

SHReg: Strictly Rotation-Equivariant Point Cloud Registration via Spherical Harmonics

Chongjian Wang¹  and Junjie Gao² *

¹ College of Mathematics and Systems Science, Shandong University of Science and Technology, Qingdao, China

202311080223@sdust.edu.cn

² School of Artificial Intelligence, Shandong Women's University, Jinan, China
gjjsdnu@163.com

Abstract. Point cloud registration critically depends on local features that are both distinctive and robust to arbitrary 3D rotations. Existing learning-based methods typically approximate rotation invariance via fragile local reference frames or extensive data augmentation, providing only empirical invariance and often degrading under unseen rotational transformations. In this paper, we propose SHReg, a strictly rotation-equivariant point cloud registration framework grounded in the representation theory of $SO(3)$. By representing local geometric features as irreducible representations of $SO(3)$, SHReg guarantees exact equivariance under arbitrary rotations without relying on local reference frames. Built upon a spherical-harmonics-based equivariant backbone, SHReg jointly learns rotation-invariant descriptors for robust correspondence matching and rotation-equivariant features that preserve fine-grained orientation information. The preserved equivariant structure enables each correspondence to directly hypothesize a rigid transformation, reducing reliance on large-scale hypothesis sampling in conventional RANSAC-based pipelines and leading to improved robustness under challenging rotational variations. Extensive experiments on 3DMatch, 3DLoMatch, and KITTI demonstrate that SHReg consistently outperforms state-of-the-art methods in registration accuracy, particularly under large rotational perturbations.

Keywords: Point Cloud Registration · Rotation-Invariant Networks · Spherical Harmonics

1 Introduction

Point cloud registration, which estimates a rigid transformation to align two partially overlapped point clouds, is a fundamental problem in 3D computer vision. It has been widely applied in autonomous driving [23, 41], robot localization [31], and 3D reconstruction [18, 25, 35]. Among various paradigms, the feature-based registration framework remains one of the most actively studied directions, typically consisting of two key steps: local descriptor extraction

* Corresponding author.

[7–9, 12, 28, 29, 32, 33, 36, 39] and robust transformation estimation [3, 11, 16, 19, 22]. A long-standing contradiction in this task lies in the fact that point clouds can exhibit large pose variations, while the extracted descriptors are expected to be invariant to such changes, especially global rotations. Therefore, learning discriminative yet rotation-invariant features is crucial for reliable correspondence establishment and accurate pose recovery.

To obtain rotation-invariant descriptors, existing methods can be roughly categorized into patch-wise [2, 8, 9, 14, 37] and scene-wise feature extractors. Patch-wise approaches learn descriptors from local neighborhoods and usually rely on explicit preprocessing to eliminate orientation ambiguity. Two representative strategies are: (1) using handcrafted rotation-invariant statistics such as distances, angles, or point-pair features [8, 9]; and (2) estimating a local reference frame [2, 14] to normalize the local patch before feature extraction. Although these designs can reduce sensitivity to rotations, the former often suffers from information loss due to aggressive invariant summarization, while the latter is vulnerable to imperfect reference frame estimation under partial overlap. In addition, patch-wise methods are typically inefficient because they process a large number of local patches independently, without effectively sharing intermediate computations across the whole point cloud.

Scene-wise extractors [7, 30, 32, 33, 36, 39], on the other hand, compute dense descriptors for all points in a unified forward pass and have achieved strong performance in recent learning-based pipelines. However, most scene-wise methods are built upon rotation-sensitive backbones (e.g., KPConv [30], FCGF [7]) and enforce rotation invariance mainly through extensive rotation augmentation during training [39]. This practice introduces several drawbacks. First, the continuous rotation space is difficult to cover sufficiently, leading to fragility when encountering unseen orientations. Second, learning rotation invariance [32, 33] implicitly often increases model capacity requirements, since the network must “memorize” invariance patterns from data. Third, forcing a rotation-sensitive architecture to become invariant can compromise its ability to focus on learning structural distinctiveness, which limits descriptor discriminability under large pose changes.

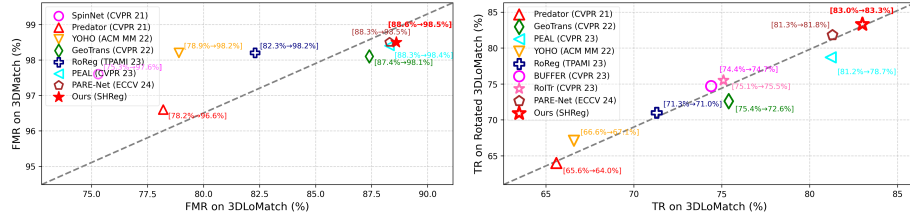


Fig. 1: Experimental results on 3DMatch, 3DLoMatch, and Rotated 3DLoMatch. Our method consistently achieves higher feature matching recall (FMR) and transformation recall (TR), demonstrating strong robustness to rotations.

In this paper, we propose SHReg, a strictly rotation-equivariant point cloud registration framework built upon spherical harmonic representations. The core idea is to explicitly model the $SO(3)$ symmetry of point clouds at the representation level and leverage it consistently in both feature extraction and transfor-

mation estimation. Unlike augmentation-driven designs that attempt to approximate rotation invariance from data, SHReg enforces equivariance by construction, enabling the network to produce representations that transform predictably under arbitrary rotations while preserving fine-grained geometric structures. As shown in Fig. 1, SHReg consistently achieves higher feature matching recall (FMR), which measures the percentage of scan pairs with sufficiently accurate correspondences, and transformation recall (TR), which evaluates the success rate of pose estimation, across 3DMatch, 3DLoMatch, and their rotated variants.

First, we introduce a spherical harmonic-based strictly rotation-equivariant feature extraction network. Spherical harmonics naturally form a complete and orthogonal basis for functions defined on the sphere and exhibit well-defined transformation behavior under rotations. By embedding local neighborhood geometry into the spherical harmonic domain, SHReg ensures that intermediate representations follow strict equivariant transformation rules. Rotation-invariant descriptors are then derived from these equivariant representations without sacrificing structural distinctiveness. In contrast to rotation-sensitive backbones trained with extensive augmentation, our design builds rotation consistency directly into the architecture, allowing the network to focus on learning intrinsic geometric characteristics rather than compensating for pose variations.

Second, we design a correspondence-driven rigid transformation estimation strategy that directly exploits the consistency of strictly rotation-equivariant features. Once reliable correspondences are established through invariant matching, the equivariant representations provide structured geometric cues for accurate rotation recovery, followed by stable translation estimation. Unlike conventional estimators such as RANSAC [11] and LGR [27], which rely on sampling correspondence triplets to generate transformation hypotheses, our method proposes hypotheses from single correspondences guided by rotation-consistent features. As shown in Fig. 2, our estimator reaches near-optimal performance with only 0.5k hypotheses, while LGR requires around 2k hypotheses to stabilize and RANSAC needs substantially more samples to achieve comparable results. These results demonstrate that leveraging strictly rotation-equivariant features enables more efficient and reliable pose hypothesis generation, leading to improved registration recall with significantly fewer hypotheses.

Overall, our main contributions are summarized as follows:

- A strictly rotation-equivariant registration framework that explicitly models $SO(3)$ symmetry to resolve the contradiction between pose variation and descriptor invariance.

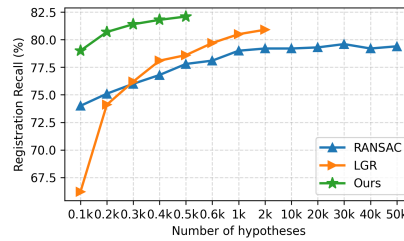


Fig. 2: Comparison of our hypothesis proposer, RANSAC and LGR on 3DLoMatch.

- A spherical-harmonics-based backbone that guarantees strict rotation equivariance and produces discriminative rotation-invariant descriptors without relying on approximate invariance.
- A transformation estimation strategy enabled by strict $SO(3)$ equivariance, allowing each correspondence to determine a rigid transformation and reducing the RANSAC search space for improved registration accuracy.

2 Related Work

Patch-wise and Scene-wise Feature Extractors. Early learning-based point cloud descriptors [2, 8, 9, 14, 37] were mainly patch-wise extractors, motivated by traditional handcrafted descriptors [10, 15, 28, 29] and constrained by the limited scalability of early point cloud convolution techniques. These methods focused on designing rotation-invariant local representations, typically through local reference frame (LRF) construction [2, 14] or handcrafted invariant encodings such as point pair features (PPFs) [8, 9]. While such strategies remove rotational ambiguity at the input level, LRF estimation is known to be unstable under noise, density variation, and partial observations, and invariant encodings often suppress orientation-dependent geometric cues that are potentially discriminative. With the development of more efficient point cloud backbones [7, 30], scene-wise feature extractors [5, 12, 27, 38] have become dominant due to their scalability and contextual reasoning capability. These methods often rely on data augmentation to improve robustness to rotations and focus on low-overlap matching via attention mechanisms and coarse-to-fine pipelines. However, augmentation-based training only provides approximate invariance and does not offer theoretical guarantees under arbitrary 3D rotations. To address this limitation, several works [32, 33, 36, 39] introduce rotation-invariant or rotation-equivariant architectures based on group convolutions, vector neurons, or invariant convolution operators. Group-based approaches discretize $SO(3)$ and extract features under multiple rotated copies of the input, leading to high computational cost and memory overhead, while lightweight equivariant models often compromise expressive power. In contrast, SHReg models local geometry using irreducible representations of $SO(3)$ via spherical harmonics, ensuring strict rotation equivariance by construction without LRF estimation, rotation discretization, or heavy augmentation, thereby preserving both efficiency and geometric fidelity.

Robust Transformation Estimators. Given noisy correspondences containing many outliers, robust transformation estimation remains a fundamental challenge in point cloud registration. RANSAC [11] is widely adopted due to its simplicity and general applicability; however, it generates hypotheses by randomly sampling correspondence triplets, which results in a cubic search space and requires a large number of iterations for reliable convergence. Recent learning based estimators [3, 6, 19, 21] improve robustness by modeling spatial consistency among correspondences, enabling more effective inlier identification or correspondence reweighting before applying SVD based transformation estimation. Although these approaches reduce the number of spurious hypotheses,

they still fundamentally depend on multiple correspondences to recover rigid transformations. A few works [34] attempt to estimate rotations from a single correspondence through direct regression, yet such strategies lack explicit geometric constraints and often generalize poorly to unseen rotations. In contrast, SHReg exploits strictly rotation equivariant features governed by irreducible representations of $SO(3)$, allowing each correspondence to directly produce a rigid transformation hypothesis in closed form. By embedding transformation structure into the feature representation itself, SHReg substantially reduces the hypothesis search space while remaining compatible with standard global verification procedures, resulting in improved efficiency and robustness under large rotational variations.

3 Method

3.1 Problem Statement

Given two partially overlapped point clouds

$$P = \{p_i \in \mathbb{R}^3 \mid i = 1, \dots, N\}, \quad Q = \{q_j \in \mathbb{R}^3 \mid j = 1, \dots, M\},$$

our goal is to estimate a rigid transformation

$$T = (R, t), \quad R \in SO(3), \quad t \in \mathbb{R}^3,$$

that aligns P to Q . We propose SHReg, a registration framework based on a strictly $SO(3)$ -equivariant spherical-harmonics network $f(\cdot)$, whose output satisfies

$$f(R \circ X) = \rho(R) f(X), \quad R \in SO(3),$$

where $\rho(\cdot)$ denotes the $SO(3)$ irrep action, implemented by Wigner- D matrices [26], on feature representations. Such rotation-equivariant features couple local structure information and orientation information induced by R , which can be decoupled into rotation-invariant descriptors for correspondence matching and rotation-equivariant features for transformation estimation.

The overall framework is illustrated in Fig. 3. Given P and Q , a hierarchical spherical-harmonics backbone extracts both patch-level and point-level features as $SO(3)$ irreps, and derives rotation-invariant descriptors for coarse-to-fine correspondence estimation. Then, the feature-based hypothesis proposer leverages the matched rotation-equivariant features $\tilde{F}_P$ and $\tilde{F}_Q$ to generate multiple reliable $SE(3)$ hypotheses from individual correspondences in closed form. The best hypothesis is selected through geometric verification and refined to produce the final alignment (R^*, t^*) .

3.2 Strictly $SO(3)$ -Equivariant Feature Backbone

Spherical-Harmonic Irreps Representation A core component of SHReg is to encode point-wise features as a direct sum of irreducible representations

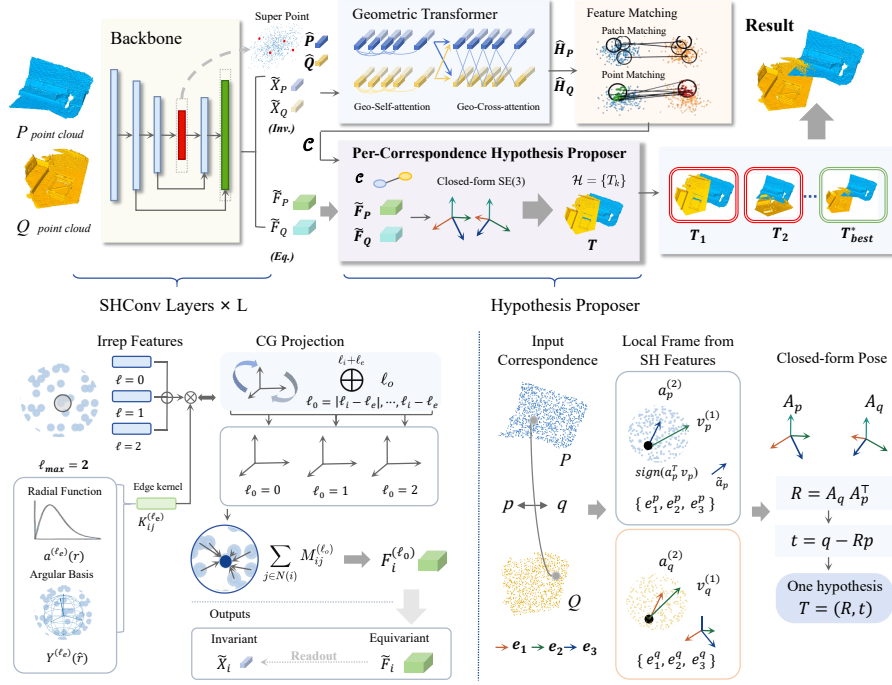


Fig. 3: Framework of SHReg. A spherical-harmonics backbone extracts rotation-invariant descriptors and rotation-equivariant features. Correspondences are obtained in a coarse-to-fine manner. Each correspondence generates one $SE(3)$ hypothesis, and the best hypothesis is selected as the final transformation.

(irreps) of $SO(3)$. For each point p_i , we decompose its feature into type- ℓ blocks:

$$F_i = \bigoplus_{\ell=0}^{\ell_{\max}} F_i^{(\ell)}, \quad F_i^{(\ell)} \in \mathbb{R}^{C_\ell \times (2\ell+1)}. \quad (1)$$

Here, $\oplus$ denotes the direct sum of $SO(3)$ irrep blocks rather than ordinary summation; each block contains C_ℓ channels with representation dimension $2\ell + 1$. The order $\ell = 0$ corresponds to scalar channels and $\ell > 0$ corresponds to higher-order geometric tensors. Under any rotation $R \in SO(3)$ applied to the input point cloud, each block transforms equivariantly:

$$F_i^{(\ell)}(R \circ P) = F_i^{(\ell)}(P) D^{(\ell)}(R)^\top, \quad (2)$$

where $D^{(\ell)}(R)$ is the Wigner- D matrix of order ℓ . This irrep structure provides strict $SO(3)$ equivariance without pose normalization. We provide a rigorous proof in the supplementary material that our spherical-harmonics encoding, tensor-product with Clebsch–Gordan projection, linear neighborhood aggregation, and scalar-gated nonlinearity jointly preserve $SO(3)$ -equivariance across layers.

$SO(3)$ -Equivariant Spherical-Harmonic Convolution Given a point cloud P , we construct a neighborhood $\mathcal{N}_i$ for each point p_i . For an edge (i, j) , we define the relative displacement

$$r_{ij} = p_j - p_i, \quad r = \|r_{ij}\|, \quad \hat{r}_{ij} = \frac{r_{ij}}{r}. \quad (3)$$

The unit direction $\hat{r}_{ij}$ is lifted to an equivariant angular basis by spherical harmonics $Y^{(\ell)}(\hat{r}_{ij})$, while the distance r is used to generate learnable radial weights:

$$a^{(\ell)}(r) = \text{MLP}_\ell(\phi(r)), \quad (4)$$

where $\phi(\cdot)$ denotes a radial basis function (RBF) expansion. The edge kernel is therefore given by $K_{ij}^{(\ell_e)} = a^{(\ell_e)}(r)Y^{(\ell_e)}(\hat{r}_{ij})$, where the radial MLP only takes the scalar distance r , while $Y^{(\ell_e)}$ provides the angular equivariant basis.

We define message passing via a tensor-product convolution followed by a Clebsch–Gordan (CG) projection:

$$M_{ij}^{(\ell_o)} = \sum_{\ell_i, \ell_e} C_{\ell_i, \ell_e \rightarrow \ell_o} \left(F_j^{(\ell_i)} \otimes (a^{(\ell_e)}(r_{ij}) Y^{(\ell_e)}(\hat{r}_{ij})) \right), \quad (5)$$

where $C_{\ell_i, \ell_e \rightarrow \ell_o}$ is the fixed CG projection from type- ℓ_i and type- ℓ_e irreps to a valid type- ℓ_o output irrep, with $|\ell_i - \ell_e| \leq \ell_o \leq \ell_i + \ell_e$. Only radial functions and channel mixing weights are learnable. Aggregating messages over the neighborhood yields the updated feature:

$$F_i^{(\ell_o)'} = \sum_{p_j \in \mathcal{N}_i} M_{ij}^{(\ell_o)}. \quad (6)$$

We use invariant $\ell = 0$ gates to modulate higher-order channels, so the scalar-gated nonlinearity preserves the corresponding $SO(3)$ transformation rules. Stacking such layers forms a hierarchical backbone that outputs multi-resolution rotation-equivariant features $\tilde{F}$, from which rotation-invariant descriptors $\tilde{X}$ are further derived for matching.

Rotation-Invariant Descriptor Readout For correspondence matching, we remove the representation dimension within each irrep block using an invariant norm:

$$X_{i,c}^{(\ell)} = \sum_{m=-\ell}^{\ell} |F_{i,c}^{(\ell,m)}|^2, \quad (7)$$

where $c = 1, \dots, C_\ell$ denotes the channel index. Concatenating invariants across all orders produces the final descriptor:

$$d_i = \text{Concat}(\{X_i^{(\ell)}\}_{\ell=0}^{\ell_{\max}}). \quad (8)$$

The descriptor dimension before readout layers is therefore $\sum_{\ell=0}^{\ell_{\max}} C_\ell$. Since $D^{(\ell)}(R)$ is orthogonal, the ℓ_2 norm in Eq. (7) is preserved under arbitrary rotations, and thus d_i is strictly rotation-invariant. A derivation of the strict invariance of Eqs. (7)–(8) is provided in the supplementary.

3.3 Coarse-to-Fine Correspondence Matching

Since the backbone produces both patch-level and point-level features, we adopt a coarse-to-fine matching strategy to suppress non-overlapping regions and establish accurate correspondences.

Superpoint Matching We first downsample the input point clouds into sparse superpoints $\hat{P}$ and $\hat{Q}$. Their rotation-invariant features $\hat{X}_P$ and $\hat{X}_Q$ are fed into a Geometric Transformer module, which captures global context via iterative self- and cross-attention. The refined superpoint features $\hat{H}_P$ and $\hat{H}_Q$ are then used to compute a similarity matrix S . The top- k reliable superpoint correspondences are selected as

$$\hat{C} = \{(\hat{p}_{x_i}, \hat{q}_{y_i})\}. \quad (9)$$

Point Matching For each matched superpoint pair, we further search point-level correspondences within their associated groups $G_{x_i}^P$ and $G_{y_i}^Q$. Following [24], we disentangle feature similarity and saliency by introducing a matchability head W_m and a saliency head W_s .

The similarity matrix is computed as

$$M_{x_i, y_i} = (W_m X_{x_i}^P)^\top (W_m X_{y_i}^Q) / \sqrt{D}, \quad (10)$$

where D denotes the feature dimension. The saliency scores are predicted by

$$\sigma_{x_i}^P = \text{Sigmoid}(W_s X_{x_i}^P), \quad \sigma_{y_i}^Q = \text{Sigmoid}(W_s X_{y_i}^Q). \quad (11)$$

We obtain a soft assignment matrix

$$Z_i = \sigma_{x_i}^P \sigma_{y_i}^Q \varphi(M_{x_i, y_i}), \quad (12)$$

where φ denotes softmax normalization. Finally, the top-ranked point correspondences across all matched superpoint pairs are selected:

$$C = \{(\tilde{p}_{x_j}, \tilde{q}_{y_j})\}. \quad (13)$$

3.4 Feature-Based Hypothesis Proposer

Given a single correspondence ($p \leftrightarrow q$), we generate one $SE(3)$ hypothesis $T = (R, t)$ in closed form from the matched rotation-equivariant features. The key idea is to construct a local orthonormal frame at each point using different irreps: $\ell = 2$ captures a stable principal axis, while $\ell = 1$ resolves its orientation.

Principal axis from $\ell = 2$. We first aggregate $\ell = 2$ features across channels:

$$u_p = w_2^\top \mathbf{F}_p^{(2)}, \quad u_q = w_2^\top \mathbf{F}_q^{(2)}, \quad u_p, u_q \in \mathbb{R}^5. \quad (14)$$

Each vector is mapped to a symmetric trace-free tensor:

$$S_p = \Phi(u_p), \quad S_q = \Phi(u_q). \quad (15)$$

The dominant eigenvectors of S_p and S_q , denoted by a_p and a_q , define unsigned principal axes.

Axis orientation from $\ell = 1$. To resolve the sign ambiguity, we extract a direction from $\ell = 1$ features:

$$v_p = w_1^\top \mathbf{F}_p^{(1)}, \quad v_q = w_1^\top \mathbf{F}_q^{(1)}. \quad (16)$$

The axis orientation is determined by

$$\tilde{a}_p = \text{sign}(a_p^\top v_p) a_p, \quad \tilde{a}_q = \text{sign}(a_q^\top v_q) a_q. \quad (17)$$

Local frame construction. We construct right-handed local frames:

$$e_1 = \text{normalize}(\tilde{a}), \quad e_2 = \text{normalize}(v - (e_1^\top v)e_1), \quad e_3 = e_1 \times e_2, \quad (18)$$

forming orthonormal matrices $A_p, A_q \in SO(3)$.

Closed-form transformation. The rotation aligning the two frames is

$$R = A_q A_p^\top, \quad (19)$$

and the translation is computed as

$$t = q - Rp. \quad (20)$$

Therefore, each correspondence directly produces one $SE(3)$ hypothesis without multi-point minimal solvers or SVD [20]. We detail the frame construction from $\ell = 1, 2$ features and degenerate-case checks in the supplementary.

3.5 Hypothesis Verification and Selection

We adopt a local-to-global hypothesis verification scheme. Given the correspondence set $\mathcal{C}$, the local phase generates a hypothesis set $\mathcal{H} = \{T_k = (R_k, t_k)\}$ by applying the above closed-form solution to each correspondence. Since each hypothesis is inferred from matched rotation-equivariant features encoding local structural orientation, the resulting transformations are already locally consistent.

In the global phase, we select the transformation that maximizes inlier support over all correspondences:

$$T^* = \arg \max_{T_k \in \mathcal{H}} \sum_{(\tilde{p}_{x_j}, \tilde{q}_{y_j}) \in \mathcal{C}} \llbracket \|R_k \tilde{p}_{x_j} + t_k - \tilde{q}_{y_j}\|_2^2 < \tau \rrbracket, \quad (21)$$

where $\llbracket \cdot \rrbracket$ denotes the Iverson bracket and τ is the inlier threshold. Compared with sampling-based estimators such as RANSAC [11], SHReg generates one closed-form hypothesis per correspondence, substantially reducing the hypothesis search space while preserving high registration accuracy.

3.6 Loss Function

We follow PARE-Net [36] and train SHReg with three losses, $\mathcal{L} = \mathcal{L}_c + \mathcal{L}_f + \mathcal{L}_r$, including a superpoint matching loss $\mathcal{L}_c$, a point matching loss $\mathcal{L}_f$, and a contrastive rotation loss $\mathcal{L}_r$. Specifically, $\mathcal{L}_c$ supervises rotation-invariant superpoint descriptors using the overlap-aware circle loss, $\mathcal{L}_f$ optimizes fine-level correspondences by supervising the soft assignment matrix and saliency prediction within ground-truth superpoint pairs, and $\mathcal{L}_r$ regularizes rotation-equivariant features to be consistent under the ground-truth rotation via a margin-based contrastive objective. For completeness, we refer readers to PARE-Net [36] for the detailed formulations and training settings of these loss terms.

4 Experiments

To evaluate the effectiveness of SHReg, we compare it with state-of-the-art point cloud registration methods on 3DMatch [42], 3DLoMatch [17], and KITTI Odometry [13]. We also conduct ablation studies to analyze the contribution of each component. All experiments are conducted on a workstation with an RTX 4090 GPU and an Intel Xeon(R) Platinum 8352V CPU. More implementation details are provided in the Appendix.

4.1 Indoor Dataset: 3DMatch & 3DLoMatch

Dataset. 3DMatch [42] and 3DLoMatch [17] are indoor RGB-D benchmarks. 3DMatch contains 62 scenes, with 46, 8, and 8 scenes used for training, validation, and testing, respectively. It consists of point cloud pairs with overlap ratios greater than 30%. 3DLoMatch is used only for testing and contains more challenging pairs with overlap ratios between 10% and 30%. We follow the standard protocols [17, 36]. Unless otherwise specified, 5000 keypoints are sampled per fragment to establish correspondences.

Baselines. We compare SHReg with strong baselines from three categories. The first category includes rotation-sensitive scene-level matchers and descriptors, such as FCGF [7], Predator [17], CoFiNet [38], GeoTransformer [27], and PEAL [40]. The second category contains rotation-robust scene-level methods, including YOHO [32], RoReg [33], RoITR [39], and PARE-Net [36]. The third category consists of patch-level rotation-invariant descriptors, including SpinNet [2] and BUFFER [1]. For methods that require a robust estimator, we use the same Open3D RANSAC implementation with optimized hyperparameters for fair comparison.

Metrics. Following [17, 27], we evaluate registration performance using Inlier Ratio (IR), Feature Matching Recall (FMR), and Registration Recall (RR). A correspondence is considered correct if its residual distance under the ground-truth transformation is smaller than τ_c . IR denotes the ratio of correct correspondences, FMR is the percentage of scan pairs whose IR exceeds 5%, and RR measures the fraction of successfully registered pairs under threshold τ_r . For rotated datasets, we additionally report Rotation Error (RE), Translation Error

Table 1: Evaluation results on 3DMatch and 3DLoMatch. Methods with the RANSAC estimator are marked by $\diamond$, which exploit 5000 points to establish correspondences.

Method	Size (MB)	3DMatch				3DLoMatch			
		FMR (% $\uparrow$)	IR (% $\uparrow$)	RR (% $\uparrow$)	Time (s $\downarrow$)	FMR (% $\uparrow$)	IR (% $\uparrow$)	RR (% $\uparrow$)	Time (s $\downarrow$)
FCGF $\diamond$ [7]	8.76	94.7	31.1	82.8	0.12	59.4	9.8	38.0	0.13
SpinNet $\diamond$ [2]	<u>1.41</u>	97.6	47.5	88.6	9.85	75.3	20.5	59.8	9.03
YOHO [32]	12.38	98.2	64.4	90.8	2.81	78.9	25.9	66.0	2.62
Predator $\diamond$ [17]	7.43	96.6	58.0	89.0	0.64	78.2	26.7	64.4	0.47
GeoTrans [27]	9.83	98.1	70.9	92.4	0.18	87.4	43.5	74.3	0.17
RoReg [33]	12.71	98.2	<u>81.6</u>	93.0	2.27	82.3	39.6	70.1	2.10
BUFFER [1]	0.92	—	—	92.1	0.22	—	—	70.3	0.21
RoITr $\diamond$ [39]	10.10	98.0	82.4	91.9	0.36	89.2	54.6	74.1	0.34
PEAL [40]	9.83	98.4	71.0	94.2	1.46	88.3	46.0	78.8	1.19
PARE-Net [36]	3.84	<u>98.5</u>	76.9	<u>95.0</u>	<u>0.17</u>	88.3	47.5	<u>80.5</u>	<u>0.17</u>
Ours	9.52	98.5	78.6	95.4	0.28	<u>88.6</u>	<u>48.8</u>	82.4	0.28

(TE), and Transformation Recall (TR) following RoReg [33]. Detailed definitions of these metrics are provided in the Appendix.

Table 2: Evaluation results on 3DLoMatch and Rotated 3DLoMatch. For easy comparison of the impact of rotation, the numerical changes w.r.t. TR on the rotated dataset are annotated in the top right corner of each value. Methods with rotation invariant/equivariant designs are marked by * symbol.

Method	Size (MB)	3DLoMatch			Rotated 3DLoMatch		
		RE ($^\circ$ $\downarrow$)	TE (cm $\downarrow$)	TR (% $\uparrow$)	RE ($^\circ$ $\downarrow$)	TE (cm $\downarrow$)	TR (% $\uparrow$)
FCGF [7]	8.76	4.84	12.87	39.6	4.74	13.39	24.5 ^{-15.1}
Predator [17]	7.43	3.61	10.65	65.6	3.55	10.30	64.0 ^{-1.6}
GeoTrans [27]	9.83	2.91	8.71	75.4	2.94	8.85	72.6 ^{-2.8}
PEAL [40]	9.83	2.84	8.64	81.2	<u>2.86</u>	8.53	78.7 ^{-2.5}
YOHO* [32]	12.38	3.54	10.34	66.6	3.61	10.16	67.1 ^{+0.5}
RoReg* [33]	12.71	3.01	9.26	71.3	3.03	9.28	71.0 ^{-0.3}
BUFFER* [1]	0.92	3.03	9.86	74.4	3.02	9.99	74.7 ^{+0.3}
RoITr* [39]	10.10	2.95	9.03	75.1	2.97	9.08	75.5 ^{+0.4}
PARE-Net* [36]	<u>3.84</u>	<u>2.87</u>	<u>8.83</u>	<u>81.3</u>	2.84	<u>8.71</u>	<u>81.8</u> ^{+0.5}
Ours*	9.52	2.92	8.96	83.0	2.88	8.94	83.3 ^{+0.3}

Quantitative results. We report the results on 3DMatch and 3DLoMatch in Table 1. SHReg achieves consistently strong performance on both datasets, especially on 3DLoMatch where low overlap makes correspondence quality and hypothesis generation more challenging. Compared with rotation-sensitive methods such as Predator [17] and GeoTransformer [27], SHReg exhibits better robustness under hard cases since it does not rely on extensive rotation augmentation to approximate invariance. Compared with rotation-invariant/equivariant competitors such as YOHO [32], RoReg [33], RoITr [39], and PARE-Net [36], SHReg yields improved IR and RR while maintaining high efficiency. This advantage mainly comes from two aspects: (1) strictly $SO(3)$ -equivariant feature learning provides stable local structure encoding under arbitrary rotations, and (2) equivariant features allow transformation hypotheses to be proposed with significantly reduced hypothesis complexity compared with triplet-based sampling.

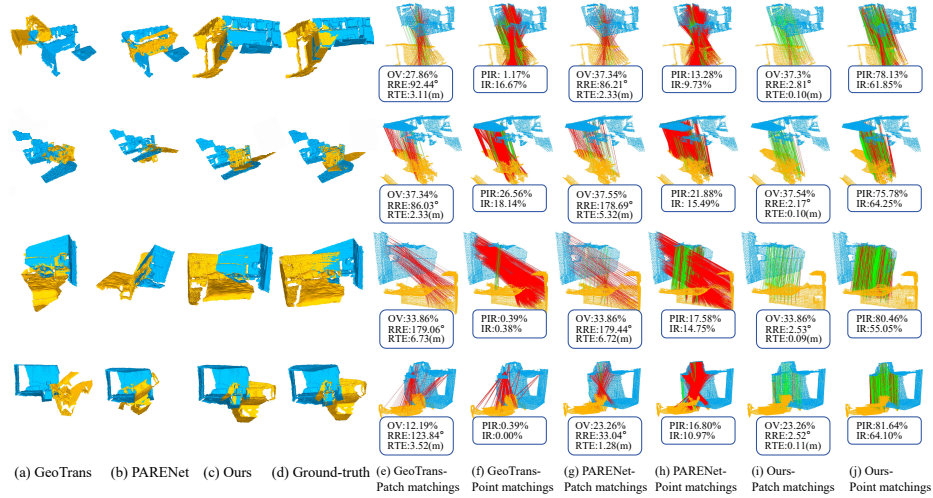


Fig. 4: Qualitative results. Our method can successfully align low-overlapped pairs with higher IR and PIR (Patch IR). Moreover, GeoTrans [27] and PARE-Net [36] incorrectly align point clouds with opposite orientations (the first and second rows), our method successfully addresses this issue because the equivariant features contain orientation information of local structure, alleviating this ambiguity.

Experiments on the Rotated Dataset. To test robustness against unseen rotations, we follow prior works [33, 36, 39] to evaluate all methods on the rotated 3DLoMatch benchmark, where arbitrary full-range rotations are independently applied to each fragment of a pair. This setting explicitly tests whether a method is intrinsically rotation-invariant/equivariant rather than relying on limited coverage of rotation augmentation. The results are shown in Table 2. Rotation-sensitive methods such as FCGF [7], Predator [17], GeoTransformer [27], and PEAL [40] present noticeable performance drops on the rotated benchmark, indicating that approximate invariance obtained by augmentation generalizes poorly to unseen rotations. In contrast, SHReg maintains stable performance and even shows slight improvements in TR, demonstrating that strict equivariance/invariance by design provides reliable robustness under severe pose variations. Moreover, SHReg surpasses previous rotation-robust methods such as YOHO [32], RoReg [33], RoITr [39] and PARE-Net [36], suggesting that spherical-harmonics-based irreps better preserve discriminative structural information.

Qualitative results. We present representative registration results in Fig. 4. SHReg produces more accurate patch-level and point-level correspondences and successfully aligns challenging fragment pairs with low overlap ratios. We observe that rotation-sensitive methods such as GeoTransformer [27] may incorrectly align scans with opposite orientations under severe pose variations. This ambiguity typically arises when point-to-point constraints alone are insufficient to disambiguate local structural symmetries, especially under large rotations. In contrast, SHReg successfully resolves such flipped alignments. This is be-

Table 3: Evaluation results on KITTI Odometry.

Method	Size(MB)	RE (°)	TE (cm)	TR (%)	Time (s)
FCGF [7]	8.76	0.30	9.5	96.6	–
D3Feat [4]	14.08	0.30	7.2	99.8	–
Predator [17]	22.77	0.27	6.8	99.8	0.77
SpinNet [2]	<u>1.41</u>	0.47	9.9	99.1	16.24
CoFiNet [38]	5.48	0.41	8.2	99.8	0.59
GeoTrans [27]	25.50	0.23	6.2	99.8	0.26
BUFFER [1]	0.92	0.26	7.1	99.8	0.27
PARE-Net [36]	2.08	0.23	<u>4.9</u>	99.8	0.21
Ours	24.82	0.23	4.7	99.8	<u>0.24</u>

cause the learned rotation-equivariant features explicitly encode the orientation of local geometric structures, allowing our hypothesis proposer to construct consistent local frames from matched correspondences and generate geometrically valid transformations. Additional qualitative results and failure case analysis are provided in the Appendix.

4.2 Outdoor Dataset: KITTI Odometry

Dataset. We further evaluate SHReg on the KITTI Odometry dataset [13], which contains outdoor LiDAR scans captured in urban driving environments. Following previous works [7, 27, 36], we use sequences 0–5 for training, 6–7 for validation, and 8–10 for testing. SHReg is trained on indoor 3DMatch and directly evaluated on KITTI without fine-tuning, demonstrating its cross-domain generalization ability.

Metrics. We adopt the standard metrics Rotation Error (RE), Translation Error (TE), and Transformation Recall (TR). Following [7, 27, 36], TR is computed with a rotation threshold of 5° and a translation threshold of 2m.

Quantitative results. Quantitative results on KITTI are reported in Table 3. On this benchmark, many methods have reached performance saturation, while SHReg achieves highly competitive accuracy with improved efficiency. These results validate that the proposed strictly equivariant representation can generalize well from indoor RGB-D scans to outdoor LiDAR scenarios.

4.3 Ablation Study

To better understand the contribution of each component in SHReg, we conduct ablation experiments on the 3DMatch and 3DLoMatch benchmarks. The quantitative results are summarized in Table 4.

Ablation of backbone. We first evaluate a rotation-sensitive baseline backbone that does not explicitly enforce rotational equivariance. We then introduce the key components of the proposed spherical-harmonics-based equivariant backbone, including spherical harmonic (SH) directional encoding, Clebsch–Gordan (CG) projection, gated nonlinearity, and rotation-invariant descriptor readout. The results show that SH representations consistently improve feature quality. In

Table 4: Ablation experiments of SHReg.

Component	Method	3DMatch			3DLoMatch		
		FMR (% $\uparrow$)	IR (% $\uparrow$)	RR (% $\uparrow$)	FMR (% $\uparrow$)	IR (% $\uparrow$)	RR (% $\uparrow$)
Backbone	Rotation-sensitive baseline (no SH/irreps)	98.0	71.0	93.6	86.7	41.5	78.2
	SH encoding w/o CG projection	98.2	73.5	94.2	87.5	43.5	79.3
	SH + CG, w/o Gate nonlinearity	<u>98.3</u>	<u>75.8</u>	<u>94.7</u>	<u>88.0</u>	<u>45.6</u>	<u>80.1</u>
	SH + CG + Gate, w/o invariant readout	97.6	68.0	92.8	84.0	36.5	75.5
	Full SHReg backbone	98.5	78.6	95.4	88.6	48.8	82.4
Pose Estimator	RANSAC (triplet sampling)	98.5	78.6	94.0	88.6	48.8	79.5
	LGR / patch-based estimator	98.5	78.6	<u>94.6</u>	88.6	48.8	<u>81.0</u>
	Ours (single-correspondence proposer)	98.5	78.6	95.4	88.6	48.8	82.4

particular, removing the invariant readout significantly degrades performance, since equivariant features still contain orientation-dependent components and are not directly suitable for correspondence matching. Removing CG projection or the gating mechanism also leads to noticeable drops, indicating the importance of valid irrep coupling and stable equivariant nonlinear transformations. The full SHReg backbone achieves the best performance on both datasets, demonstrating the effectiveness of strictly $SO(3)$ -equivariant feature learning.

Ablation of pose estimator. We further evaluate different pose estimation strategies using the same correspondence sets, including standard RANSAC with randomly sampled correspondence triplets, a patch-based estimator similar to LGR, and our single-correspondence hypothesis proposer. Since all estimators share identical correspondences, FMR and IR remain unchanged, while RR reflects the pose estimation quality. As shown in Table 4, our hypothesis proposer achieves the highest RR on both datasets, especially on 3DLoMatch. This is because low-overlap pairs make triplet-based sampling more likely to produce unstable hypotheses, whereas SHReg constructs transformations directly from individual correspondences using orientation information encoded in rotation-equivariant features. This reduces the reliance on extensive random sampling and improves pose estimation robustness.

5 Conclusion

In this paper, we presented SHReg, a point cloud registration framework built upon strictly $SO(3)$ -equivariant spherical-harmonic representations. By modeling local features as irreducible representations of $SO(3)$, SHReg guarantees consistent feature transformations under arbitrary rotations without relying on local reference frames or rotation augmentation. The proposed backbone produces both rotation-invariant descriptors for reliable correspondence matching and rotation-equivariant features that preserve structural orientation cues. Leveraging this property, we introduced a correspondence-driven hypothesis proposer that estimates rigid transformations directly from individual correspondences, significantly reducing the hypothesis search space compared with conventional sampling-based estimators. Extensive experiments on indoor and outdoor benchmarks demonstrate that SHReg achieves strong registration accuracy and robustness, particularly under large rotational variations.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China under Grant No. 62373164, the Project of Central Government Guides Local Program under Grant No. YDZX2024075, the Natural Science Foundation of Shandong Province Youth Program under Grant No. ZR2025QC2246Z, the Taishan Scholars Program of Shandong Province under Grant No. tsqn202507271, and the Taishan Experts Program under Grant No. tscy20241154.

References

1. Ao, S., Hu, Q., Wang, H., Xu, K., Guo, Y.: Buffer: Balancing accuracy, efficiency, and generalizability in point cloud registration. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 1255–1264 (2023), <https://api.semanticscholar.org/CorpusID:260055164>
2. Ao, S., Hu, Q., Yang, B., Markham, A., Guo, Y.: Spinnet: Learning a general surface descriptor for 3d point cloud registration. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 11748–11757 (2020), <https://api.semanticscholar.org/CorpusID:227151932>
3. Bai, X., Luo, Z., Zhou, L., Chen, H., Li, L., Hu, Z., Fu, H., Tai, C.L.: Pointdsc: Robust point cloud registration using deep spatial consistency. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 15854–15864 (2021), <https://api.semanticscholar.org/CorpusID:232168842>
4. 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. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 6358–6366 (2020), <https://api.semanticscholar.org/CorpusID:212628458>
5. Chen, H., Yan, P., Xiang, S., Tan, Y.: Dynamic cues-assisted transformer for robust point cloud registration. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 21698–21707 (2024), <https://api.semanticscholar.org/CorpusID:272722265>
6. Chen, Z., Sun, K., Yang, F., Tao, W.: Sc2-pcr: A second order spatial compatibility for efficient and robust point cloud registration. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 13211–13221 (2022), <https://api.semanticscholar.org/CorpusID:252571579>
7. Choy, C.B., Park, J., Koltun, V.: Fully convolutional geometric features. 2019 IEEE/CVF International Conference on Computer Vision (ICCV) pp. 8957–8965 (2019), <https://api.semanticscholar.org/CorpusID:201686687>
8. Deng, H., Birdal, T., Ilic, S.: Ppf-foldnet: Unsupervised learning of rotation invariant 3d local descriptors. ArXiv **abs/1808.10322** (2018), <https://api.semanticscholar.org/CorpusID:52131369>
9. Deng, H., Birdal, T., Ilic, S.: Ppfnet: Global context aware local features for robust 3d point matching. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition pp. 195–205 (2018), <https://api.semanticscholar.org/CorpusID:3703761>
10. Dong, Z., Dong, Z., Yang, B., Liu, Y., Liang, F., Li, B., Zang, Y.: A novel binary shape context for 3d local surface description. *Isprs Journal of Photogrammetry and Remote Sensing* **130**, 431–452 (2017), <https://api.semanticscholar.org/CorpusID:125337993>

11. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* **24**, 381–395 (1981), <https://api.semanticscholar.org/CorpusID:972888>
12. Gao, J., Dong, Q., Wang, R., Chen, S., Xin, S., Tu, C., Wang, W.: Oaaformer: Robust and efficient point cloud registration through overlapping-aware attention in transformer. *Journal of Computer Science and Technology* **39**, 755 – 770 (2023), <https://api.semanticscholar.org/CorpusID:264146030>
13. Geiger, A., Lenz, P., Urtasun, R.: Are we ready for autonomous driving? the kitti vision benchmark suite. 2012 IEEE Conference on Computer Vision and Pattern Recognition pp. 3354–3361 (2012), <https://api.semanticscholar.org/CorpusID:6724907>
14. Gojcic, Z., Zhou, C., Wegner, J.D., Wieser, A.: The perfect match: 3d point cloud matching with smoothed densities. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 5540–5549 (2018), <https://api.semanticscholar.org/CorpusID:53670229>
15. Guo, Y., Soheli, F., Bennamoun, Wan, J., Lu, M.: Rops: A local feature descriptor for 3d rigid objects based on rotational projection statistics. 2013 1st International Conference on Communications, Signal Processing, and their Applications (ICC-SPA) pp. 1–6 (2013), <https://api.semanticscholar.org/CorpusID:14670920>
16. Hu, S., Mellado, N., Mitra, N.J., Aiger, D.: Super 4pcs fast global pointcloud registration via smart indexing. *Computer Graphics Forum* **33** (2014), <https://api.semanticscholar.org/CorpusID:45860720>
17. Huang, S., Gojcic, Z., Usvyatsov, M.M., Wieser, A., Schindler, K.: Predator: Registration of 3d point clouds with low overlap. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 4265–4274 (2020), <https://api.semanticscholar.org/CorpusID:227209593>
18. Jaramillo, P., Sipiran, I.: Cultural heritage 3d reconstruction with diffusion networks. *ArXiv abs/2410.10927* (2024), <https://api.semanticscholar.org/CorpusID:273350603>
19. Jiang, H., Dang, Z., Wei, Z., Xie, J., Yang, J., Salzmann, M.: Robust outlier rejection for 3d registration with variational bayes. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 1148–1157 (2023), <https://api.semanticscholar.org/CorpusID:257921545>
20. Kabsch, W.: A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A* **32**, 922–923 (1976), <https://api.semanticscholar.org/CorpusID:97383637>
21. Lee, J., Kim, S., Cho, M., Park, J.: Deep hough voting for robust global registration. 2021 IEEE/CVF International Conference on Computer Vision (ICCV) pp. 15974–15983 (2021), <https://api.semanticscholar.org/CorpusID:237453312>
22. Li, J., Hu, Q., Ai, M.: Gesac: Robust graph enhanced sample consensus for point cloud registration. *ISPRS Journal of Photogrammetry and Remote Sensing* (2020), <https://api.semanticscholar.org/CorpusID:225238083>
23. Li, Y., Ma, L., Zhong, Z., Liu, F., Chapman, M.A., Cao, D., Li, J.: Deep learning for lidar point clouds in autonomous driving: A review. *IEEE Transactions on Neural Networks and Learning Systems* **32**(8), 3412–3432 (2020)
24. Lindenberger, P., Sarlin, P.E., Pollefeys, M.: Lightglue: Local feature matching at light speed. 2023 IEEE/CVF International Conference on Computer Vision (ICCV) pp. 17581–17592 (2023), <https://api.semanticscholar.org/CorpusID:259243929>

25. Liu, G., Rong, Y., Sheng, L.: Votehmr: Occlusion-aware voting network for robust 3d human mesh recovery from partial point clouds. *Proceedings of the 29th ACM International Conference on Multimedia* (2021), <https://api.semanticscholar.org/CorpusID:239011881>
26. Mattis, D.C.: Quantum theory of angular momentum (1981), <https://api.semanticscholar.org/CorpusID:115850593>
27. Qin, Z., Yu, H., Wang, C., Guo, Y., Peng, Y., Ilic, S., Hu, D., Xu, K.: Geotransformer: Fast and robust point cloud registration with geometric transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **45**, 9806–9821 (2023), <https://api.semanticscholar.org/CorpusID:257646283>
28. Rusu, R.B., Blodow, N., Beetz, M.: Fast point feature histograms (fpfh) for 3d registration. *2009 IEEE International Conference on Robotics and Automation* pp. 3212–3217 (2009), <https://api.semanticscholar.org/CorpusID:15022990>
29. Salti, S., Tombari, F., di Stefano, L.: Shot: Unique signatures of histograms for surface and texture description. *Comput. Vis. Image Underst.* **125**, 251–264 (2014), <https://api.semanticscholar.org/CorpusID:17799553>
30. Thomas, H., Qi, C., Deschaud, J.E., Marcotegui, B., Goulette, F., Guibas, L.J.: Kpconv: Flexible and deformable convolution for point clouds. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* pp. 6410–6419 (2019), <https://api.semanticscholar.org/CorpusID:121328056>
31. Ullah, I., Adhikari, D., Khan, H., Anwar, M.S., Ahmad, S., Bai, X.: Mobile robot localization: Current challenges and future prospective. *Comput. Sci. Rev.* **53**, 100651 (2024), <https://api.semanticscholar.org/CorpusID:271020621>
32. Wang, H., Liu, Y., Dong, Z., Wang, W., Yang, B.: You only hypothesize once: Point cloud registration with rotation-equivariant descriptors. *Proceedings of the 30th ACM International Conference on Multimedia* (2021), <https://api.semanticscholar.org/CorpusID:237371232>
33. Wang, H., Liu, Y., Hu, Q., Wang, B., Chen, J., Dong, Z., Guo, Y., Wang, W., Yang, B.: Roreg: Pairwise point cloud registration with oriented descriptors and local rotations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **45**, 10376–10393 (2023), <https://api.semanticscholar.org/CorpusID:256969892>
34. Xing, X., Lu, Z., Wang, Y., Xiao, J.: Efficient single correspondence voting for point cloud registration. *IEEE Transactions on Image Processing* **33**, 2116–2130 (2024), <https://api.semanticscholar.org/CorpusID:268373790>
35. Xu, Y., Shi, Z., Yifan, W., Chen, H., Yang, C., Peng, S., Shen, Y., Wetzstein, G.: Grm: Large gaussian reconstruction model for efficient 3d reconstruction and generation. *ArXiv abs/2403.14621* (2024), <https://api.semanticscholar.org/CorpusID:268554137>
36. Yao, R., Du, S., Cui, W., Tang, C., Yang, C.: Pare-net: Position-aware rotation-equivariant networks for robust point cloud registration. In: *European Conference on Computer Vision* (2024), <https://api.semanticscholar.org/CorpusID:271212761>
37. Yew, Z.J., Lee, G.H.: 3dfeat-net: Weakly supervised local 3d features for point cloud registration. *ArXiv abs/1807.09413* (2018), <https://api.semanticscholar.org/CorpusID:50770050>
38. Yu, H., Li, F., Saleh, M., Busam, B., Ilic, S.: Cofinet: Reliable coarse-to-fine correspondences for robust point cloud registration. In: *Neural Information Processing Systems* (2021), <https://api.semanticscholar.org/CorpusID:239998094>
39. Yu, H., Qin, Z., Hou, J., Saleh, M., Li, D., Busam, B., Ilic, S.: Rotation-invariant transformer for point cloud matching. *2023 IEEE/CVF Conference on Computer*

- Vision and Pattern Recognition (CVPR) pp. 5384–5393 (2023), <https://api.semanticscholar.org/CorpusID:257532279>
40. Yu, J., Ren, L., Zhou, W., Zhang, Y., Lin, L., Dai, G.: Peal: Prior-embedded explicit attention learning for low-overlap point cloud registration. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 17702–17711 (2023), <https://api.semanticscholar.org/CorpusID:260964522>
 41. Yue, X., Wu, B., Seshia, S.A., Keutzer, K., Sangiovanni-Vincentelli, A.L.: A lidar point cloud generator: from a virtual world to autonomous driving. Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval (2018), <https://api.semanticscholar.org/CorpusID:4557768>
 42. Zeng, A., Song, S., Nießner, M., Fisher, M., Xiao, J., Funkhouser, T.A.: 3dmatch: Learning local geometric descriptors from rgb-d reconstructions. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp. 199–208 (2016), <https://api.semanticscholar.org/CorpusID:11446141>