SE3ET: SE(3)-Equivariant Transformer for Low-Overlap Point Cloud Registration

Anonymous Authors[1]

Abstract

Partial point cloud registration is a challenging problem, especially when the robot undergoes a large transformation, causing a significant initial pose error and a low overlap. This work proposes exploiting equivariant learning from 3D point clouds to improve registration robustness. We propose SE3ET, an SE(3)-equivariant registration framework that employs equivariant point convolution and equivariant transformer design to learn expressive and robust geometric features. We tested the proposed registration method on indoor and outdoor benchmarks where the point clouds are under arbitrary transformations and low overlapping ratios. We also provide generalization tests and run-time performance.

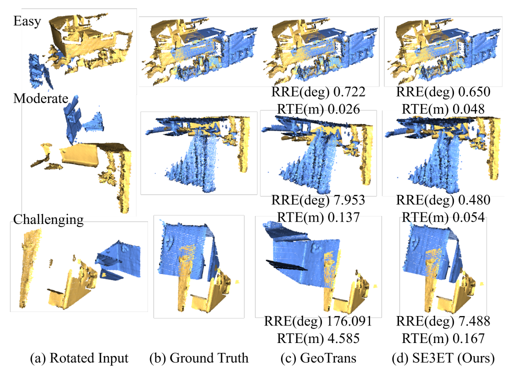

Figure 1. SE3ET can register two low-overlap point clouds with significant rotations and translations. This qualitative result is performed on rotated 3DLoMatch, where the first row is an easy example (28.72 % overlapping ratio with multiple overlapping surfaces), the middle row is a moderate example (10.51 % overlapping ratio with multiple overlapping surfaces), and the last row is a challenging example (26.75 % overlapping ratio with only one overlapping surface).

1. Introduction

Point cloud registration is a fundamental problem in computer vision and robotics. It aims to find the optimal transformation estimation between two point clouds. Recent works such as Predator (Huang et al., 2021) and GeoTransformer (Qin et al., 2022) look into point cloud registration with partial-to-partial situations, especially when two point clouds have a low overlap rate.

However, these methods are not optimized for cases where the point clouds are under significant initial pose error, which is common in robotic scenarios. The current state-of-the-art works' limitations are demonstrated in Figure 1.

YOHO (Wang et al., 2022a) applies equivariant feature learning in the point cloud registration task by using icosahedral-group convolution to learn rotation-equivariant point descriptors. However, the framework is not optimized for low-overlap registration.

Few existing learning-based methods consider arbitrary transformation situations in the network architecture. For example, YOHO (Wang et al., 2022a) applies equivariant feature learning in the point cloud registration task by using icosahedral-group convolution to learn rotation-equivariant point descriptors. However, the framework can be further optimized for low-overlap registration.

In this work, we propose SE3ET, a more robust and efficient SE(3)-equivariant low-overlap point cloud registration framework. The main contributions are summarized as follows:

- Our feature learning based on equivariant convolutions and transformers improves the robustness against point clouds with low overlap and large pose changes.
- We propose four designs of equivariant transformers, each offering unique properties and potential uses.

2. Related Work

Correspondence-based point cloud registration methods (Yu et al., 2021; Qin et al., 2022; Fu et al., 2021; Wang & Solomon, 2019; Cao et al., 2021; Yew & Lee, 2020), generally follow a two-step procedure consisting of correspondence matching and finding the optimal transformation that aligns the pairs. While correspondence-based methods have been widely used for point cloud registration, they are susceptible to incorrect correspondence pairs. We propose that using equivariant features can determine more accurate correspondence pairs within our correspondence-based method.

A common challenge in the point cloud registration task is the partial overlap situation, where the overlapping ratio between the two point clouds is low, especially when dealing with noisy data, occlusions, or outliers. While various research (Huang et al., 2021; Wang et al., 2022b; Mei et al., 2023; Vaswani et al., 2017; Wang et al., 2022b; Qin et al., 2022; Zhu et al., 2021; Yew & Lee, 2022; Yu et al., 2021; Qin et al., 2022; Huang et al., 2022; Yu et al., 2023) progress in tackling the partial overlap challenge, when it comes to substantial rotational-wise transformation, most of the above methods rely on data augmentation during training to increase robustness to arbitrary transformation.

Recent studies (Deng et al., 2021; Chen et al., 2021; Zhu et al., 2023) focus on learning equivariant features from 3D point clouds offer valuable insights into the relevance of network architecture for point cloud registration across various transformation scenarios. Moreover, (Fuchs et al., 2020) and (Chatzipantazis et al., 2022) research into the integration of equivariant learning within the transformer mechanism demonstrates its applicability to tasks such as classification and reconstruction. However, despite these advancements, there remains a scarcity of learning-based methods for point cloud registration that adequately address arbitrary transformation situations within the network architecture. YOHO ((Wang et al., 2022a)), a notable example, employs rotation-equivariant feature learning and has been extended to RoReg ((Wang et al., 2023)). While these methods effectively leverage equivariant feature learning for point cloud registration, further optimization is needed to bolster processing robustness, particularly in scenarios with low overlap.

Building upon these varied algorithmic approaches, this paper proposes a novel framework that potentially and effectively resolves issues of low overlapping point clouds in registration procedures robust to arbitrary transformation. In this paper, the optimal performance of E2PN (Zhu et al., 2023) in learning SE(3)-equivariant features has been harnessed by incorporating it in our feature learning process. Improvements in feature capabilities are achieved via the transformer mechanism's implicit learning of the overlapping points. By leveraging equivariant and invariant features, a more robust registration of point clouds makes arbitrary transformation possible.

3. Methodology

Point cloud registration aims to compute an optimal rigid transformation $\mathbf{T} = \{(\mathbf{R}, \mathbf{t}) \mid \mathbf{R} \in \mathrm{SO}(3), \mathbf{t} \in \mathbb{R}^3\} \in \mathrm{SE}(3)$ that aligns two given partially overlapped point clouds $\hat{\mathcal{P}} = \{\mathbf{p_i} \in \mathbb{R}^3 \mid i = 1, \ldots, N\}$ and $\hat{\mathcal{Q}} = \{\mathbf{q_j} \in \mathbb{R}^3 \mid j = 1, \ldots, M\}$.

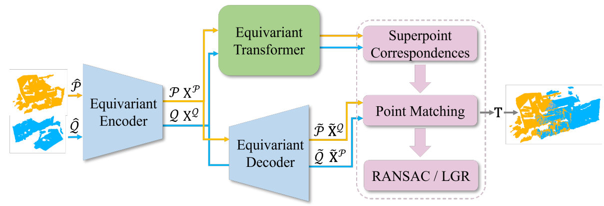

Figure 2. The proposed point cloud registration framework includes a SE(3)-equivariant feature encoder and decoder and an equivariant transformer design for learning the point correspondences of superpoints.

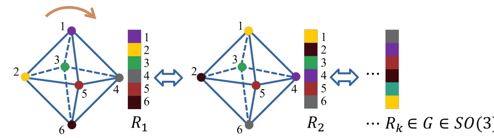

Figure 3. The geometric (octahedron shape) and algebraic (color bricks) illustration of using permutation to recover discretized rotation group. Each color represents the feature of one anchor (vertex), and the different order of the combination of features represents the discretized rotation defined in the network. If the octahedron is rotated 90 degrees along the arrow direction, the order of the features changes accordingly. The discretization of the rotation groups is derived from the permutation of the discrete anchors. For $A = 6$, the rotation group contains 24 rotations.

The proposed framework is shown in Figure 2, we will introduce each components in the following subsections.

3.1. Equivariant Feature Encoder and Decoder

We adopt GeoTransformer's multi-stage encoder-decoder structure (Qin et al., 2022), following a Feature Pyramid Network (Lin et al., 2017) to extract multi-scale features from downsampled point clouds. Diverging from conventional approaches, we employ E2PN (Zhu et al., 2023), an SE(3)-equivariant point convolutional network, for SE(3)-equivariant convolution. E2PN discretizes the 3D rotation group $\mathrm{SO}(3)$ into a polyhedral rotation group G, approximating $\mathrm{SE}(3)$ as $\mathbb{R}^3 \times G$, where feature maps reside on $\mathbb{R}^3 \times V$, with V representing polyhedral vertices. These maps, equivariant to $\mathrm{SE}(3)$ transformations, are computed efficiently via quotient-space convolution. We refer to polyhedron vertices as *anchors* for equivariant feature learning. Point clouds downscaled to the coarsest level (termed "superpoints") are denoted as $\mathcal{P} \in \mathbb{R}^{N' \times 3}$, $\mathcal{Q} \in \mathbb{R}^{M' \times 3}$, with their equivariant features $\mathbf{X}^{\mathcal{P}} \in \mathbb{R}^{N' \times A \times C}$, $\mathbf{X}^{\mathcal{Q}} \in \mathbb{R}^{M' \times A \times C}$, where N', M' denote the downsampled points, C represents the feature channels, and $A = |V|$ denotes the anchor size.

By design, we can express each rotation in the discretized rotation group G using a permutation of the vertices V. A geometric illustration is in Figure 3.

3.2. Equivariant Transformer Design

We propose equivariant self-attention and cross-attention modules to enhance features of superpoints by gathering information across various spatial locations and orientations. Self-attention modules facilitate feature interaction within a point cloud, while cross-attention modules enable feature communication between pairs of point clouds.

3.2.1. Equivariant Self-Attention (ESA) Module

We introduce an equivariant self-attention module, extending self-attention methods in (Qin et al., 2022) while ensuring equivariance to $\mathrm{SE}(3)$ transformations. This module allows consistent behavior under rigid body transformations.

Equivariant superpoint features $\mathbf{X}^{\mathcal{P}}$ serve as query, key, and value inputs. We follow (Qin et al., 2022) to use geometric information to provide geometric structure embedding $\mathbf{P}^{\mathcal{P}} \in \mathbb{R}^{N' \times N' \times C}$.

For a point indexed i in $\mathcal{P}$ at anchor coordinate r, attention between such elements is computed using trainable weight matrices $\mathbf{W}^Q$, $\mathbf{W}^K$, and $\mathbf{W}^P \in \mathbb{R}^{C \times C}$. Output features $\mathbf{x}^{\mathcal{P}}_{\mathtt{SA},ir}$ are obtained via weighted summation over points. A subsequent feed-forward layer, as in (Vaswani et al., 2017), refines learned features, resulting in $\mathbf{X}^{\mathcal{P}}_{\mathtt{SA}} \in \mathbb{R}^{N' \times A \times C}$.

$$\mathbf{a}_{\mathtt{SA},ir,jr} = \frac{(\mathbf{x}_{ir}\mathbf{W}^Q)(\mathbf{x}_{jr}\mathbf{W}^K + \mathbf{p}_{i,j}\mathbf{W}^P)^\intercal}{\sqrt{C}}, \tag{1}$$

$$\mathbf{x}^{\mathcal{P}}_{\mathtt{SA},ir} = \sum_{j=1}^{N'} \mathrm{Softmax}_j(\mathbf{a}_{\mathtt{SA},ir,jr})\mathbf{x}_{jr}\mathbf{W}^V, \tag{2}$$

Compared with conventional self-attention among the point features, our equivariant self-attention module allows different attention values at different anchor coordinates for the same pair of points, similar to multi-head self-attention, but with the equivariant property.

3.2.2. Invariant Cross-Attention Module

The cross-attention mechanism integrates two separate inputs, typically observed in point cloud registration tasks with paired point clouds $\mathcal{P}$ and $\mathcal{Q}$. We focus on operations within $\mathcal{P}$; analogous operations apply to $\mathcal{Q}$.

Invariant Cross-Attention (ICA). In this module, attention is conducted on $\mathrm{SE}(3)$-invariant features derived from equivariant self-attention features $\mathbf{X}^{\mathcal{P}}_{\mathtt{SA}}$ and $\mathbf{X}^{\mathcal{Q}}_{\mathtt{SA}}$ for $\mathcal{P}$ and $\mathcal{Q}$, respectively. Pooling on the anchor dimension yields invariant features $\mathbf{X}^{\mathcal{P}}_{\mathtt{SA-inv}} \in \mathbb{R}^{N' \times C}$ and $\mathbf{X}^{\mathcal{Q}}_{\mathtt{SA-inv}} \in \mathbb{R}^{M' \times C}$.

Features $\mathbf{X}^{\mathcal{P}}_{\mathtt{SA-inv}}$ serve as queries, and $\mathbf{X}^{\mathcal{Q}}_{\mathtt{SA-inv}}$ as keys. The attention value between point i in $\mathcal{P}$ and point j in $\mathcal{Q}$ is computed using trainable weight matrices.

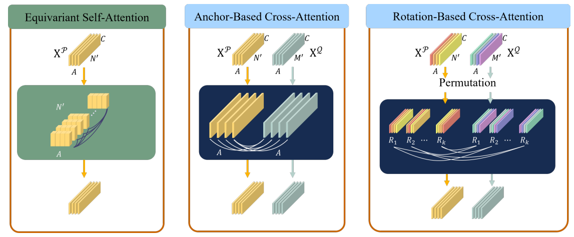

Figure 4. An illustration of the proposed equivariant self-attention, equivariant anchor-based, and rotation-based cross-attention modules.

$$\mathbf{a}_{\mathtt{ICA},i,j} = \frac{(\mathbf{x}^{\mathcal{P}}_{\mathtt{SA-inv},i}\mathbf{W}^Q)(\mathbf{x}^{\mathcal{Q}}_{\mathtt{SA-inv},j}\mathbf{W}^K)^\intercal}{\sqrt{C}} \tag{3}$$

Depending on the desired output—equivariant or invariant—the values $\mathcal{Q}$ can be equivariant features $\mathbf{X}^{\mathcal{Q}}_{\mathtt{SA}}$ or invariant features $\mathbf{X}^{\mathcal{Q}}_{\mathtt{SA-inv}}$. The collection of equivariant output features is denoted as $\mathbf{X}^{\mathcal{P}}_{\mathtt{ICA}} \in \mathbb{R}^{N' \times A \times C}$, while invariant output features are denoted as $\mathbf{X}^{\mathcal{P}}_{\mathtt{ICA-inv}} \in \mathbb{R}^{N' \times C}$. Take $\mathbf{X}^{\mathcal{P}}_{\mathtt{ICA},i}$ for point i in $\mathcal{P}$ as an example,

$$\mathbf{x}^{\mathcal{P}}_{\mathtt{ICA},i} = \sum_{j=1}^{M'} \mathrm{Softmax}_j(\mathbf{a}_{\mathtt{ICA},i,j})\mathbf{x}^{\mathcal{Q}}_{\mathtt{SA},j}\mathbf{W}^V \tag{4}$$

3.2.3. Equivariant Cross-Attention Modules

The two types of *Equivariant* Cross-Attention modules are featured to learn equivariant attention over the rotation anchor dimension and the point dimension to enable equivariant learning for various equivariant features.

Equivariant Anchor-Based Cross-Attention (ACA). In this module, we learn cross-attention scores for each anchor dimension. Operating on equivariant self-attention features $\mathbf{X}^{\mathcal{P}}_{\mathtt{SA}}$ and $\mathbf{X}^{\mathcal{Q}}_{\mathtt{SA}}$, for $\mathcal{P}$ and $\mathcal{Q}$ respectively, $\mathbf{X}^{\mathcal{P}}_{\mathtt{SA}}$ serves as query and $\mathbf{X}^{\mathcal{Q}}_{\mathtt{SA}}$ as key and value. Attention computation between two points in $\mathcal{P}$ and $\mathcal{Q}$ is performed via trainable weight matrices.

$$\mathbf{a}_{\mathtt{ACA\_raw},ir,js} = \frac{(\mathbf{x}^{\mathcal{P}}_{\mathtt{SA},ir}\mathbf{W}^Q)(\mathbf{x}^{\mathcal{Q}}_{\mathtt{SA},js}\mathbf{W}^K)^\intercal}{\sqrt{C}} \tag{5}$$

Normalization is applied on both anchor and spatial dimensions to stabilize feature learning. Softplus ensures the non-negativity of raw attention, preventing false point-matching from distracting the anchor-wise attention between two point clouds. Average pooling on the point dimension focuses on anchor features from both inputs. Global anchor-based attention, denoted as $\mathbf{a}_{\mathtt{ACA\_anchor}}$, captures correlations between anchors across all points, normalized over the anchor dimension. Softmax normalization is applied to the spatial dimension.

$$\mathbf{a}_{\texttt{ACA\_anchor},r,s} = \frac{1}{N'M'} \sum_{i=1}^{N'} \sum_{j=1}^{M'} \text{Softplus}(\mathbf{a}_{\texttt{ACA\_raw},ir,js}) \tag{6}$$

$$\mathbf{a}_{\texttt{ACA\_norm\_anchor},r,s} = \frac{\mathbf{a}_{\texttt{ACA\_anchor},r,s}}{\sum_{s=1}^{A} \mathbf{a}_{\texttt{ACA\_anchor},r,s}} \tag{7}$$

$$\mathbf{a}_{\texttt{ACA\_norm\_spatial},ir,js} = \text{Softmax}_j(\mathbf{a}_{\texttt{ACA\_raw},ir,js}) \tag{8}$$

$$\mathbf{a}_{\texttt{ACA},ir,js} = \mathbf{a}_{\texttt{ACA\_norm\_spatial},ir,js} \mathbf{a}_{\texttt{ACA\_norm\_anchor},r,s} \tag{9}$$

Output features for each point in $\mathcal{P}$ are obtained through weighted summation over points in $\mathcal{Q}$, followed by a feed-forward layer.

$$\mathbf{x}^{\mathcal{P}}_{\texttt{ACA},ir} = \sum_{j=1}^{M'} \sum_{s=1}^{A} \mathbf{a}_{\texttt{ACA},ir,js} \mathbf{x}^{\mathcal{Q}}_{\texttt{SA},js} \mathbf{W}^V, \tag{10}$$

ACA preserves equivariance, as the attention value depends solely on feature contents, regardless of anchor indices. This design ensures equivariant preservation, which is crucial for consistent behavior under rotations.

Equivariant Rotation-Based Cross-Attention (RCA). In this version of cross-attention, we learn the cross-attention scores for each discretized rotation in the rotation group.

First, we use the permutation layer from E2PN (Zhu et al., 2023), which is mentioned in Section 3.1, to reconstruct feature maps defined on the discretized rotation group G from the features defined on anchors V. We denote the permuted feature corresponding to the rotation $g \in G$ as:

$$\mathbf{x}^{\mathcal{Q}}_{\texttt{Permute},jg} = \text{Permute}_g(\{\mathbf{x}^{\mathcal{Q}}_{\texttt{SA},js}\}_{s=1,\dots,A}) \tag{11}$$

After obtaining the feature corresponding to the rotation groups, the raw attention between two input features can be computed as:

$$\mathbf{a}_{\texttt{RCA\_raw},i,jg} = \frac{(\mathbf{x}^{\mathcal{P}}_{\texttt{SA},i}\mathbf{W}^Q)(\mathbf{x}^{\mathcal{Q}}_{\texttt{Permute},jg}\mathbf{W}^K)^\intercal}{\sqrt{C}} \tag{12}$$

We carry out normalization in both the rotation and spatial dimensions for consistent feature learning. Rotation-wise, we calculate the normalized global rotation attention:

$$\mathbf{a}_{\texttt{RCA\_rot},g} = \frac{1}{N'M'} \sum_{i=1}^{N'} \sum_{j=1}^{M'} \text{Softplus}(\mathbf{a}_{\texttt{RCA\_raw},i,jg}) \tag{13}$$

$$\mathbf{a}_{\texttt{RCA\_norm\_rot},g} = \frac{\mathbf{a}_{\texttt{RCA\_rot},g}}{\sum_{g=1}^{|P|} \mathbf{a}_{\texttt{RCA\_rot},g}} \tag{14}$$

Table 1. Evaluation on the rotated 3DLoMatch benchmark with 5000 sample points and 1k iterations when performing RANSAC. RR, IR, and FMR represent Registration Recall, Inlier Ratio, and Feature Matching Recall. The best is shown in bold font. * results are from (Wang et al., 2022a).

	RR (%) ↑	IR (%) ↑	FMR (%) ↑
Predator (Huang et al., 2021)*	57.7	26.2	75.7
GeoTransformer (Qin et al., 2022)	62.4	34.5	84.0
YOHO (Wang et al., 2022a)*	65.9	26.4	79.2
Ours	**69.0**	**43.0**	**87.1**

Spatial-wise, we use the softmax function on the j dimension to normalize across the spatial dimension.

$$\mathbf{a}_{\texttt{RCA\_norm\_spatial},i,jg} = \text{Softmax}_j(\mathbf{a}_{\texttt{RCA\_raw},i,jg}) \tag{15}$$

We multiply Equation (14) and Equation (15) to obtain attention for both the spatial and rotation dimensions.

$$\mathbf{a}_{\texttt{RCA},i,jg} = \mathbf{a}_{\texttt{RCA\_norm\_spatial},i,jg}\, \mathbf{a}_{\texttt{RCA\_norm\_rot},g} \tag{16}$$

The output feature for the i'th point in $\mathcal{P}$, denoted as $\mathbf{x}^{\mathcal{P}}_{\texttt{RCA},i} \in \mathbb{R}^{A \times C}$, can be written as:

$$\mathbf{x}^{\mathcal{P}}_{\texttt{RCA},i} = \sum_{g=1}^{|P|} \sum_{j=1}^{M'} \mathbf{a}_{\texttt{RCA},i,jg}(\mathbf{x}^{\mathcal{Q}}_{\texttt{Permute},jg}\mathbf{W}^V) \tag{17}$$

After passing through the feed-forward layer, we denote the collection of the output features as $\mathbf{X}^{\mathcal{P}}_{\texttt{RCA}} \in \mathbb{R}^{N' \times A \times C}$.

In this section, we proposed different equivariant self-attention and cross-attention modules. Each model aims to enhance its respective features, fortifying the robustness of point cloud registration tasks. We believe such an inclusion leverages the full potential of the equivariant features, thereby optimizing the overall performance of our proposed framework.

4. Experimental Results

We conducted experiments on rotated 3DLoMatch (Huang et al., 2021; Wang et al., 2022a) benchmark, which contains point cloud pairs with 10% to 30% overlap and arbitrary rotations. As shown in Table 1, our method performs the best among all the baselines, showing superior robustness to low overlap and arbitrary rotations.

5. Conclusion

We have designed *SE3ET*, a low-overlap point cloud registration framework that exploits SE(3)-equivariant feature learning. We show that SE(3)-equivariant features improve robustness to large transformations in the low-overlap situation. Our experimental results on Rotated 3DLoMatch show that the proposed method achieves promising results in challenging scenarios.

References

Cao, A.-Q., Puy, G., Boulch, A., and Marlet, R. Pcam: Product of cross-attention matrices for rigid registration of point clouds. In *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 13229–13238, 2021.

Chatzipantazis, E., Pertigkiozoglou, S., Dobriban, E., and Daniilidis, K. Se(3)-equivariant attention networks for shape reconstruction in function space. *arXiv preprint arXiv:2204.02394*, 2022.

Chen, H., Liu, S., Chen, W., Li, H., and Hill, R. Equivariant point network for 3d point cloud analysis. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 14514–14523, 2021.

Deng, C., Litany, O., Duan, Y., Poulenard, A., Tagliasacchi, A., and Guibas, L. J. Vector neurons: A general framework for so (3)-equivariant networks. In *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 12200–12209, 2021.

Fu, K., Liu, S., Luo, X., and Wang, M. Robust point cloud registration framework based on deep graph matching. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 8893–8902, 2021.

Fuchs, F., Worrall, D., Fischer, V., and Welling, M. Se(3)-transformers: 3d roto-translation equivariant attention networks. *Proc. Advances Neural Inform. Process. Syst. Conf.*, 33:1970–1981, 2020.

Huang, S., Gojcic, Z., Usvyatsov, M., Wieser, A., and Schindler, K. Predator: Registration of 3d point clouds with low overlap. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 4267–4276, 2021.

Huang, X., Wang, Y., Li, S., Mei, G., Xu, Z., Wang, Y., Zhang, J., and Bennamoun, M. Robust real-world point cloud registration by inlier detection. *Comput. Vis. Image Understanding*, 224:103556, 2022.

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2117–2125, 2017.

Mei, G., Tang, H., Huang, X., Wang, W., Liu, J., Zhang, J., Van Gool, L., and Wu, Q. Unsupervised deep probabilistic approach for partial point cloud registration. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 13611–13620, 2023.

Qin, Z., Yu, H., Wang, C., Guo, Y., Peng, Y., and Xu, K. Geometric transformer for fast and robust point cloud registration. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 11143–11152, 2022.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Wang, H., Liu, Y., Dong, Z., and Wang, W. You only hypothesize once: Point cloud registration with rotation-equivariant descriptors. In *Proceedings of the 30th ACM International Conference on Multimedia*, pp. 1630–1641, 2022a.

Wang, H., Liu, Y., Hu, Q., Wang, B., Chen, J., Dong, Z., Guo, Y., Wang, W., and Yang, B. Roreg: Pairwise point cloud registration with oriented descriptors and local rotations. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2023.

Wang, Y. and Solomon, J. M. Deep closest point: Learning representations for point cloud registration. In *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 3523–3532, 2019.

Wang, Y., Yan, C., Feng, Y., Du, S., Dai, Q., and Gao, Y. Storm: Structure-based overlap matching for partial point cloud registration. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2022b.

Yew, Z. J. and Lee, G. H. Rpm-net: Robust point matching using learned features. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 11824–11833, 2020.

Yew, Z. J. and Lee, G. H. Regtr: End-to-end point cloud correspondences with transformers. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 6677–6686, 2022.

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

Yu, J., Ren, L., Zhang, Y., Zhou, W., Lin, L., and Dai, G. Peal: Prior-embedded explicit attention learning for low-overlap point cloud registration. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 17702–17711, 2023.

Zhu, L., Liu, D., Lin, C., Yan, R., Gómez-Fernández, F., Yang, N., and Feng, Z. Point cloud registration using representative overlapping points. *arXiv preprint arXiv:2107.02583*, 2021.

Zhu, M., Ghaffari, M., Clark, W. A., and Peng, H. E2pn: Efficient se (3)-equivariant point network. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2023.