

Enhancing Rotation-Invariant 3D Learning with Global Pose Awareness and Attention Mechanisms

Jiaxun Guo¹, Manar Amayri¹, Nizar Bouguila¹, Xin Liu², Wentao Fan³*

¹CIISE, Concordia University, Montreal, Canada

²Department of Computer Science and Technology, Huaqiao University, Xiamen, China

³Guangdong Provincial/Zhuhai Key Laboratory IRADS and Department of Computer Science, Beijing Normal-Hong Kong Baptist University, Zhuhai, China

jiaxun.guo@mail.concordia.ca, {manar.amayri, nizar.bouguila}@concordia.ca, xliu@hqu.edu.cn, wentaofan@bnbu.edu.cn

Abstract

Recent advances in rotation-invariant (RI) learning for 3D point clouds typically replace raw coordinates with handcrafted RI features to ensure robustness under arbitrary rotations. However, these approaches often suffer from the loss of global pose information, making them incapable of distinguishing geometrically similar but spatially distinct structures. We identify that this limitation stems from the restricted receptive field in existing RI methods, leading to *Wing-tip feature collapse*, a failure to differentiate symmetric components (e.g., left and right airplane wings) due to indistinguishable local geometries. To overcome this challenge, we introduce the Shadow-informed Pose Feature (SiPF), which augments local RI descriptors with a globally consistent reference point (referred to as the “shadow”) derived from a learned shared rotation. This mechanism enables the model to preserve global pose awareness while maintaining rotation invariance. We further propose Rotation-invariant Attention Convolution (RIAttnConv), an attention-based operator that integrates SiPFs into the feature aggregation process, thereby enhancing the model’s capacity to distinguish structurally similar components. Additionally, we design a task-adaptive shadow locating module based on the Bingham distribution over unit quaternions, which dynamically learns the optimal global rotation for constructing consistent shadows. Extensive experiments on 3D classification and part segmentation benchmarks demonstrate that our approach substantially outperforms existing RI methods, particularly in tasks requiring fine-grained spatial discrimination under arbitrary rotations.

Code — <https://github.com/jiaxunguo/EnRI-GAM>

Extended version — <https://arxiv.org/abs/2511.08833>

Introduction

The rapid advancement of 3D sensing technologies has led to the widespread adoption of point cloud data in various 3D computer vision tasks, including shape classification, part segmentation, and object recognition (Guo et al. 2020; Afham et al. 2022; Duan et al. 2023; Jin et al. 2023). While recent learning-based models (Qi et al. 2017b; Wang et al. 2019; Liu et al. 2019) have achieved notable success on pre-aligned datasets, their generalization to real-world scenar-

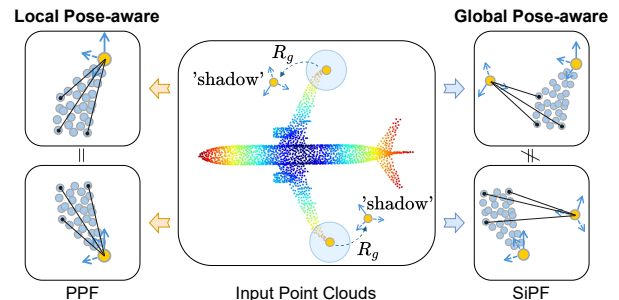


Figure 1: Illustration of pose ambiguity caused by symmetric local structures. Traditional Point Pair Features (PPF) capture only local geometric relationships and thus yield identical representations for symmetric regions with similar local geometry (e.g., left and right wing tips). In contrast, the proposed SiPF incorporates global pose awareness by introducing a ‘shadow’ reference point derived from a shared rotation matrix R_g to incorporate pose-aware context, enabling consistent and discriminative RI representations across arbitrary orientations.

ios remains limited. A common and often unrealistic assumption in these models is that input objects are consistently aligned in a canonical orientation. However, in practical applications such as autonomous driving, robotics, and augmented reality, 3D objects naturally appear under arbitrary rotations. Models that lack true *rotation-invariant* (RI) capabilities experience substantial performance degradation when confronted with such orientation variations, even when extensively trained with rotation-augmented data. This limitation fundamentally restricts the deployment of point cloud models in unconstrained environments.

To overcome this issue, a growing body of research has focused on designing RI models by replacing raw 3D coordinates with handcrafted or learned geometric features that are invariant to global rotations (Xu et al. 2021; Gu et al. 2022; Li et al. 2021b). These approaches demonstrate improved robustness to orientation changes by leveraging local descriptors such as point pair features or RI tensors. However, a critical drawback remains: by discarding absolute coordinate information, these methods also lose global pose context, which is essential for differentiating spatially dis-

*Corresponding author.

inct but geometrically similar regions. For instance, in the case of an airplane model, symmetric components such as the left and right wings may be indistinguishable when only local geometry is considered—regardless of rotation invariance. This leads to what we refer to as the *wing-tip feature collapse* phenomenon. Recent attempts to mitigate this issue involve expanding the receptive field and incorporating pairwise relationships within local neighbourhoods to capture broader context while maintaining RI properties (Chen and Cong 2022; Zhang, Yang, and Xiang 2024). Despite these advances, existing methods still struggle with structural ambiguity in repetitive or symmetric components, primarily due to their limited capacity to encode spatial distinctiveness without access to a consistent global reference.

We argue that the inherent limitation of the receptive field in current RI methods is a fundamental bottleneck. Specifically, for geometrically similar regions, such as opposing airplane wings, the local features (even aggregated from larger neighbourhoods) lack the discriminative information necessary to encode their distinct global poses. As shown in Figure 1, conventional RI descriptors fail to preserve relative global positioning, resulting in ambiguous representations that impede fine-grained recognition tasks like part segmentation. Furthermore, these limitations explain the suboptimal performance of many RI pipelines, which focus on preserving local geometric consistency but neglect global relational cues critical for downstream tasks.

To address these challenges, in this work we propose a novel feature representation termed *Shadow-informed Pose Feature* (SiPF) that augments local RI features with global pose cues while preserving rotation invariance. Inspired by the classical point pair feature (PPF) (Drost et al. 2010), SiPF introduces a globally consistent shadow reference for each point, derived via a learned shared rotation. This “shadow” serves as an anchor point for encoding relative position in a globally consistent manner. Specifically, we introduce a task-adaptive shadow locating module that estimates a shared optimal rotation via the Bingham distribution over unit quaternions. Using the learned rotation, each point is paired with its shadow to construct SiPF descriptors that retain local geometric relations while incorporating global spatial awareness. We further design a novel attention-based operator, termed *Rotation-invariant Attention Convolution* (RIAttnConv), that integrates SiPF into a cross-attention mechanism fused with convolutional operations. This design enables dynamic aggregation of spatially-aware features, thereby improving the model’s ability to distinguish structurally similar but spatially distinct components. We validate our method on multiple benchmark datasets across classification and part segmentation tasks. Experimental results demonstrate that our framework consistently outperforms state-of-the-art RI methods, particularly in challenging settings that require fine-grained spatial discrimination under arbitrary orientations.

Our main contributions are summarized as follows:

- We identify the restricted receptive field as a key limitation in current RI methods and address it by introducing the SiPF, which incorporates global pose context while maintaining rotation invariance.

- We develop RIAttnConv, a novel convolutional module that integrates SiPF through an attention-augmented feature aggregation strategy, significantly enhancing the discriminative capacity of RI representations.
- We design a task-adaptive shadow locating module that learns a globally consistent reference rotation using the Bingham distribution over unit quaternions, enabling dynamic and robust shadow construction.

Related Works

Rotation-Sensitive 3D Point Cloud Analysis

PointNet (Qi et al. 2017a) is a pioneering work in point cloud analysis, introducing a straightforward yet effective framework for directly processing raw 3D point sets in tasks such as classification and segmentation. To address its limitations in capturing local geometric structures, PointNet++ (Qi et al. 2017b) extends this architecture with hierarchical feature learning over local neighbourhoods. Subsequent works, such as PointCNN (Li et al. 2018) and DGCNN (Wang et al. 2019), further enhance local context modeling through specialized convolutional and dynamic graph-based operations, respectively. Despite these advances, a shared limitation among these approaches is their sensitivity to input orientation. Because they operate directly on absolute coordinates, their performance tends to degrade significantly when encountering arbitrarily rotated inputs, which limits their applicability in real-world scenarios.

Rotation-Robust Methods

To mitigate the sensitivity to rotations, early approaches utilize principal component analysis (PCA) to align input point clouds to a canonical orientation. For instance, Li et al. (Li et al. 2021a) address pose ambiguity through hybrid canonicalization strategies. However, such methods often rely heavily on data augmentation or heuristic approximations, failing to achieve true rotation invariance. An alternative direction involves equivariant neural networks that encode predictable transformations of features under rotations via structured linear mappings (Deng et al. 2021; Poulenard and Guibas 2021; Fuchs et al. 2020). These models preserve pose information while improving rotation robustness. Nonetheless, maintaining equivariance typically imposes strict architectural constraints, such as enforcing linear compositions in encoders (Deng et al. 2021; Jing et al. 2020), which can limit model expressiveness and hinder learning capacity. Recently, constructing RI features within local neighbourhoods has emerged as a prevalent strategy. Most of these methods employ handcrafted features derived from geometric priors such as pairwise distances and inter-point angles (Li et al. 2021b; Xu et al. 2021). For instance, RI-CNN (Zhang et al. 2019) and PaRI-Conv (Chen and Cong 2022) design 4D and 8D local descriptors, respectively, to enhance geometric encoding. However, these designs often sacrifice global contextual information, thereby constraining their ability to capture complex spatial semantics (Zhang et al. 2019). To alleviate this drawback, several methods have attempted to expand the receptive field or increase connectivity. ClusterNet (Chen et al. 2019), for

example, constructs a densely connected k NN graph (with $k = 80$) to retain richer relational information. LocoTrans (Chen et al. 2024) enhances local RI features by leveraging representations derived from an equivariant backbone.

Recent advances such as RIGA (Yu et al. 2024), and rotation-invariant transformers (Yu et al. 2023) show that incorporating global context or attention is beneficial for rotation-robust point matching. Attention-based sampling strategies (Wu et al. 2023, 2025) further highlight the importance of globally informed feature construction. Building on these insights, our approach explicitly integrates global pose information with local geometric encoding, yielding more expressive RI representations and consistently strong performance across rotation-robust benchmarks.

Problem Definition and Background

Rotation-Invariant Functions. Let a point cloud be denoted by $P = \{p_1, \dots, p_N\} \in \mathbb{R}^{N \times 3}$ and let $R \in SO(3)$ represents an arbitrary 3D rotation matrix. The rotated point cloud is thus $P_R = P R$. A function Φ is RI if and only if:

$$\Phi(P_R) = \Phi(P), \quad \forall R \in SO(3). \quad (1)$$

Receptive Field in 3D Point Cloud Convolutions. Given a point cloud $P = [p_1, \dots, p_N] \in \mathbb{R}^{N \times 3}$ with associated input features $X = [x_1, \dots, x_N] \in \mathbb{R}^{N \times c_{in}}$, consider a reference point p_r with its local neighbourhood denoted by $\mathcal{N}(p_r)$. The output feature at p_r , written as part of $X' = [x'_1, \dots, x'_N] \in \mathbb{R}^{N \times c_{out}}$, is typically obtained by aggregating transformed features of its neighbours:

$$f(p_r) = \bigwedge_{j \in \mathcal{N}(p_r)} W_j^T \cdot h(p_j). \quad (2)$$

where $h : \mathbb{R}^3 \rightarrow \mathbb{R}^c$ is a nonlinear point-encoding function, W_j^T denotes the learnable dynamic kernel weight associated with the pair (p_r, p_j) , and \bigwedge represents an aggregation operator such as MAX, SUM, or AVG.

Let the local patch around point p_j be defined as $\Omega(p_j) = \{p_k \mid k \in \mathcal{N}(p_j)\}$. Since h may also depend on the neighbourhood of p_j , the effective receptive field at p_r becomes the union of its neighbours' patches:

$$\Gamma(p_r) = \bigcup_{j \in \mathcal{N}(p_r)} \Omega(p_j). \quad (3)$$

To make the dependence on the receptive field explicit, Equation 2 can be reformulated as:

$$f(p_r | \Gamma(p_r)) = \bigwedge_{j \in \mathcal{N}(p_r)} W_j^T \cdot h(p_j | \Omega(p_j)). \quad (4)$$

Limitations of the Receptive Field in Existing RI Methods. Consider an airplane point cloud and let $p_{\text{left}}, p_{\text{right}} \in P$ be the two wing-tip points. Due to their geometric symmetry, there exists a rotation $R_{\text{sym}} \in SO(3)$ such that:

$$\Omega(p_{\text{right}}) = \Omega(p_{\text{left}}) R_{\text{sym}}. \quad (5)$$

Patch-Swapping Transformation. For any reference point p_r with neighbourhood $\mathcal{N}(p_r)$, we define a transformation on the receptive field as:

$$T(\Gamma(p_r)) = \bigcup_{j \in \mathcal{N}(p_r)} \Omega(p_j) R_j, \quad (6)$$

where each $R_j \in SO(3)$ maps the local patch $\Omega(p_j)$ to its geometrically equivalent counterpart (e.g. R_{sym} swaps the left and right wing-tip patches). Applying Equation 4 to the transformed data yields:

$$f(p_r | T(\Gamma(p_r))) = \bigwedge_{j \in \mathcal{N}(p_r)} W_j^T \cdot h(p_j | \Omega(p_j) R_j). \quad (7)$$

Since h is rotation-invariant by definition (Equation 1), we have $h(p_j | \Omega(p_j) R_j) = h(p_j | \Omega(p_j))$. Therefore, Equations 4 and 7 are equivalent, leading to:

$$f(p_r | \Gamma(p_r)) = f(p_r | T(\Gamma(p_r))). \quad (8)$$

Wing-Tip Feature Collapse. By setting $p_r = p_{\text{left}}$ and choosing T to swap the left and right wing-tip patches, Equation (8) simplifies to: $f(p_{\text{left}}) = f(p_{\text{right}})$. This result implies that the network assigns identical features to symmetric yet globally opposite regions. As a consequence, crucial pose information (e.g. distinction between left and right) is lost, introducing ambiguity in downstream tasks and limiting the discriminative power of the learned representations.

The Proposed Method

To address the global pose ambiguity that arises among geometrically equivalent local patches, we propose a novel RI point descriptor, termed SiPF. It augments each point with a global pose-aware reference, known as the *shadow*, obtained by applying a shared rotation to the entire point cloud. This augmentation allows local patches to encode relative pose information without compromising rotation invariance, thereby enabling the network to distinguish globally distinct yet locally similar structures.

Building upon SiPF, we further introduce RIAttnConv, a rotation-invariant convolutional operator that incorporates shadow-informed relative pose information into the feature aggregation process. Specifically, for a given reference point p_r , the output feature is computed as:

$$x'_r = \bigwedge_{j \in \mathcal{N}(p_r)} \text{Attn}(\{\mathcal{W}(\mathcal{P}_r^j) \cdot x_j\}_{j=1}^k), \quad (9)$$

where \mathcal{P}_r^j encodes the relative pose among the reference point p_r , its shadow p'_r , and neighbour p_j . The kernel weight W_j^T adapts dynamically to the global context introduced by p'_r , allowing the model to distinguish symmetric structures based on their global orientation. As a result, W_j^T in Equations (4) and (7) varies under local patch rotations, enabling the attention module $\text{Attn}(\cdot)$ to differentiate identical local geometries in distinct global poses while preserving overall rotation invariance. The full architectural pipeline of our method is illustrated in Figure 2.

Shadow-Informed Pose Feature

LRF Construction. Local Reference Frame (LRF) is a fundamental component for constructing RI representations. While the encoding on the relative pose \mathcal{P}_r^j between point pairs, the LRF \mathcal{L}_r serves as a local coordinate system independently constructed at each point $p_r \in P$. Formally, $\mathcal{L}_r = [\partial_r^1, \partial_r^2, \partial_r^3] \in \mathbb{R}^{3 \times 3}$ consisting of three orthonormal

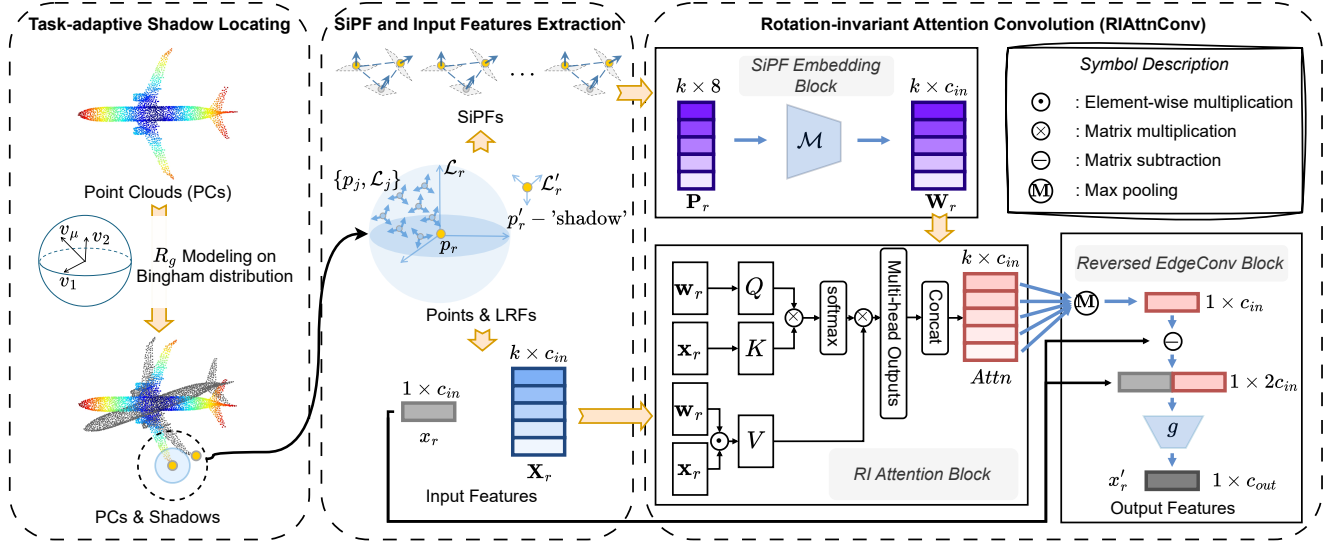


Figure 2: Overview of the proposed SiPF and RIAttnConv pipeline. Our framework consists of three main components: (1) Task-adaptive Shadow Locating estimates a globally consistent reference rotation R_g by modeling unit quaternions with a Bingham distribution, generating ‘shadow’ points for each input; (2) SiPF and Input Feature Extraction constructs LRFs $\{\mathcal{L}_j\}_{j=1}^k$ and computes SiPFs \mathbf{P}_r by combining local geometric descriptors with global shadow-based pose differences; (3) RIAttnConv embeds SiPFs to produce adaptive kernel weights \mathbf{W}_r , which are used in a RI attention block to guide feature aggregation. A reversed EdgeConv module then fuses neighbourhood and center features x_r to produce the final output x'_r .

basis vectors, typically derived from two direction vectors e_r^1, e_r^2 via Gram–Schmidt orthonormalization:

$$\partial_r^1 = e_r^1, \quad \partial_r^3 = \frac{\partial_r^1 \times e_r^2}{\|\partial_r^1 \times e_r^2\|}, \quad \partial_r^2 = \partial_r^3 \times \partial_r^1. \quad (10)$$

Recent works (Chen and Cong 2022; Zhang, Yang, and Xiang 2024) adopted a robust and purely local LRF construction strategy, wherein the vector from the barycenter of point neighbours to the point itself defines one axis. Specifically, the barycenter is computed as $m_r = \frac{1}{k} \sum_{j=1}^k p_j^r$, where p_j^r are the k -nearest neighbours of p_r , and the axis e_r^2 is defined as $\overrightarrow{p_r m_r} = m_r - p_r$. In this work, we adopt this LRF construction strategy based on local geometry ($e_r^1 = n_r, e_r^2 = \overrightarrow{p_r m_r}$) as input features X to ensure essential robustness and consistency in real-world point clouds.

SiPF Extraction. To describe the relative pose between the LRFs of a reference point p_r and its neighbour p_j , we design our RI feature based on the standard Point Pair Feature (PPF) (Drost et al. 2010), a 4D vector constructed from the relative position and the primary LRF axes $\partial_r^1, \partial_j^1$ of the two points:

$$\begin{aligned} \text{PPF}(p_r, p_j) &= (\|d\|_2, \cos(\alpha_1), \cos(\alpha_2), \cos(\alpha_3)), \\ \alpha_1 &= \angle(\partial_r^1, d), \quad \alpha_2 = \angle(\partial_j^1, d), \quad \alpha_3 = \angle(\partial_r^1, \partial_j^1), \end{aligned} \quad (11)$$

where $d = p_j - p_r$ and $\angle(\cdot, \cdot)$ denotes the angle between two vectors. While PPF is sufficient to represent the pairwise pose, it becomes ambiguous in the context of multiple neighbours. As shown in prior works (Deng, Birdal, and Ilic 2018; Chen and Cong 2022), for two neighbours p_j and p'_j with identical normals and equal distances to p_r , if both lie

on a circle around the axis ∂_r^1 , their PPF values with respect to p_r will be identical, i.e., $\text{PPF}(p_r, p_j) = \text{PPF}(p_r, p'_j)$, despite having distinct spatial arrangements.

To resolve this ambiguity and incorporate global pose awareness, we introduce the *shadow* point p'_r of p_r , derived from a shared global rotation. We then define the *Shadow-informed Point Pair Feature* (SiPPF) as:

$$\text{SiPPF}(p_r, p'_r, p_j) = \frac{\text{PPF}(p_r, p'_r) - \text{PPF}(p_j, p'_r)}{\|\text{PPF}(p_r, p'_r) - \text{PPF}(p_j, p'_r)\|_2}, \quad (12)$$

where $\text{PPF}(p_r, p'_r)$ and $\text{PPF}(p_j, p'_r)$ encode the relative geometry between each point and the shadow. We normalize the difference using the ℓ_2 -norm to emphasize the pose difference of p_j under the shared global reference.

Finally, we concatenate the local geometric encoding with the global-aware difference to obtain our 8D *Shadow-informed Pose Feature* (SiPF):

$$\mathcal{P}_r^j = (\text{PPF}(p_r, p_j), \text{SiPPF}(p_r, p'_r, p_j)) \in \mathbb{R}^8. \quad (13)$$

Task-Adaptive Shadow Locating. As previously discussed, a shared global rotation is applied to generate the shadow point p'_r , which serves as a consistent reference for global-relative feature extraction. However, in certain geometric configurations, the shadow may fail to enhance the discriminability of local features due to inherent ambiguities. Two representative failure cases are outlined below:

- When the shadow p'_r lies along the primary LRF axis ∂_r^1 , and the primary axes ∂_r^1 and ∂_j^1 coincide, the SiPF degenerates to standard PPF due to unresolved degrees of freedom of p_j .

- If the global rotation coincides with the local rotation path R_j used in Equation 6, the resulting SiPF only cancels the internal degrees of freedom of p_j , but fails to distinguish the receptive field $\Omega(p_j)$ from its rotated variant $\Omega(p_j)R_j$.

These cases highlight the necessity of a task-adaptive, ambiguity-resilient global rotation strategy. To account for the ambiguity in selecting \mathbf{R}_g over the entire dataset, we model the uncertainty using a *Bingham distribution* on the hypersphere S^3 of unit quaternions (Bingham 1974). Let q represent candidate quaternions corresponding to potential global rotations. The Bingham probability density function is defined as:

$$B(q | \mathbf{V}, \mathbf{\Lambda}) = \frac{1}{F(\mathbf{\Lambda})} \exp(q^\top \mathbf{V} \mathbf{\Lambda} \mathbf{V}^\top q), \quad (14)$$

where $\mathbf{V} \in \mathbb{R}^{4 \times 4}$ is an orthogonal matrix, and $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \lambda_3, 0)$ is a diagonal matrix with $\lambda_1 \leq \lambda_2 \leq \lambda_3 < 0$. The matrix \mathbf{V} defines the mode and dispersion axes, while $\mathbf{\Lambda}$ controls the strength of dispersion along these axes. $F(\mathbf{\Lambda})$ is the normalization constant. To connect \mathbf{R}_g with the learned mode of the Bingham distribution, we extract the mode vector in \mathbf{V} as the optimal global rotation candidate for generating $p'_r = p_r \mathbf{R}_g$ at each iteration.

To ensure that the selected global rotation aligns with the task-specific objectives, we embed the optimization of \mathbf{R}_g into the end-to-end training. Specifically, we jointly optimize the rotation with network parameters by minimizing a composite loss that integrates both the task objective and the probabilistic alignment with the Bingham distribution:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{task}} + \delta \cdot \sqrt{(\mathcal{L}_{\text{bingham}} - 0.1 \cdot \mathcal{L}_{\text{task}})^2}, \quad (15)$$

where $\mathcal{L}_{\text{bingham}}$ denotes the negative log-likelihood of the Bingham distribution evaluated at the quaternion $q = \text{Quat}(\mathbf{R}_g)$, and δ is a balancing hyperparameter, empirically set to 0.8. The subtraction term encourages to maintain consistency between the uncertainty captured by the Bingham model and the alignment needs of the downstream task.

The RIAttnConv Operator

To enhance RI feature learning, we propose the RIAttnConv operator. Unlike prior approaches that rely only on relative pose for determining kernel weights, RIAttnConv employs an attention mechanism that adaptively weights neighbour features based on both geometric cues and feature similarity. This design enables the model to capture broader contextual dependencies while facilitating effective aggregation.

RI Attention. Given a reference feature $x_r \in \mathbb{R}^{c_{in}}$ and its k -nearest neighbouring features $\{x_j\}_{j=1}^k$, we define the stacked neighbour feature matrix as $\mathbf{X}_r = [x_1, \dots, x_k]^\top \in \mathbb{R}^{k \times c_{in}}$. Similarly, the corresponding SiPF descriptors are stacked as $\mathbf{P}_r = [\mathcal{P}_r^1, \dots, \mathcal{P}_r^k]^\top \in \mathbb{R}^{k \times 8}$. RIAttnConv constructs a scaled dot-product attention mechanism using both \mathbf{X}_r and \mathbf{P}_r :

$$\text{Attention}(Q, K, V) = \text{Softmax} \left(\frac{QK^\top}{\sqrt{c_{in}}} \right) V, \quad (16)$$

$$Q = \mathbf{W}_r, \quad K = \mathbf{X}_r, \quad V = \mathbf{W}_r \cdot \mathbf{X}_r,$$

where $\mathbf{W}_r = [W_1^\top, \dots, W_k^\top]^\top \in \mathbb{R}^{k \times c_{in}}$, and $W_j^\top = \mathcal{M}(\mathcal{P}_r^j)$ as $\mathcal{W}(\mathcal{P}_r^j)$ in Equation (9). \mathcal{M} is a Multi-layer Perceptron (MLP) that maps the relative pose \mathcal{P}_r^j into the respective dimension.

This design enables the Softmax(\cdot) in Equation (16) to produce a $k \times k$ attention matrix over the k -nearest neighbours, enhancing contextual sensitivity. Compared to prior methods (Chen and Cong 2022; Zhou et al. 2021), this formulation offers a favourable trade-off between computational complexity and expressive power. A detailed efficiency analysis is provided in ablation study (see Table 5).

Reversed EdgeConv. To retain the feature information of the reference point x_r , we propose a reversed variant of EdgeConv (Wang et al. 2019), where neighbour features are first aggregated and then fused with the reference feature. Specifically, given the attention mechanism, we compute:

$$\text{Attn}(\{\mathcal{W}(\mathcal{P}_r^j) \cdot x_j\}_{j=1}^k) = \text{Attention}(Q, K, V), \quad (17)$$

$$\hat{x}_r = \text{MAX}_{j \in \mathcal{N}(p_r)} \text{Attn}(\{\mathcal{W}(\mathcal{P}_r^j) \cdot x_j\}_{j=1}^k),$$

where Attn is the output from the RI attention module defined in Equation (16). Then, the updated feature x'_r is obtained by fusing \hat{x}_r and the original reference feature x_r as:

$$x'_r = g((\hat{x}_r - x_r) \oplus x_r), \quad (18)$$

where $g(\cdot)$ denotes a one-layer MLP and \oplus represents feature concatenation.

Network Architecture

To validate the generalizability and effectiveness of the proposed RIAttnConv operator, we integrate it into two representative backbone architectures widely adopted in point cloud learning: DGCNN (Wang et al. 2019) for classification, and AdaptConv (Zhou et al. 2021) for segmentation.

Shape Classification. We adopt the DGCNN backbone and replace all EdgeConv layers with RIAttnConv. The k -nearest neighbour graph is constructed in Euclidean space at each feature layer, with $k = 20$ as the neighbourhood size.

Shape Part Segmentation. For part segmentation, we follow the AdaptConv-based design. Specifically, we stack four RIAttnConv layers followed by a graph convolution layer. Three intermediate pooling stages are inserted to progressively enlarge the receptive field. The k -nearest neighbour graph is constructed with $k = 40$ for all layers, using Euclidean distance in feature space.

Experiments

We evaluate our method on three benchmarks: ModelNet40 (Wu et al. 2015) for shape classification, ShapeNetPart (Yi et al. 2016) for part segmentation, and ScanObjectNN (Uy et al. 2019) for real-world classification. To assess rotation invariance, we adopt three standard train/test protocols: z/z , $z/\text{SO}(3)$, and $\text{SO}(3)/\text{SO}(3)$, where z and $\text{SO}(3)$ represent vertical and arbitrary rotations. For fair comparison, RIAttnConv is integrated into standard backbones without additional modifications. Extensive ablation studies further validate the contribution of each component, with more results provided in the supplementary material.

Methods	Inputs	z/z	z/SO(3)	SO(3)/SO(3)
Rotation-sensitive				
PointNet (Qi et al. 2017a)	pc	89.2	16.2	75.5
PointNet++ (Qi et al. 2017b)	pc+n	89.3	28.6	85.0
PointCNN (Li et al. 2018)	pc	92.2	41.2	84.5
DGCNN (Wang et al. 2019)	pc	92.2	20.6	81.1
Rotation-robust				
RIConv (Zhang et al. 2019)	pc	86.5	86.4	86.4
ClusterNet (Chen et al. 2019)	pc	87.1	87.1	87.1
RI-GCN (Kim, Park, and Han 2020)	pc+n	91.0	91.0	91.0
VN-DGCNN (Deng et al. 2021)	pc	89.5	89.5	90.2
SGMNet (Xu et al. 2021)	pc	90.0	90.0	90.0
OrientedMP (Luo et al. 2022)	pc+n	88.4	88.4	88.4
ELGANet (Gu et al. 2022)	pc	90.3	90.3	90.3
PaRot (Zhang et al. 2023)	pc	90.9	91.0	90.8
TetraSphere (Melnyk et al. 2024)	pc	90.5	90.5	90.3
Ours	pc	91.8	91.8	91.8
Ours	pc+n	92.6	92.6	92.6

Table 1: Shape classification accuracy (%) on ModelNet40 dataset. ‘pc’ and ‘n’ represent the 3D coordinates and normal vectors of input data, respectively.

Methods	z/SO(3)	SO(3)/SO(3)
Input pc only		
PointNet (Qi et al. 2017a)	16.7	54.7
PointNet++ (Qi et al. 2017b)	15.0	47.4
PointCNN (Li et al. 2018)	14.6	63.7
DGCNN (Wang et al. 2019)	17.7	71.8
RIConv (Zhang et al. 2019)	75.3	75.5
RI-GCN (Kim, Park, and Han 2020)	80.5	80.6
VN-DGCNN (Deng et al. 2021)	79.8	80.3
OrientedMP (Luo et al. 2022)	76.7	77.2
LGR-Net (Zhao et al. 2022)	81.2	81.4
PaRI-Conv (Chen and Cong 2022)	83.3	83.3
PaRot (Zhang et al. 2023)	82.1	82.6
Ours	84.0	84.0

Table 2: Real-world classification accuracy (%) on ScanObjectNN OBJ_BG dataset.

Implementation Details

We use stochastic gradient descent (SGD) with an initial learning rate of 0.1, decayed to 0.001 via cosine annealing (Loshchilov and Hutter 2016). Models are trained for 300 epochs with a batch size of 32 for classification and 16 for segmentation. A dropout rate of 0.5 is applied to fully connected layers.

To ensure rotation-invariant input representations, we initialize a point-wise descriptor for each point p_i based on its relative position to the global centroid O (Cohen et al. 2018), including $\|\overrightarrow{Op_i}\|_2$, $\sin(\angle(\partial_i^1, \overrightarrow{Op_i}))$, and $\cos(\angle(\partial_i^1, \overrightarrow{Op_i}))$, where ∂_i^1 is the local reference axis. By default, we use surface normals as ∂_i^1 (denoted as pc+normal). For methods using only coordinates, we replace ∂_i^1 with the vector $\overrightarrow{p_r m_r}$, from the barycenter to the point.

3D Shape Classification

Dataset. We evaluate our method on the widely used ModelNet40 (Wu et al. 2015) benchmark, which contains 12,311 CAD models from 40 categories, with 9,843 samples for training and 2,468 for testing. Following (Qi et al. 2017a), we randomly sample 1,024 points for each object and augment the training data with random scaling and translation.

Results. Table 1 reports classification accuracy under three rotation protocols: z/z, z/SO(3), and SO(3)/SO(3). While

Method	Input	z/SO(3)		SO(3)/SO(3)	
		C. mIoU	I. mIoU	C. mIoU	I. mIoU
PointNet (Qi et al. 2017a)	pc	37.8	-	74.4	-
PointNet++ (Qi et al. 2017b)	pc	48.2	-	76.7	-
DGCNN (Wang et al. 2019)	pc	37.4	-	73.3	-
RI-CNN (Zhang et al. 2019)	pc	75.3	-	75.5	-
RI-GCN (Kim, Park, and Han 2020)	pc+n	77.2	-	77.3	-
VN-DGCNN (Deng et al. 2021)	pc	-	81.4	-	81.4
LGR-Net (Zhao et al. 2022)	pc+n	-	-	-	82.8
PaRI-Conv (Chen and Cong 2022)	pc+n	-	84.6	-	84.6
CRIN (Lou et al. 2023)	pc	80.5	-	80.5	-
PaRot (Zhang et al. 2023)	pc	79.2	-	79.5	-
RI-PCR (Yu, Zhang, and Cai 2023)	pc	-	80.3	-	80.4
LocoTrans (Chen et al. 2024)	pc	80.1	84.0	80.0	83.8
TetraSphere (Melnyk et al. 2024)	pc	-	82.3	-	82.3
RISurConv (Zhang et al. 2024)	pc+n	81.5	-	81.5	-
Ours	pc	81.7	84.4	82.4	85.1
Ours	pc+n	82.9	85.0	83.8	85.5

Table 3: Part segmentation results on ShapeNetPart dataset. ‘C. mIoU’ stands for mIoU (%) over all classes while ‘I. mIoU’ denotes mIoU over all instances.

rotation-sensitive methods (e.g., PointCNN, DGCNN) perform well under aligned conditions, their accuracy drops sharply (up to 70%) under unseen rotations. In contrast, RI methods exhibit stable performance across settings but often suffer from limited expressiveness without normals; for instance, RIConv and ClusterNet achieve only 86–87% accuracy in such cases.

Our proposed RIAttnConv achieves 91.8% accuracy using only point coordinates, surpassing several normal-dependent methods such as OrientedMP (88.4%) and RI-GCN (91.0%). With surface normals, performance further improves to 92.6%, yielding a 1.6% gain over RI-GCN and outperforming recent approaches including TetraSphere and PaRot. These results highlight the effectiveness of our attention-based design in learning pose-aware yet rotation-invariant features.

Real-World Shape Classification

Dataset. We further evaluate our method on the ScanObjectNN dataset (Uy et al. 2019), a real-world benchmark comprising 2,902 objects from 15 categories captured in cluttered and noisy environments. Following standard protocol, we use the OBJ_BG subset, consisting of 2,890 samples with 2,312 for training and 578 for testing. As surface normals are unavailable, only raw point coordinates are used.

Results. As shown in Table 2, our method achieves the highest accuracy under both z/SO(3) and SO(3)/SO(3) settings, despite using only point coordinates. It outperforms PaRI-Conv by 0.7% in both cases, highlighting the effectiveness of our global pose-aware attention in capturing robust local geometry under real-world noise and occlusion.

Shape Part Segmentation

Dataset. We evaluate our method on the widely adopted ShapeNetPart dataset (Yi et al. 2016), which contains 16,881 3D models across 16 object categories. Each object is annotated with 2 to 6 parts, yielding a total of 50 distinct part labels. Following the standard protocol in (Qi et al. 2017b), we

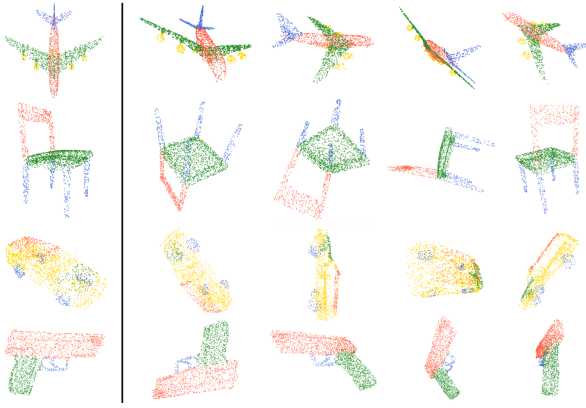


Figure 3: Visualization of part segmentation results on ShapeNetPart Dataset under $z/SO(3)$ setting. The leftmost column is the ground truth. The testing results under arbitrary rotations are on the rest columns.

use the official train/test split and uniformly sample 2,048 points from each object for input.

Results. As shown in Table 3, our method achieves state-of-the-art performance under both $z/SO(3)$ and $SO(3)/SO(3)$ settings. Under the more challenging $z/SO(3)$ setup, it achieves a class mean Intersection-over-Union (C. mIoU) of 82.9% and an instance mean IoU (I. mIoU) of 85.0% with normals, outperforming recent methods. Without normals, it still achieves 84.4% and 85.1% instance mIoU, surpassing existing RI baselines such as LocoTrans and TetraSphere, and rivaling normal-dependent approaches like PaRI-Conv and RISurConv. Furthermore, our method consistently achieves strong class-level and instance-level segmentation performance under unseen rotations. This improvement stems from the proposed SiPF and global pose-aware reference, which enable the model to disambiguate locally similar parts by leveraging global spatial context. Figure 3 visualizes representative segmentation outputs under the $z/SO(3)$ setting, further illustrating the model’s robustness to arbitrary spatial rotations.

Ablation Studies

In this section, we present a series of ablation experiments to assess the effectiveness of the key components in our proposed framework. Unless otherwise stated, all experiments are performed on ShapeNetPart dataset with normal vector under the $z/SO(3)$ setting and reported with C. mIoU (%) and I. mIoU (%).

Effectiveness of SiPF. To assess the effectiveness of SiPF

No.	RI Representation	Dim	$z/SO(3)$	
			C. mIoU	I. mIoU
1	PPF	4	81.1	84.1
2	Aug. PPF	8	81.8	84.2
3	SiPF-w/oDirection	5	82.4	84.5
4	SiPF	8	82.9	85.0

Table 4: Ablation study on the relative pose feature \mathcal{P}_r^j .

RI Convolution	Params	FLOPs	$z/SO(3)$	
			C. mIoU	I. mIoU
Concat.	3.00M	4895M	82.1	84.2
AdaptConv	3.91M	18896M	82.0	84.2
PaRI-Conv	2.95M	4574M	82.4	84.5
RIAttnConv	3.01M	4795M	82.9	85.0

Table 5: Ablation study on the RI convolution.

	LocoTrans	TetraSphere	RISurConv	Ours
Params	6.72M*	1.31M	4.06M	3.01M
FLOPs	7998M*	4092M	9726M	4795M
C. mIoU	80.1	-	81.3	81.7
I. mIoU	84.0	82.3	-	84.4

Table 6: Comparison of FLOPs, parameter count, and mIoU among different models with point coordinates only. (* indicates the results are estimated based on the classification architecture.)

in capturing both rotation invariance and global pose cues, we compare it with the standard Point Pair Feature (PPF) and the augmented PPF (Aug. PPF) from (Chen and Cong 2022), which introduces an extra local axis to resolve PPF ambiguity. As shown in Table 4, SiPF clearly outperforms both variants, underscoring the importance of incorporating a global pose-aware reference. We also report ablations on SiPF’s directional components, confirming their contribution to enhanced robustness and discriminability.

Effectiveness of RIAttnConv. We evaluate the effectiveness of RIAttnConv by comparing it with alternative convolutional strategies, as shown in Table 5. All variants employ SiPF as the RI descriptor for fair comparison. In addition to two recent convolutional baselines (Zhou et al. 2021; Chen and Cong 2022), we include a variant that directly concatenates SiPF with latent features without structural modification. RIAttnConv achieves the best performance while maintaining comparable or lower complexity on parameters and FLOPs, highlighting its efficiency in utilizing SiPF.

Computational Efficiency. Table 6 reports the computational cost of our method compared to recent state-of-the-art approaches. While some methods improve accuracy by increasing parameters and FLOPs, our model achieves the best segmentation performance with substantially lower complexity, demonstrating superior efficiency and scalability.

Conclusion

In this paper, we proposed the Shadow-informed Pose Feature (SiPF), a novel rotation-invariant descriptor that enriches local geometric features with global pose context via a learned shadow reference. Together with the introduced RI-AttnConv operator, our framework effectively distinguishes symmetric or spatially ambiguous structures and achieves strong performance across classification and segmentation benchmarks under arbitrary rotations. A current limitation is that our evaluation is conducted on object-level datasets. Scene-level assessment remains an appealing future direction, which may further highlight pose ambiguity and validate the broader applicability of our framework.

Acknowledgments

The completion of this work was supported in part by Natural Sciences and Engineering Research Council of Canada (NSERC), the National Natural Science Foundation of China (62276106, 62576143), Guangdong Basic and Applied Basic Research Foundation (2024A1515011767), Guangdong Provincial Key Laboratory IRADS (2022B1212010006), and Guangdong Higher Education Upgrading Plan (2021-2025) of “Rushing to the Top, Making Up Shortcomings and Strengthening Special Features” (2024KTSCX222, 2024GXJK695).

References

- Afham, M.; Dissanayake, I.; Dissanayake, D.; Dharmasiri, A.; Thilakarathna, K.; and Rodrigo, R. 2022. Crosspoint: Self-supervised cross-modal contrastive learning for 3d point cloud understanding. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 9902–9912.
- Bingham, C. 1974. An antipodally symmetric distribution on the sphere. *The Annals of Statistics*, 1201–1225.
- Chen, C.; Li, G.; Xu, R.; Chen, T.; Wang, M.; and Lin, L. 2019. Clusternet: Deep hierarchical cluster network with rigorously rotation-invariant representation for point cloud analysis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 4994–5002.
- Chen, R.; and Cong, Y. 2022. The devil is in the pose: Ambiguity-free 3d rotation-invariant learning via pose-aware convolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7472–7481.
- Chen, Y.; Duan, L.; Zhao, S.; Ding, C.; and Tao, D. 2024. Local-consistent transformation learning for rotation-invariant point cloud analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5418–5427.
- Cohen, T. S.; Geiger, M.; Köhler, J.; and Welling, M. 2018. Spherical cnns. *arXiv preprint arXiv:1801.10130*.
- Deng, C.; Litany, O.; Duan, Y.; Poulenard, A.; Tagliasacchi, A.; and Guibas, L. J. 2021. Vector neurons: A general framework for so(3)-equivariant networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 12200–12209.
- Deng, H.; Birdal, T.; and Ilic, S. 2018. Ppf-foldnet: Unsupervised learning of rotation invariant 3d local descriptors. In *Proceedings of the European conference on computer vision (ECCV)*, 602–618.
- Drost, B.; Ulrich, M.; Navab, N.; and Ilic, S. 2010. Model globally, match locally: Efficient and robust 3D object recognition. In *2010 IEEE computer society conference on computer vision and pattern recognition*, 998–1005. Ieee.
- Duan, L.; Zhao, S.; Xue, N.; Gong, M.; Xia, G.-S.; and Tao, D. 2023. Condaformer: Disassembled transformer with local structure enhancement for 3d point cloud understanding. *Advances in Neural Information Processing Systems*, 36: 23886–23901.
- Fuchs, F.; Worrall, D.; Fischer, V.; and Welling, M. 2020. Se(3)-transformers: 3d roto-translation equivariant attention networks. *Advances in neural information processing systems*, 33: 1970–1981.
- Gu, R.; Wu, Q.; Li, Y.; Kang, W.; Ng, W. W.; and Wang, Z. 2022. Enhanced local and global learning for rotation-invariant point cloud representation. *IEEE MultiMedia*, 29(4): 24–37.
- Guo, Y.; Wang, H.; Hu, Q.; Liu, H.; Liu, L.; and Bennamoun, M. 2020. Deep learning for 3d point clouds: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 43(12): 4338–4364.
- Jin, Z.; Hayat, M.; Yang, Y.; Guo, Y.; and Lei, Y. 2023. Context-aware alignment and mutual masking for 3d-language pre-training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10984–10994.
- Jing, B.; Eismann, S.; Suriana, P.; Townshend, R. J.; and Dror, R. 2020. Learning from protein structure with geometric vector perceptrons. *arXiv preprint arXiv:2009.01411*.
- Kim, S.; Park, J.; and Han, B. 2020. Rotation-invariant local-to-global representation learning for 3d point cloud. *Advances in Neural Information Processing Systems*, 33: 8174–8185.
- Li, F.; Fujiwara, K.; Okura, F.; and Matsushita, Y. 2021a. A closer look at rotation-invariant deep point cloud analysis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 16218–16227.
- Li, X.; Li, R.; Chen, G.; Fu, C.-W.; Cohen-Or, D.; and Heng, P.-A. 2021b. A rotation-invariant framework for deep point cloud analysis. *IEEE transactions on visualization and computer graphics*, 28(12): 4503–4514.
- Li, Y.; Bu, R.; Sun, M.; Wu, W.; Di, X.; and Chen, B. 2018. Pointcnn: Convolution on x-transformed points. *Advances in neural information processing systems*, 31.
- Liu, Y.; Fan, B.; Xiang, S.; and Pan, C. 2019. Relation-shape convolutional neural network for point cloud analysis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 8895–8904.
- Loshchilov, I.; and Hutter, F. 2016. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*.
- Lou, Y.; Ye, Z.; You, Y.; Jiang, N.; Lu, J.; Wang, W.; Ma, L.; and Lu, C. 2023. CRIN: rotation-invariant point cloud analysis and rotation estimation via centrifugal reference frame. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 1817–1825.
- Luo, S.; Li, J.; Guan, J.; Su, Y.; Cheng, C.; Peng, J.; and Ma, J. 2022. Equivariant point cloud analysis via learning orientations for message passing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 18932–18941.
- Melnyk, P.; Robinson, A.; Felsberg, M.; and Wadenbäck, M. 2024. Tetrasphere: A neural descriptor for o(3)-invariant point cloud analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5620–5630.

- Poulenard, A.; and Guibas, L. J. 2021. A functional approach to rotation equivariant non-linearities for Tensor Field Networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 13174–13183.
- Qi, C. R.; Su, H.; Mo, K.; and Guibas, L. J. 2017a. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 652–660.
- Qi, C. R.; Yi, L.; Su, H.; and Guibas, L. J. 2017b. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, 30.
- Uy, M. A.; Pham, Q.-H.; Hua, B.-S.; Nguyen, T.; and Yeung, S.-K. 2019. Revisiting point cloud classification: A new benchmark dataset and classification model on real-world data. In *Proceedings of the IEEE/CVF international conference on computer vision*, 1588–1597.
- Wang, Y.; Sun, Y.; Liu, Z.; Sarma, S. E.; Bronstein, M. M.; and Solomon, J. M. 2019. Dynamic graph cnn for learning on point clouds. *ACM Transactions on Graphics (tog)*, 38(5): 1–12.
- Wu, C.; Wan, Y.; Fu, H.; Pfrommer, J.; Zhong, Z.; Zheng, J.; Zhang, J.; and Beyerer, J. 2025. SAMBLE: Shape-Specific Point Cloud Sampling for an Optimal Trade-Off Between Local Detail and Global Uniformity. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 1342–1352.
- Wu, C.; Zheng, J.; Pfrommer, J.; and Beyerer, J. 2023. Attention-based point cloud edge sampling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5333–5343.
- Wu, Z.; Song, S.; Khosla, A.; Yu, F.; Zhang, L.; Tang, X.; and Xiao, J. 2015. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1912–1920.
- Xu, J.; Tang, X.; Zhu, Y.; Sun, J.; and Pu, S. 2021. Sgmnet: Learning rotation-invariant point cloud representations via sorted gram matrix. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 10468–10477.
- Yi, L.; Kim, V. G.; Ceylan, D.; Shen, I.-C.; Yan, M.; Su, H.; Lu, C.; Huang, Q.; Sheffer, A.; and Guibas, L. 2016. A scalable active framework for region annotation in 3d shape collections. *ACM Transactions on Graphics (ToG)*, 35(6): 1–12.
- Yu, H.; Hou, J.; Qin, Z.; Saleh, M.; Shugurov, I.; Wang, K.; Busam, B.; and Ilic, S. 2024. Riga: Rotation-invariant and globally-aware descriptors for point cloud registration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(5): 3796–3812.
- Yu, H.; Qin, Z.; Hou, J.; Saleh, M.; Li, D.; Busam, B.; and Ilic, S. 2023. Rotation-invariant transformer for point cloud matching. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5384–5393.
- Yu, J.; Zhang, C.; and Cai, W. 2023. Rethinking rotation invariance with point cloud registration. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 3313–3321.
- Zhang, D.; Yu, J.; Zhang, C.; and Cai, W. 2023. Parot: Patch-wise rotation-invariant network via feature disentanglement and pose restoration. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 3418–3426.
- Zhang, Z.; Hua, B.-S.; Rosen, D. W.; and Yeung, S.-K. 2019. Rotation invariant convolutions for 3d point clouds deep learning. In *2019 International conference on 3d vision (3DV)*, 204–213. IEEE.
- Zhang, Z.; Yang, L.; and Xiang, Z. 2024. Risurconv: Rotation invariant surface attention-augmented convolutions for 3d point cloud classification and segmentation. In *European Conference on Computer Vision*, 93–109. Springer.
- Zhao, C.; Yang, J.; Xiong, X.; Zhu, A.; Cao, Z.; and Li, X. 2022. Rotation invariant point cloud analysis: Where local geometry meets global topology. *Pattern Recognition*, 127: 108626.
- Zhou, H.; Feng, Y.; Fang, M.; Wei, M.; Qin, J.; and Lu, T. 2021. Adaptive graph convolution for point cloud analysis. In *Proceedings of the IEEE/CVF international conference on computer vision*, 4965–4974.