature descriptors
should have high similarity between matching pairs and
should have low similarity between non-matching pairs.
Let assume that the correspondence matrix outputted from
the feature matching layer is as follows:

\[
P = \{p_{jk}\}^{J \times K}, \quad 0 \leq p_{jk} \leq 1
\]

(5)

Without loss of generality, \( p_{jk} \) is scaled similarity between
\( F_{x_j} \) and \( F_{y_k} \). Because each element of the equilibrium-state
correspondence matrix, \( p_{jk} \), represents zero for all \( j \) and \( k \)
except that \( x_j \) and \( y_k \) are matching pair (i.e. \( p_{jk} = 1 \) if
\( x_j \) and \( y_k \) are matching pair and \( p_{jk} = 0 \) if not) which
are identical to the elements of ground-truth correspondence
matrix, we thus supervise our network to learn ground-truth
correspondences as:

\[
\mathcal{L}_{corr} = -\sum_{j=1}^{J} \log(p_{jk}^*),
\]

(6)
where $y_{k^*}$ is the ground truth target point corresponded with source point $x_k$. The correspondence loss, $L_{corr}$, is a cross-entropy loss used in [5, 2]. We employ this loss specifically because the correspondence loss further strengthens the distinctiveness of feature descriptors between non-matching pairs and the matchability between matching pairs ($\sum_{k=1}^{K} p_{jk} = 1$).

The correspondence loss is prone to the overfitting problem because it guides the point-to-point correspondence about all matching pairs. To alleviate the overfitting, we additionally add the transformation loss $L_{trans}$ to the total loss of network as the regularization term as follows:

$$L_{total} = L_{corr} + \lambda L_{trans}, \quad (7)$$

where $L_{trans} = \|\hat{R}^T R - I\|^2 + \|\hat{t} - t^*\|^2$ is also used by existing methods [18, 19, 24, 2, 25]. As mentioned in Section 3.2, we fuse our DoPE into the other registration methods and compute the total loss at every $t$-th iteration.

5. Experiments

For following sections, we evaluate the performance of our DoPE by inserting it into the various baseline registration methods with the various datasets. The learning-based registration methods such as Deep Closest Point (DCP [18]), PRNet [19], RPM-Net [24], Iterative Distance-Aware Matrix convolution (IDAM [5]), DeepGMR [25], and Deep Global Registration (DGR [2]) are considered as our baseline networks. The object-level datasets (ModelNet40 [22], ScanObjectNN [15]) and the scene-level dataset (3DMatch [26]) are employed. We reported the superior result between the original paper and our reruns for the results of baseline methods.

5.1. Object-level Dataset

The ModelNet40 dataset consists of 12,311 models from 40 categories. As with the DCP [18], we divide 12,311 models into 9,843 for training and 2,468 for testing. During training, we pick rotation $R$ and translation $t$ randomly at $[0, 45^\circ]$ and $[-0.5, 0.5]$, respectively. Then, we measure the root-mean-square error (RMSE), and the mean absolute error (MAE) between the ground-truth ($R^*, t^*$), and the predicted ($R, t$). The rotation measurements is the degree. We use DCP, PRNet, IDAM, RPM-Net as our baseline methods of the object-level dataset.

**Full Data** Full data setting implies that source and target point clouds have exact one-to-one correspondences for all points. Specifically, we follow the experimental settings of DCP [18] for the full data of ModelNet40, sampling 1,024 points from the surface of each model of ModelNet40. Table 1 shows the registration results on full data without any perturbation of points (full+clean dataset). DoPE remarkably enhances the performance of baseline methods. RPM-Net+DoPE especially achieves dozens of times performance improvement in terms of rotational metric compared to baseline performance. Moreover, we also investigate the robustness to Gaussian noise in Table 2. We add random Gaussian noise with the distribution of $\mathcal{N}(0, 0.01)$ to each point of source and target point clouds independently so that some points could not have exact matching points. Table 2 shows that DoPE still noticeably improves the registration performance across all benchmarks in the presence of noise, although the positional information of matching points could not be identical due to the noise.

**Partial Data** Because point cloud registration mostly occurs between partially overlapped point clouds in real-world applications, we generate the partially overlapped data of ModelNet40. We randomly pick one point from each source and target point clouds and then compute 768 nearest-neighbor points out of the full 1,024 points as in PRNet [19]. Table 3 shows the results of partial+clean ModelNet40. DoPE enhances all baseline’s performance even better than the existing SOTA performance. Especially, DCP with DoPE surprisingly outperforms other baseline methods about the all performance metrics, although DCP does not explicitly handle the partially-overlapped registration problem. This indicates that the distinctiveness of features is significant for point cloud registration. We also experiment with Gaussian noise (partial-noisy dataset) and show that our proposed module is also robust to the noise in Table 4.
Unfamiliar Data To compare each method’s generalizability, we test on ScanObjectNN dataset using the models that are trained on full-noisy ModelNet40 dataset in Table 4. Because the ScanObjectNN is a real-world point cloud object dataset extracted from scanned indoor scene data consisting of 15,000 objects categorized into 15 categories, ScanObjectNN contains objects which are significantly different from the synthetic CAD dataset-MoNet40. Because the ScanObjectNN data are extracted from scanned scene data, background elements or parts of nearby objects could be included in each object data, and even the density of point clouds is non-uniform. The results on ScanObjectNN is shown in Table 5, indicating that the DoPE module works well even in the unfamiliar data.

5.2. Scene-level Dataset

For the scene-level dataset, we use the real-world indoor 3DMatch dataset [26], which consists of 3D point cloud pairs from eight different scenes with ground truth transformations estimated from RGB-D reconstruction and use DGR as our baseline. A single point is subsampled within each 5cm voxel to generate point clouds with uniform density. We follow the train/test split and the standard procedure to generate pairs with at least 30% overlap for training and testing. Different from the error metrics in synthetic dataset, we use the error metric as DGR does for fair comparison: rotation error (RE) as \( \arccos \frac{Tr(R^T R) - 1}{2} \), translational error (TE) as \( \| t - \hat{t} \|_2 \), and recall. Recall is the ratio of successful registrations, and we define a successful registration as the case in which RE is less than 15 degrees, and TE is less than 0.3m. Table 6 summarizes the experimental results of the 3DMatch dataset. DGR+DoPE outperforms the baseline significantly, demonstrating that DoPE can be scalable well in the scene-level dataset.
6. Analysis

6.1. Ablation Study

Effects of Different Components Table 7 shows that the performance improvement induced by DoPE is largely attributed to the use of positional embeddings w.r.t. the joint origin rather than mere vanilla adoptions of existing operations such as self-attention and positional embeddings. Table 7 demonstrates that the use of the joint-origin is not arbitrary and is in fact an crucial aspect of DoPE’s design.

Effects of $L_{corr}$

Table 8 shows that the increase in registration accuracy is mainly attributed to DoPE as opposed to the correspondence loss $L_{corr}$. Simply adding the correspondence loss to the baselines actually hampers performance, whereas the DoPE module improves registration accuracy by enhancing the point-wise correspondences and thereby enabling the backbone network to better leverage the correspondence loss.

6.2. Visualization Analysis

In Figure 4, we visualize influence of the DoPE module through its point-to-point correspondence matrix. Specifically, we plot the line connecting points $x_j$ and $y_k$ of source and target point clouds, respectively, when the matching score between the two points exceeds 0.01 (i.e., $j_k > 0.01$). The line is colored green if the correspondence is correct and red if it is incorrect. By comparing the correspondence matrices of DCP and DCP+DoPE for several ModelNet40 objects, we see that DCP+DoPE predicts more accurate matching matrices than does DCP for all objects.

Table 7: Effects of each component on registration performance. LA: Local attention, N-LA: Non-local attention, Indiv.: Positional embeddings computed w.r.t. individual origin (centroid of each point cloud), Joint: Positional embeddings computed w.r.t. joint-origin.

<table>
<thead>
<tr>
<th>GNN(DCP)</th>
<th>LA</th>
<th>N-LA</th>
<th>Indiv.</th>
<th>Joint</th>
<th>MAE(R)</th>
<th>MAE(h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>5.34</td>
<td>0.022</td>
</tr>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>4.74</td>
<td>0.022</td>
</tr>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>2.38</td>
<td>0.021</td>
</tr>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>1.04</td>
<td>0.006</td>
</tr>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>4.45</td>
<td>0.020</td>
</tr>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>3.86</td>
<td>0.025</td>
</tr>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>0.80</td>
<td>0.006</td>
</tr>
</tbody>
</table>

Table 8: Effects of $L_{corr}$

<table>
<thead>
<tr>
<th>Models</th>
<th>$L_{corr}$</th>
<th>MAE(R)</th>
<th>MAE(h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCP[18]</td>
<td>No</td>
<td>4.45</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>5.61</td>
<td>0.03</td>
</tr>
<tr>
<td>DCP+DoPE</td>
<td>No</td>
<td>1.33</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>0.80</td>
<td>0.006</td>
</tr>
<tr>
<td>PRNet[19]</td>
<td>No</td>
<td>1.45</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>1.72</td>
<td>0.018</td>
</tr>
<tr>
<td>PRNet+DoPE</td>
<td>No</td>
<td>0.31</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>0.20</td>
<td>0.002</td>
</tr>
<tr>
<td>RPMNet[24]</td>
<td>No</td>
<td>0.84</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>1.41</td>
<td>0.012</td>
</tr>
<tr>
<td>RPMNet+DoPE</td>
<td>No</td>
<td>0.14</td>
<td>0.0013</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>0.12</td>
<td>0.0011</td>
</tr>
</tbody>
</table>

Table 9: Efficiency on various methods

The leftmost object in Figure 4 shows in greater detail that the lines on the horizontal stabilizer represent exact one-to-one matching between source and target point clouds in DCP+DoPE whereas the lines in DCP represent ambiguous and incorrect matching. Through these qualitative visualization results, we further illustrate the problems induced by oversmoothing and structural ambiguity of the GNN backbone, and how DoPE alleviates their effects.

6.3. Efficiency Analysis

We estimate the efficiency of various models using the number of network parameters, the number of FLOPs, and the inference time. The FLOPs and inference time are estimated for processing one pair of input point clouds. We use the same hyper-parameter settings as reported by each method. Table 9 demonstrates that the the DoPE module incurs little additional complexity compared to the baseline models, indicating the potential scaleability of DoPE to models that handle large data settings.

7. Conclusion

In this paper, we call to attention the shortcomings of GNN-based features for registration, namely the indistinguishable feature problem associated with oversmoothing and structural ambiguity. These constituent issues motivate DoPE, a novel positional embedding module that significantly enhances the intra-set distinctiveness of the per-point features generated by prominent GNN backbones, and hence the resulting point-to-point correspondences. DoPE computes positional information with respect to the joint-origin of the combined point clouds, iteratively refining the joint-origin and the correspondence matrix until convergence to an equilibrium where the positional embedding for both point clouds become essentially identical. We demonstrate that the DoPE module significantly increases the registration performance across all combinations of experimental settings. Furthermore, we hope that our analysis of the indistinguishable features problem motivates the future design of a stand-alone GNN backbone specifically tailored to point cloud registration.
References


