This CVPR paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

# 4D Local Modeling Toward Dynamic Global Perception for Ambiguity-free Rotation-Invariant Point Cloud Analysis

Jiaxun Guo[1] Wentao Fan[2*] Manar Amayri[1] Nizar Bouguila[1]
[1]CIISE, Concordia University
[2]Guangdong Provincial/Zhuhai Key Laboratory IRADS and Department of Computer Science, Beijing Normal-Hong Kong Baptist University
jiaxun.guo@mail.concordia.ca, wentaofan@bnbu.edu.cn,
{manar.amayri, nizar.bouguila}@concordia.ca

## Abstract

*Rotation invariance remains a core challenge in point cloud analysis, where existing methods often struggle with structural ambiguities and insufficient global context. Most rotation-invariant (RI) representations are derived from local coordinate systems, which inherently suffer from point-pair ambiguities and fail to capture discriminative features in symmetric or repetitive structures, while discarding informative global pose cues. To overcome these limitations, we propose* ***Ga4DPF****, a novel framework that offers a robust, global-aware RI representation by converting rotation-equivariant geometric representations into invariant ones, while concurrently integrating global pose awareness. Specifically, Ga4DPF introduces a learnable steerable transform that equivariantly lifts point clouds into 4D space, facilitating robust local feature construction and mitigating point-pair ambiguities. Concurrently, we model a dynamic global pose reference using the Bingham distribution, which adaptively estimates a consistent global rotation and enhances global feature discriminability. Extensive experiments on multiple benchmark datasets demonstrate that Ga4DPF achieves state-of-the-art performance with high computational efficiency, offering a new paradigm for rotation-invariant point cloud analysis. The source code is available at:* https://github.com/jiaxunguo/ga4dpf.

## 1. Introduction

The rapid advancement of 3D sensing technologies (such as LiDAR, stereo, and time-of-flight sensors) has led to the widespread adoption of point clouds in a variety of 3D vision tasks, including shape classification, segmentation, and object recognition [1, 10, 16]. As a compact geometric representation, point clouds are especially favoured in safety-critical domains such as autonomous driving and robotics [14, 33]. However, most existing deep learning models implicitly assume that input point clouds are consistently aligned to a canonical orientation [21, 31, 36]. In real-world settings, objects are frequently observed under arbitrary and unpredictable rotations, and even extensive data augmentation with random rotations cannot guarantee true robustness. This lack of inherent rotation invariance significantly degrades model performance and limits their deployment in unconstrained environments.

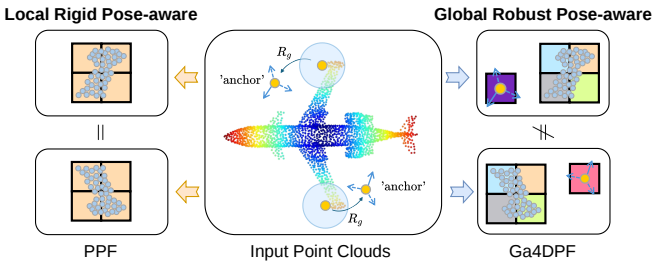


Figure 1. Traditional local pose-aware descriptors (e.g., PPF [9]) exhibit limited expressiveness and are prone to ambiguity in symmetric structures. In contrast, our approach enhances local geometric representation by lifting point-pair features into a higher-dimensional 4D space, enabling more robust and distinctive feature encoding. By further incorporating a global anchor derived from a dynamic global rotation $R_g$, Ga4DPF achieves more discriminative and rotation-consistent representations.

To address this issue, recent studies have focused on constructing rotation-invariant representations that eliminate the dependency on canonical alignment. These approaches often utilize auxiliary modules to transform coordinate frames [3, 41] or encode pairwise geometric relationships within enlarged local neighborhoods [19, 39].

*Corresponding author.

Although effective at achieving rotation invariance, these methods typically incur increased computational overhead and reduced performance on already aligned data. A fundamental challenge that persists is the issue of *structural ambiguity* in RI feature construction. First, rigid local RI encoding schemes lack robustness. The different spatial arrangements may yield nearly identical RI descriptors, which motivates continued refinement of local representations [4, 44]. Second, the use of local reference frames inherently discards global pose information, which is critical for disambiguating structures that are locally similar but globally distinct. While advances in expressive equivariant architectures [7, 27], existing RI frameworks remain limited by unstable local features and global ambiguity.

Our method is an equivariance-to-invariance framework that directly addresses the structural ambiguity of existing RI methods. Drawing inspiration from the observation that equivariant representations retain more structural information than invariant ones [11, 18], we design a representation pipeline that lifts point clouds into a 4D space using steerable 3D spherical neurons [26]. This lifting operation retains critical geometric structures while offering flexibility across varying viewpoints. Crucially, the lifting is fully learnable, enabling adaptive and elastic local feature construction beyond rigid coordinate dependence, thus enhancing the stability of RI representations. To further resolve global ambiguity, we introduce a novel mechanism that assigns each point a learnable global *anchor*, derived from a statistically consistent global rotation modeled by the Bingham distribution. As illustrated in Figure 1, this anchor injects global context into local neighborhoods, enabling the model to disambiguate symmetric or locally similar but globally distinct structures. By integrating robust 4D local representations with dynamic global pose awareness, our framework produces ambiguity-free RI features while remaining compatible with standard RI learning pipelines. We validate the effectiveness of our method across multiple benchmark datasets on both classification and part segmentation tasks. Experimental results demonstrate that our approach significantly outperforms existing state-of-the-art RI methods in both performance and generalization capability.

Our main contributions are summarized as follows:

- We identify and formalize the intrinsic ambiguities in existing RI representations, both at local and global levels, revealing their limitations in learning robust and discriminative features under arbitrary rotations.
- We propose a novel 4D statistical framework that leverages steerable spherical neurons to construct rotation-equivariant local features in a lifted space, offering a principled solution for point-pair encoding in 4D and significantly reducing the ambiguity of spatial arrangements.
- We introduce a learnable global pose reference based on the Bingham distribution, jointly optimized with task-specific objectives, to dynamically capture and utilize global pose information.

## 2. Related Works

### 2.1. Rotation-Sensitive 3D Point Cloud Analysis

Early point cloud processing methods, such as PointNet [30] and its hierarchical extension PointNet++ [31], operate directly on raw 3D coordinates but are inherently sensitive to rotations due to their reliance on absolute spatial positions. Subsequent models like PointCNN [20] and DGCNN [36] introduced learnable operators and graph-based neighborhood aggregation to capture local geometry. More recent approaches, such as CurveNet [38] and methods grounded in conformal geometry [26], integrate stronger geometric priors. However, they still process in the original coordinate space, limiting their robustness to arbitrary orientation changes.

A complementary line of research attempts to address rotation sensitivity via canonical pose alignment. These methods predict a transformation that aligns each input to a consistent canonical orientation, often using modules like Spatial Transformer Networks (STNs) [15], self-consistent regressors [46], or PCA-based normalization [19, 41]. While effective to some extent, these methods typically assume category-level consistency and rely on extensive rotation augmentation during training, which limits their generalizability to unseen categories or partial observations.

Although these rotation-sensitive models perform well under aligned conditions, their performance deteriorates significantly under arbitrary rotations. This motivates the development of rotation-aware approaches that are inherently robust to orientation changes.

### 2.2. Rotation-Aware Methods

To overcome the challenges posed by arbitrary rotations in 3D point clouds, a range of rotation-aware approaches have been proposed. These techniques eliminate the assumption of pre-aligned input and can be broadly categorized into rotation-invariant and rotation-equivariant methods.

Rotation-invariant methods aim to produce features that remain unchanged under any 3D rotation. A dominant paradigm in RI research is local patch modeling, which constructs invariant features by encoding pairwise or patch-level geometric relationships. One common approach defines Local Reference Frames (LRFs) for each point and computes descriptors based on relative geometry [19, 39]. Classical examples include Point Pair Features (PPF) [9] and their variants, which encode distances and angles between point pairs in a rotation-independent manner. These methods eliminate the need for alignment and provide finer local geometric characterization. However, the construction of rigid local frames is sensitive to noise and perturbations,

often leading to structural ambiguity in symmetric or repetitive patterns and reduced robustness of RI representations.

Rotation-equivariant methods, on the other hand, preserve rotation-dependent information by ensuring that features transform consistently with the input under rotations. Tensor field networks [12, 29] achieve strict equivariance through spherical harmonics, while recent vector-based approaches [32, 34] offer more flexible and lightweight alternatives. These methods maintain local structural information and exhibit strong expressive power. However, they often lack explicit mechanisms for enforcing full rotation invariance, which may limit their effectiveness in applications requiring orientation-independent predictions.

Despite the complementary strengths of RI and equivariant methods, few works have successfully integrated both perspectives. A more effective strategy, motivated by recent theoretical insights [11, 18], is to first build a rotation-equivariant representation that preserves geometric structure and then derive an invariant representation from it. Following this principle, we propose a unified framework that bridges the gap between equivariant and RI learning, enabling the construction of ambiguity-free representations.

## 3. Preliminaries

This section introduces essential concepts of equivariance and invariance, along with prior theoretical foundations. These preliminaries provide insight into point cloud transformations and the inherent limitations of local receptive fields in point cloud convolutional architectures.

Let a point cloud be denoted as $P = \{\mathbf{p}_1, \dots, \mathbf{p}_N\} \in \mathbb{R}^{N\times 3}$, and let $\boldsymbol{R} \in \mathrm{SO}(3)$ represent an arbitrary 3D rotation matrix. The rotated point cloud is defined as $P_R = P\boldsymbol{R}$.

### 3.1. Equivariance and Invariance

Let $G$ be a group, and let $\{T_g : P \to P\}_{g\in G}$ be a family of transformations acting on the point cloud. A mapping $f : P \to P^{'}$ is said to be *G-equivariant* if, for every $g \in G$, there exists a transformation $V_g : P^{'} \to P^{'}$ such that:

$$V_g\big[f(\mathbf{p})\big] \;=\; f\big(T_g[\mathbf{p}]\big) \quad \text{for all } g \in G,\ \mathbf{p} \in P. \tag{1}$$

*G-invariance* is the special case of equivariance in which $V_g$ is the identity map for all $g \in G$, i.e.,

$$f(\mathbf{p}) \;=\; f\big(T_g[\mathbf{p}]\big) \quad \text{for all } g \in G,\ \mathbf{p} \in P. \tag{2}$$

To apply transformation $T_g$ on a 3D point $\mathbf{p} \in \mathbb{R}^3$, the group elements $g \in \mathrm{SO}(3)$ are typically represented by $3 \times 3$ rotation matrices $\boldsymbol{R}$ [5].

### 3.2. Receptive Fields in Point Cloud Convolutions

Local *receptive fields* are commonly used in point cloud learning to capture geometric context. Let $X = \{x_1, ..., x_N\} \in \mathbb{R}^{c_{\text{in}}}$ denote the input feature set associated with point cloud $P$. For a reference point $\mathbf{p}_r$, its local neighborhood is denoted by $\mathcal{N}(\mathbf{p}_r)$, and the stacked neighbor feature matrix is defined as $\mathbf{X}_r = [x_1, \dots, x_k]^\top \in \mathbb{R}^{k\times c_{\text{in}}}$.

Point cloud convolutions compute output features $X^{'} = \{x_1^{'}, ..., x_N^{'}\} \in \mathbb{R}^{c_{\text{out}}}$ at each point $\mathbf{p}_r$ by aggregating transformed features from its local neighborhood:

$$f(\mathbf{p}_r) = \bigwedge_{j\in\mathcal{N}(\mathbf{p}_r)} g(W_j^r \cdot h(\mathbf{p}_j \mid \Omega(\mathbf{p}_j)),\ x_r), \tag{3}$$

where $h$ is a nonlinear function that maps the $\mathbf{p}_j$ to feature $x_j$, $\Omega(\mathbf{p}_j)$ denote the local patch centered at point $\mathbf{p}_j$, $g$ is a one-layer MLP for aggregation [36], $W_j^r$ denotes a learnable or dynamic kernel weight for the pair $(\mathbf{p}_r, \mathbf{p}_j)$, and $\wedge$ represents a permutation-invariant aggregation operator such as MAX, SUM, or AVG.

Since the function $h$ also depend on the neighbourhood of $\mathbf{p}_j$, the effective receptive field of $\mathbf{p}_r$ becomes the union of its neighbors' patches:

$$\Gamma(\mathbf{p}_r) \;=\; \bigcup_{j\in\mathcal{N}(\mathbf{p}_r)} \Omega(\mathbf{p}_j). \tag{4}$$

While local receptive fields capture rich geometric structure, they are inherently limited in distinguishing globally identical configurations. For instance, consider two symmetric points $\mathbf{p}_{\text{left}}, \mathbf{p}_{\text{right}} \in P$ located on opposite sides of an airplane wing. There exists a rotation $\boldsymbol{R}_{\text{sym}} \in \mathrm{SO}(3)$ such that their local receptive fields are rotationally equivalent:

$$\Gamma(\mathbf{p}_{\text{right}}) = \Gamma(\mathbf{p}_{\text{left}})\,\boldsymbol{R}_{\text{sym}}. \tag{5}$$

As a result, the network produces identical features for both points, i.e., $f(\mathbf{p}_{\text{left}}) = f(\mathbf{p}_{\text{right}})$, leading to *global ambiguity* and the loss of critical global pose information.

## 4. The Proposed Method

In this section, we propose an ambiguity-free framework for RI point cloud analysis that explicitly addresses both local and global symmetry challenges. Our method consists of two key components: (1) A lifting mechanism that projects 3D point clouds into 4D via steerable spherical neurons, enabling robust and expressive point-pair representations; and (2) A global-aware module based on a Bingham-inspired formulation that complements local receptive fields by resolving global symmetry ambiguity while preserving rotation invariance. An overview of the full architectural pipeline is illustrated in Figure 2.

### 4.1. 4D Robust Modeling for Local Representation

**LRF Construction.** LRFs are fundamental to RI feature extraction. For each point $\mathbf{p}_r \in P$, we construct an LRF $\mathcal{L}_r = \{\partial_r^1, \partial_r^2, \partial_r^3\} \in \mathbb{R}^{3\times 3}$ via Gram–Schmidt orthonormalization using two directional vectors $e_r^1$ and $e_r^2$:

$$\partial_r^1 = e_r^1, \quad \partial_r^3 = \frac{\partial_r^1 \times e_r^2}{\|\partial_r^1 \times e_r^2\|}, \quad \partial_r^2 = \partial_r^3 \times \partial_r^1. \tag{6}$$

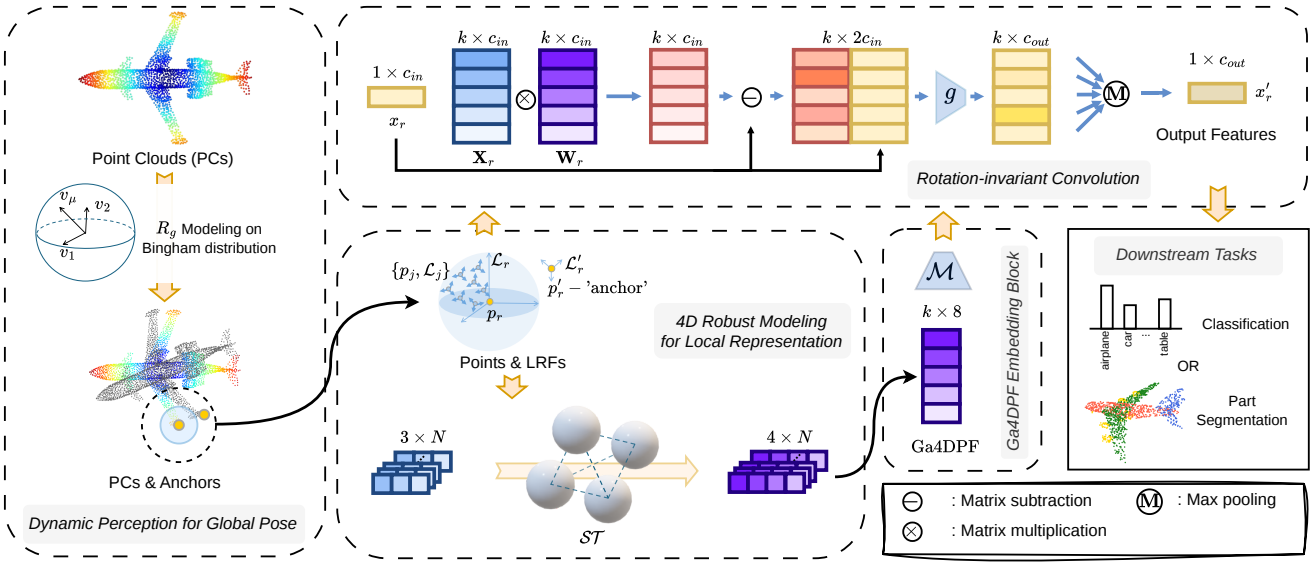


Figure 2. The pipeline of the proposed method. Our framework consists of two key components: (1) Dynamic Perception for Global Pose, which addresses the inherent *ambiguity of symmetric local structures* by estimating a globally consistent reference rotation $R_g$ through Bingham distribution modeling, generating adaptive anchor points that provide additional global cues; (2) 4D Robust Modeling for Local Representation, where constructs LRFs $\{\mathcal{L}_j\}_{j=1}^{k}$ and a learnable steerable transform ($\mathcal{ST}$) lifts the point cloud into 4D space, enabling fine-grained geometric encoding and enhancing the robustness of local representations against *point-pair ambiguity*. Finally, a Rotation-Invariant Convolution module embeds and aggregates the resulting Ga4DPF features, producing discriminative representations suitable for downstream tasks such as shape classification and part segmentation.

Following prior works [3, 44], we adopt a purely local construction where $e_r^2$ is defined as the vector from the barycenter of the $k$-nearest neighbours $\mathbf{m}_r = \frac{1}{k}\sum_{j=1}^{k} \mathbf{p}_j^r$ to the reference point $\mathbf{p}_r$, i.e., $e_r^2 = \mathbf{m}_r - \mathbf{p}_r$, and $e_r^1 = n_r$ is the surface normal. This construction ensures geometric consistency and forms the input feature $X$ in our method.

**Steerable 3D Spherical Neurons.** A steerable spherical neuron [26] consists of a learnable spherical decision surface $\mathbf{S} \in \mathbb{R}^5$ [25, 28] and three rotated variants located at the vertices of a regular tetrahedron. Let $\boldsymbol{R}_O$ denote the geodesic rotation that moves the learned center $\mathbf{c}_0$ to $\frac{\|\mathbf{c}_0\|}{\sqrt{3}}(1,1,1)$, and let $\{\boldsymbol{R}_{T_i}\}_{i=0}^{3}$ denote the rotation isomorphisms in $\mathbb{R}^5$ that rotate from $(1,1,1)$ to each tetrahedral vertex (with $\boldsymbol{R}_{T_0} = \mathbf{I}_5$). The resulting $4 \times 5$ filter bank is defined as:

$$B(\mathbf{S}) = \left[(\boldsymbol{R}_O^\top \boldsymbol{R}_{T_i} \boldsymbol{R}_O\, \mathbf{S})^\top\right]_{i=0\ldots3}, \tag{7}$$

which can be viewed as a mapping $f_{4S}(\cdot;\mathbf{S}) : \mathbb{R}^5 \to \mathbb{R}^4$, parameterized by five learnable weights. Given an embedded input point $\mathbf{P} \in \mathbb{R}^5$ and a 3D rotation representation $\boldsymbol{R} \in \mathbb{R}^5$, the filter bank satisfies the equivariance condition [25]:

$$V_{\boldsymbol{R}}\, B(\mathbf{S})\, \mathbf{P} \;=\; B(\mathbf{S})\, \boldsymbol{R}\, \mathbf{P}, \tag{8}$$

where the induced $4 \times 4$ action on the output is:

$$V_{\boldsymbol{R}} = \mathbf{M}^\top \boldsymbol{R}_O\, \boldsymbol{R}\, \boldsymbol{R}_O^\top \mathbf{M},$$
$$\mathbf{M} = \frac{1}{2}\begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}. \tag{9}$$

Thus, $V_{\boldsymbol{R}} \in G \subset \mathrm{SO}(4)$ represents the rotation group acting in the output space defined by the tetrahedral basis. We use this equivariant filter output as a 4D replacement for the original 3D points.

**4D Point Pair Features (4DPPF).** We define the 4D lifting operation as a steerable transform $\mathcal{ST}(\cdot;\mathbf{S}) : \mathbb{R}^{N\times3} \to \mathbb{R}^{N\times4}$. Following standard point cloud processing pipelines [30, 36], the operation $B(\mathbf{S})\,\mathbf{P}$ is applied uniformly across points, yielding a rotation-equivariant output $\mathbf{P}' = \mathcal{ST}(\mathbf{P};\mathbf{S}) \in \mathbb{R}^{N\times4}$.

To encode local geometric relationships in this 4D space, we define the 4D point pair feature between a reference point $\mathbf{p}'_r$ and its neighbour $\mathbf{p}'_j$ as:

$$\text{4DPPF}(\mathbf{p}'_r, \mathbf{p}'_j) = (\|d\|_2,\ \cos(\alpha_1),\ \cos(\alpha_2),\ \cos(\alpha_3)),$$
$$\alpha_1 = \angle(\Delta_r^1, d), \quad \alpha_2 = \angle(\Delta_j^1, d), \quad \alpha_3 = \angle(\Delta_r^1, \Delta_j^1), \tag{10}$$

where $d = \mathbf{p}'_j - \mathbf{p}'_r$, $\angle(\cdot,\cdot)$ denotes the angle between two 4D vectors, and $\Delta = \mathcal{ST}(\partial;\mathbf{S})$ represents the orthonormal

basis vectors lifted to 4D space.

While Equation 10 resembles the classic 3D PPF formulation [9], it fundamentally enhances discriminability by leveraging 4D geometry. Unlike 3D PPFs, which often yield identical descriptors for symmetric point pairs [3, 8], our 4DPPF naturally resolves these ambiguities due to its richer representational capacity.

**Theorem 1** (Disambiguation Guarantee of 4DPPF.)**.** *Let $\boldsymbol{p}_r$ be the reference point with local axis $\partial_r^1$. Consider a set of neighbors $\boldsymbol{p}_j(\theta)$ obtained by rotating a fixed point $\boldsymbol{p}_j(0)$ around $\partial_r^1$ by angle $\theta$, the lifted 4D point pair feature satisfies:*

$$4DPPF\big(\boldsymbol{p}_r', \boldsymbol{p}_j'\big) \neq 4DPPF\big(\boldsymbol{p}_r', \boldsymbol{p}_j'(\theta)\big), \quad \theta \neq 0, \tag{11}$$

*where $\boldsymbol{p}' = \mathcal{ST}(\boldsymbol{p})$ and $V_R(\theta) = M^\top R_O R(\theta) R_O^\top M$ is the induced 4D rotation.*

*Proof.* For points $\mathbf{p}_r, \mathbf{p}_j \in \mathbb{R}^3$, if $\mathbf{p}_r$ lies on the axis of rotation, then:

$$\langle R(\theta)\mathbf{p}_j,\, \mathbf{p}_r\rangle = \langle \mathbf{p}_j,\, R(\theta)^\top \mathbf{p}_r\rangle = \langle \mathbf{p}_r,\, \mathbf{p}_j\rangle, \tag{12}$$

implying that $\|d(\theta)\|_2$ and the 3D PPF remain unchanged with $\theta$, and $\langle\cdot,\cdot\rangle$ denotes the inner product of two vectors. In contrast, for lifted 4D points $\mathbf{p}_r', \mathbf{p}_j' \in \mathbb{R}^4$, we have:

$$\langle V_R(\theta)\mathbf{p}_j',\, \mathbf{p}_r'\rangle = \langle \hat{\mathbf{p}}_j,\, R_O R(-\theta) R_O^\top \hat{\mathbf{p}}_r\rangle, \tag{13}$$

where $\hat{\mathbf{p}} = M\mathbf{p}'$. Unlike the 3D case, $\mathbf{p}_r'$ is generally *not* an eigenvector of $V_R(\theta)$, hence $R_O R(-\theta) R_O^\top \hat{\mathbf{p}}_r \neq \hat{\mathbf{p}}_r$ for $\theta \neq 0$, and the inner product varies with $\theta$, leading to:

$$\|d(\theta)\|_2 = \sqrt{\|\mathbf{p}_j'\|_2^2 + \|\mathbf{p}_r'\|_2^2 - 2\langle \mathbf{p}_j',\, V_R(\theta)^T \mathbf{p}_r'\rangle}. \tag{14}$$

Likewise, the basis vectors $\Delta_j^1(\theta) = V_R(\theta)\Delta_j^1(0)$ and $\hat{\Delta}_r^1$ satisfy $R_O R(-\theta) R_O^\top \hat{\Delta}_r^1 \neq \hat{\Delta}_r^1$, so the angles $\cos(\alpha_1)$, $\cos(\alpha_2)$, $\cos(\alpha_3)$ also vary with $\theta$. Therefore,

$$\text{4DPPF}\big(\mathbf{p}_r', \mathbf{p}_j'\big) \neq \text{4DPPF}\big(\mathbf{p}_r', \mathbf{p}_j'(\theta)\big) \tag{15}$$

for $\theta \neq 0$, except on a degenerate set of measure zero. □

### 4.2. Dynamic Perception for Global Pose

**Global-Aware Pose Feature.** While local 4DPPF features effectively capture fine-grained geometric relationships, they are inherently constrained by the local receptive fields. As a result, they may fail to distinguish symmetric configurations due to a lack of global context, an issue we refer to as *global ambiguity*. To overcome this limitation, we propose the *Global-aware 4D Pose Feature* (Ga4DPF), which augments local descriptors with a global reference. Formally, Ga4DPF is defined as:

$$\text{Ga4DPF}(\mathbf{p}_r) = \big(\text{4DPPF}(\mathbf{p}_r', \mathbf{p}_j'),\ \text{4DPPF}(\mathbf{p}_r^\dagger, \mathbf{p}_j')\big) \in \mathbb{R}^8, \tag{16}$$

where the *anchor point* is defined as $\mathbf{p}_r^\dagger = \mathcal{ST}(\mathbf{p}_r \boldsymbol{R}_g; \mathbf{S})$, with $\boldsymbol{R}_g \in \mathrm{SO}(3)$ a consistent global rotation defined later. Analysis of how Ga4DPF$(\mathbf{p}_r)$ achieves rotation invariance can be found in the Supplementary.

By incorporating the anchor $p_r^\dagger$, the receptive field $\Gamma(\mathbf{p}_r)$ is no longer restricted to the local neighbourhood, but aligned with a global reference. This breaks the local equivariance described in Equation 5, i.e., $\Gamma(\mathbf{p}_r, \mathbf{p}_r^\dagger) \neq \Gamma(\mathbf{p}_r, \mathbf{p}_r^\dagger)\boldsymbol{R}_{\text{sym}}$, and injects global pose awareness into the representation:

$$\text{Ga4DPF}(\mathbf{p}_r) \neq \text{Ga4DPF}(\mathbf{p}_r \boldsymbol{R}_{\text{sym}}). \tag{17}$$

**Dynamic Global Rotation.** The anchor point $\mathbf{p}_r^\dagger$ is constructed through a global rotation $\boldsymbol{R}_g$, which is applied consistently across the dataset. However, using a fixed $\boldsymbol{R}_g$ may not improve discriminability, particularly when $\boldsymbol{R}_g$ coincides with a scene symmetry transformation $\boldsymbol{R}_{\text{sym}}$ as shown in Equation 5, in which case Ga4DPF cannot effectively disambiguate symmetric local regions.

To overcome this, we propose a *dynamic strategy* that models uncertainty in choosing $\boldsymbol{R}_g$ by employing the *Bingham distribution* $\mathcal{B}(q \mid \mathbf{V}, \mathbf{\Lambda})$ over the unit quaternion hypersphere $\mathbb{S}^3$, where $q$ denotes a global rotation encoded as a unit quaternion. The Bingham distribution is defined as:

$$\mathcal{B}(q \mid \mathbf{V}, \mathbf{\Lambda}) = \frac{1}{F(\mathbf{\Lambda})} \exp\left(q^\top \mathbf{V}\mathbf{\Lambda}\mathbf{V}^\top q\right), \tag{18}$$

where $F(\mathbf{\Lambda})$ is the normalization constant, $\mathbf{V} \in \mathbb{R}^{4\times4}$ is an orthogonal matrix that specifies the mode and dispersion axes, and $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \lambda_3, 0)$ controls the concentration with $\lambda_1 < \lambda_2 \leq \lambda_3 < 0$.

To couple global pose learning with downstream task performance, we jointly optimize the Bingham parameters and network using the following objective:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{task}} + \delta \cdot \sqrt{(\mathcal{L}_{\text{bingham}} - 0.1 \cdot \mathcal{L}_{\text{task}})^2}, \tag{19}$$

where $\mathcal{L}_{\text{bingham}}$ is the negative log-likelihood of sampled quaternions, and $\delta$ is a balancing hyperparameter (empirically set to 0.8). The subtraction term encourages consistency between global pose adaptivity and task performance.

To enable gradient-based learning, we adopt a sampling-based approximation of the Bingham distribution. Given learned $\mathbf{V}$ and $\mathbf{\Lambda}$, we sample $N_s$ quaternions $\{q_i\}_{i=1}^{N_s} \sim \mathcal{B}(q \mid \mathbf{V}, \mathbf{\Lambda})$, and compute the global rotation $\boldsymbol{R}_g = \mathrm{Rot}(q_g)$, where the representative quaternion $q_g$ is estimated via the geometric mean:

$$q_g = \arg\max_{q\in\mathbb{S}^3} \sum_{i=1}^{N_s} \langle q, q_i\rangle^2, \tag{20}$$

which corresponds to the dominant eigenvector of the matrix $\sum_i q_i q_i^\top$. In practice, we use $N_s = 10$.

### 4.3. Network Architecture

To validate the robustness and effectiveness of the proposed 4DPPF and dynamic global rotation, we integrate them into the PaRI-Conv framework [3] and evaluate them using two widely adopted point cloud backbones for fair evaluation.

**Shape Classification.** For classification, we employ DGCNN [36] as the backbone and replace all EdgeConv layers with PaRI-Conv modules enhanced by Ga4DPF. At each layer, we construct a $k$-nearest neighbor graph in feature space using Euclidean distance, with $k = 20$.

**Shape Part Segmentation.** For part segmentation, we follow the AdaptConv design [45]. Specifically, we stack five Ga4DPF-augmented PaRI-Conv layers, interleaved with three intermediate pooling layers to progressively expand the receptive field. Across all layers, the $k$-nearest neighbor graph is constructed in Euclidean space with $k = 40$.

## 5. Experiments

We evaluate the proposed method on three widely used benchmarks: ModelNet40 [37] for shape classification, ShapeNetPart [40] for part segmentation, and ScanObjectNN [35] for real-world object classification. To systematically assess rotation robustness, we adopt three standard evaluation protocols: z/z, z/SO(3), and SO(3)/SO(3), where z and SO(3) denote upright and arbitrary 3D rotations, respectively. In addition, we perform comprehensive ablation studies to investigate the contribution of each key component in our framework. Further experimental results and analysis are provided in the Supplementary.

### 5.1. Implementation Details

We adopt stochastic gradient descent (SGD) for optimization, with an initial learning rate of 0.1, decayed to 0.001 via cosine annealing [22]. All models are trained for 300 epochs. The batch size is set to 32 for classification and 16 for segmentation, with a dropout rate of 0.5 applied to fully connected layers for regularization.

To ensure RI initialization, we define a point-wise descriptor for each point $\mathbf{p}_i$ based on its relative geometry to the global centroid $O$ [6], comprising three elements: $\|\overrightarrow{O\mathbf{p}_i}\|_2$, $\sin(\angle(\partial_i^1, \overrightarrow{O\mathbf{p}_i}))$, and $\cos(\angle(\partial_i^1, \overrightarrow{O\mathbf{p}_i}))$. Here, $\partial_i^1$ denotes the local axis, typically the surface normal under pc+n setting. For coordinate-only settings, we replace $\partial_i^1$ with $\overrightarrow{\mathbf{p}_r\mathbf{m}_r}$, the direction from the barycenter to the point.

### 5.2. 3D Shape Classification

**Dataset.** We first evaluate our approach on the widely adopted ModelNet40 benchmark [37], which comprises 12,311 synthetic CAD models across 40 object categories, with 9,843 samples for training and 2,468 for testing. Following the standard protocol of [31], we uniformly sample 1,024 points from each shape. Random scaling and translation are applied to improve generalization during training.

| Methods | Inputs | z/z | z/SO(3) | SO(3)/SO(3) |
|---|---|---|---|---|
| Rotation-sensitive | | | | |
| PointNet [30] | pc | 89.2 | 16.2 | 75.5 |
| PointNet++ [31] | pc+n | 89.3 | 28.6 | 85.0 |
| DGCNN [36] | pc | 92.2 | 20.6 | 81.1 |
| Rotation-aware | | | | |
| RIConv [43] | pc | 86.5 | 86.4 | 86.4 |
| ClusterNet [2] | pc | 87.1 | 87.1 | 87.1 |
| RI-GCN [17] | pc+n | 91.0 | 91.0 | 91.0 |
| OrientedMP [24] | pc+n | 88.4 | 88.4 | 88.4 |
| ELGANet [13] | pc | 90.3 | 90.3 | 90.3 |
| PaRot [42] | pc | 90.9 | 91.0 | 90.8 |
| TetraSphere [27] | pc | 90.5 | 90.5 | 90.3 |
| **Ours** | pc | **91.9** | **91.9** | **91.9** |
| **Ours** | pc+n | **92.8** | **92.8** | **92.8** |

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

**Results.** Table 1 summarizes the classification accuracy on ModelNet40 under three protocols: z/z, z/SO(3), and SO(3)/SO(3), reflecting increasing rotational difficulty. Rotation-sensitive methods (e.g., PointNet++ [31], DGCNN [36]) perform well under z/z but degrade sharply under unseen rotations (e.g., PointNet++ drops from 89.3% to 28.6%). In contrast, RI-based methods such as RIConv [43] and ClusterNet [2] maintain stable accuracy (86%–87%) but are limited in capturing rich geometry using coordinates alone. More recent methods like ELGANet [13], PaRot [42], and TetraSphere [27] further improve performance (90.3%–90.5%) without requiring normals. Our method achieves 91.9% using only coordinates, outperforming several normal-dependent methods such as OrientedMP [24] (88.4%) and RI-GCN [17] (91.0%). With normals, performance increases to 92.8%, surpassing PaRot and RI-GCN by up to 1.8%. These results highlight the effectiveness of our 4D representation and global pose integration in learning expressive RI representations.

### 5.3. Real-World Shape Classification

**Dataset.** To evaluate performance under real-world conditions, we conduct experiments on the ScanObjectNN dataset [35], which contains over 15,000 point clouds scanned from 2,902 real-world indoor objects. We adopt the challenging OBJ_BG subset, consisting of 2,890 samples from 15 categories, split into 2,312 for training and 578 for testing. This subset introduces real-world challenges such as noise, occlusion, and background clutter, which significantly impact rotation stability. As normals are unavailable in the processed data, only raw point coordinates are used.

**Results.** Table 2 presents classification accuracy under z/SO(3) and SO(3)/SO(3) protocols. Our method achieves the highest accuracy across both settings, outperforming the previous best model, LocoTrans [4], by 2.4% and 2.8%,

| Methods | z/SO(3) | SO(3)/SO(3) |
|---|---|---|
| Input pc only | | |
| PointNet [30] | 16.7 | 54.7 |
| PointNet++ [31] | 15.0 | 47.4 |
| DGCNN [36] | 17.7 | 71.8 |
| RIConv [43] | 75.3 | 75.5 |
| RI-GCN [17] | 80.5 | 80.6 |
| OrientedMP [24] | 76.7 | 77.2 |
| PaRI-Conv [3] | 83.3 | 83.3 |
| LocoTrans [4] | 85.0 | 84.5 |
| **Ours** | **87.4** | **87.3** |

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

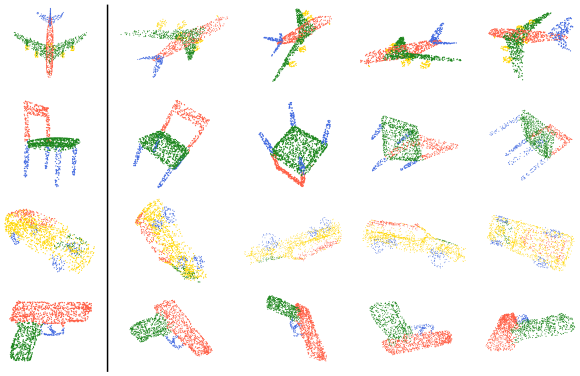

Figure 3. Visualization of part segmentation results on ShapeNet-Part Dataset under z/SO(3) setting. The leftmost column is the ground truth. The testing results under arbitrary rotations are on the rest columns.

respectively. This demonstrates strong generalization and robustness in real-world conditions. Compared to PaRI-Conv [3], which relies on rigid, handcrafted local reference frames and suffers under noise and occlusion (83.3%), our method dynamically lifts point–normal pairs into a 4D space and learns features using steerable spherical neurons. This flexible and learnable design enhances both local stability and geometric expressiveness. Moreover, the minimal performance gap (0.1%) between z/SO(3) and SO(3)/SO(3) further confirms the rotation invariance of our model.

### 5.4. Shape Part Segmentation

**Dataset.** We further evaluate our method on the ShapeNet-Part dataset [40], a large-scale benchmark for fine-grained 3D part segmentation. It contains 16,881 models across 16 object categories, each annotated with 2 to 6 semantic parts, resulting in a total of 50 unique part labels. Following the standard protocol of [31], we adopt the official train/test split and uniformly sample 2,048 points per shape as input.

**Results.** As shown in Table 3, our method achieves the highest mean Intersection-over-Union (mIoU) across all classes and all instances. Under the pc+n and z/SO(3) setting, our model achieves a class mIoU of 82.2%, surpassing RISurConv [44] (81.5%), demonstrating superior segmentation quality with normal vectors. Under the pc and z/SO(3) setting, our method attains a class mIoU of 81.3%, outperforming LocoTrans [4] by 1.2% (80.1%), highlighting the robustness of our learned rotation-invariant features even without normal input. Figure 3 visualizes representative segmentation results under the z/SO(3) protocol, further illustrating the model's robustness to arbitrary rotations.

| Method | Input | z/SO(3) | | SO(3)/SO(3) | |
|---|---|---|---|---|---|
| | | C. mIoU | I. mIoU | C. mIoU | I. mIoU |
| PointNet [30] | pc | 37.8 | - | 74.4 | - |
| PointNet++ [31] | pc | 48.2 | - | 76.7 | - |
| DGCNN [36] | pc | 37.4 | - | 73.3 | - |
| RI-GCN [17] | pc+n | 77.2 | - | 77.3 | - |
| PaRI-Conv [3] | pc+n | - | 84.6 | - | 84.6 |
| CRIN [23] | pc | 80.5 | - | 80.5 | - |
| PaRot [42] | pc | 79.2 | - | 79.5 | - |
| LocoTrans [4] | pc | 80.1 | 84.0 | 80.0 | 83.8 |
| TetraSphere [27] | pc | - | 82.3 | - | 82.3 |
| RISurConv [44] | pc+n | 81.5 | - | 81.5 | - |
| **Ours** | pc | **81.3** | **84.0** | **81.2** | **83.9** |
| **Ours** | pc+n | **82.2** | **85.1** | **82.6** | **85.0** |

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

### 5.5. Model Complexity

We compare the computational efficiency of our model against recent RI methods in Table 4. Our method achieves competitive segmentation performance while maintaining a favorable balance between accuracy and complexity. Specifically, the proposed model has only 2.25M parameters and 3841M FLOPs, significantly lower than LocoTrans [4] and RISurConv [44]. Compared to TetraSphere [27], which has a similar parameter count, our method improves I. mIoU by 1.7% while using fewer FLOPs. These results highlight the efficiency and scalability of our model in resource-constrained scenarios.

| | LocoTrans | TetraSphere | RISurConv | **Ours** |
|---|---|---|---|---|
| Params | 6.72M* | 1.31M | 4.06M | **2.25M** |
| FLOPs | 7998M* | 7996M | 12120M | **3841M** |
| C. mIoU | 80.1 | - | 81.3 | **81.3** |
| I. mIoU | 84.0 | 82.3 | - | **84.0** |

Table 4. Comparison of FLOPs, parameter count, and mIoU on ShapeNetPart (z/SO(3)) with point coordinates only. (* indicates the results are estimated based on the classification architecture.)

### 5.6. Ablation Studies

In this section, we conduct ablation analysis to assess the effectiveness of key components in our framework. Specifically, we analyze the influence of (i) the relative pose representation and (ii) the number of samples $N_s$ used in geometric mean estimation for global pose aggregation. All experiments are conducted under the z/SO(3) protocol for consistency and fairness.

| No. | RI Representation | Dim | MN40. | ScanObj. |
|---|---|---|---|---|
| 1 | PPF | 4 | 91.8 | 82.8 |
| 2 | Aug.PPF [3] | 8 | 92.4 | 83.3 |
| 3 | 4DPPF($\mathbf{p}'_r, \mathbf{p}'_j$) | 4 | 92.4 | 84.4 |
| 4 | (PPF, 4DPPF($\mathbf{p}^\dagger_r, \mathbf{p}'_j$)) | 8 | 92.5 | 86.7 |
| 5 | (4DPPF($\mathbf{p}'_r, \mathbf{p}'_j$), 4DPPF($\mathbf{p}^\dagger_r, \mathbf{p}'_j$)) | 8 | **92.8** | **87.4** |

Table 5. Ablation study on the relative pose feature. Reported results are classification accuracy under the z/SO(3) setting.

**Effect of Relative Pose Representation.** Table 5 presents a detailed comparison of different RI representations. The baseline PPF uses the classic point pair feature formulation [9]. Aug.PPF [3] incorporates handcrafted enhancements to mitigate ambiguity from symmetric point pairs (Equation 12). In contrast, our learned formulation 4DPPF($\mathbf{p}'_r$, $\mathbf{p}'_j$), derived from lifted point pairs using steerable spherical neurons, replaces handcrafted descriptors and achieves superior performance on ScanObjectNN, despite operating in a lower-dimensional feature space. The hybrid design (PPF, 4DPPF($\mathbf{p}^\dagger_r$, $\mathbf{p}'_j$)) combines the classic PPF with our global-aware variant and significantly outperforms both PPF and Aug.PPF alone. Finally, our full representation Ga4DPF($\mathbf{p}_r$), defined as (4DPPF($\mathbf{p}'_r$, $\mathbf{p}'_j$), 4DPPF($\mathbf{p}^\dagger_r$, $\mathbf{p}'_j$)), integrates both local and global-aware 4DPPFs. It consistently achieves the highest accuracy across both ModelNet40 and ScanObjectNN, validating the effectiveness of unifying local and global relational cues.

**Effect of Sampling Number $N_s$.** Table 6 investigates the impact of the sampling number $N_s$ used for estimating the geometric mean. We report classification accuracy on both ModelNet40 and ScanObjectNN under the z/SO(3) setting. Increasing $N_s$ from 1 to 10 yields consistent performance improvements, with the highest accuracy achieved at $N_s$ = 10, with 92.8% on ModelNet40 and 87.4% on ScanObjectNN. However, further increasing $N_s$ yields no improvement or slight degradation, indicating that $N_s$ = 10 offers an effective balance between accuracy and efficiency.

| $N_s$ | 1 | 5 | 10 | 15 | 20 |
|---|---|---|---|---|---|
| Acc.(ModelNet40) | 92.3 | 92.5 | **92.8** | 92.8 | 92.8 |
| Acc.(ScanObjectNN) | 85.5 | 85.9 | **87.4** | 87.2 | 86.9 |

Table 6. Ablation study on the sampling number $N_s$ for geometric mean estimation. Reported results are classification accuracy under the z/SO(3) setting.

### 5.7. Visualization of Learned RI Features

We visualize the dense point-wise features learned under various rotations in Figure 4. Compared to DGCNN [36], whose features exhibit significant variation under different poses, and PaRI-Conv [3], which shows improved consistency, our method produces the most rotation-invariant and semantically discriminative representations. Notably, our features remain stable across arbitrary rotations and demonstrate fine-grained distinctions between symmetric structures (such as airplane wings and engines) indicating enhanced local expressiveness. While PaRI-Conv also leverages LRFs to achieve rotation invariance, its reliance on handcrafted frame construction restricts its capacity to resolve structural ambiguities in symmetric geometries. In contrast, our model integrates a global pose anchor within the lifted 4D feature space, allowing more consistent feature alignment across poses and enabling compact yet expressive representation learning.

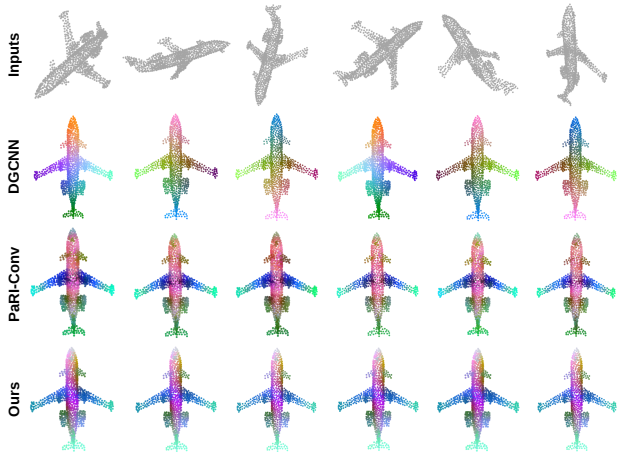


Figure 4. Visualization of learned dense features for shape part segmentation under different rotations. Our method (bottom row) yields more consistent and discriminative representations, especially for symmetric parts, while preserving rotation invariance.

## 6. Conclusion

In this work, we presented Ga4DPF, a robust and global-aware RI representation that directly addresses the long-standing challenge of structural ambiguity in RI point cloud analysis. Our key insight lies in bridging rotation equivariance and invariance through joint local–global modeling: a learnable steerable transform enables 4D lifting for robust local feature construction against point-pair ambiguity, while a dynamic global pose reference modeled by the Bingham distribution provides consistent orientation cues. By jointly eliminating both local and global ambiguities, Ga4DPF yields a compact, discriminative, and truly RI representation. Extensive experiments on multiple benchmarks validate the effectiveness and efficiency of our method, significantly outperforming existing RI methods and highlighting its potential as a principled direction for future RI representation learning. In future work, we plan to extend this representation to more complex scenes, explore its integration with generative models, and investigate its application in multimodal and spatiotemporal frameworks.

## 7. 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

[1] Mohamed Afham, Isuru Dissanayake, Dinithi Dissanayake, Amaya Dharmasiri, Kanchana Thilakarathna, and Ranga Rodrigo. 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*, pages 9902–9912, 2022. 1

[2] Chao Chen, Guanbin Li, Ruijia Xu, Tianshui Chen, Meng Wang, and Liang Lin. 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*, pages 4994–5002, 2019. 6

[3] Ronghan Chen and Yang Cong. 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*, pages 7472–7481, 2022. 1, 4, 5, 6, 7, 8

[4] Yiyang Chen, Lunhao Duan, Shanshan Zhao, Changxing Ding, and Dacheng Tao. Local-consistent transformation learning for rotation-invariant point cloud analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5418–5427, 2024. 2, 6, 7

[5] Gregory S Chirikjian and Alexander B Kyatkin. *Engineering applications of noncommutative harmonic analysis: with emphasis on rotation and motion groups*. CRC press, 2000. 3

[6] Taco S Cohen, Mario Geiger, Jonas Köhler, and Max Welling. Spherical cnns. In *International Conference on Learning Representations*, 2018. 6

[7] Congyue Deng, Or Litany, Yueqi Duan, Adrien Poulenard, Andrea Tagliasacchi, and Leonidas J Guibas. Vector neurons: A general framework for so (3)-equivariant networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12200–12209, 2021. 2

[8] Haowen Deng, Tolga Birdal, and Slobodan Ilic. Ppf-foldnet: Unsupervised learning of rotation invariant 3d local descriptors. In *Proceedings of the European conference on computer vision (ECCV)*, pages 602–618, 2018. 5

[9] Bertram Drost, Markus Ulrich, Nassir Navab, and Slobodan Ilic. Model globally, match locally: Efficient and robust 3d object recognition. In *2010 IEEE computer society conference on computer vision and pattern recognition*, pages 998–1005. Ieee, 2010. 1, 2, 5, 8

[10] Lunhao Duan, Shanshan Zhao, Nan Xue, Mingming Gong, Gui-Song Xia, and Dacheng Tao. Condaformer: Disassembled transformer with local structure enhancement for 3d point cloud understanding. *Advances in Neural Information Processing Systems*, 36:23886–23901, 2023. 1

[11] Carlos Esteves, Christine Allen-Blanchette, Ameesh Makadia, and Kostas Daniilidis. Learning so (3) equivariant representations with spherical cnns. In *Proceedings of the european conference on computer vision (ECCV)*, pages 52–68, 2018. 2, 3

[12] Fabian Fuchs, Daniel Worrall, Volker Fischer, and Max Welling. Se (3)-transformers: 3d roto-translation equivariant attention networks. *Advances in neural information processing systems*, 33:1970–1981, 2020. 3

[13] Ruibin Gu, Qiuxia Wu, Yuqiong Li, Wenxiong Kang, Wing WY Ng, and Zhiyong Wang. Enhanced local and global learning for rotation-invariant point cloud representation. *IEEE MultiMedia*, 29(4):24–37, 2022. 6

[14] Hanjiang Hu, Zuxin Liu, Sharad Chitlangia, Akhil Agnihotri, and Ding Zhao. Investigating the impact of multi-lidar placement on object detection for autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2550–2559, 2022. 1

[15] Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. Spatial transformer networks. *Advances in neural information processing systems*, 28, 2015. 2

[16] Zhao Jin, Munawar Hayat, Yuwei Yang, Yulan Guo, and Yinjie Lei. Context-aware alignment and mutual masking for 3d-language pre-training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10984–10994, 2023. 1

[17] Seohyun Kim, Jaeyoo Park, and Bohyung Han. Rotation-invariant local-to-global representation learning for 3d point cloud. *Advances in Neural Information Processing Systems*, 33:8174–8185, 2020. 6, 7

[18] Risi Kondor and Shubhendu Trivedi. On the generalization of equivariance and convolution in neural networks to the action of compact groups. In *International conference on machine learning*, pages 2747–2755. PMLR, 2018. 2, 3

[19] Xianzhi Li, Ruihui Li, Guangyong Chen, Chi-Wing Fu, Daniel Cohen-Or, and Pheng-Ann Heng. A rotation-invariant framework for deep point cloud analysis. *IEEE transactions on visualization and computer graphics*, 28 (12):4503–4514, 2021. 1, 2

[20] Yangyan Li, Rui Bu, Mingchao Sun, Wei Wu, Xinhan Di, and Baoquan Chen. Pointcnn: Convolution on x-transformed points. *Advances in neural information processing systems*, 31, 2018. 2

[21] Yongcheng Liu, Bin Fan, Shiming Xiang, and Chunhong Pan. Relation-shape convolutional neural network for point cloud analysis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8895–8904, 2019. 1

[22] Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts. In *International Conference on Learning Representations*, 2017. 6

[23] Yujing Lou, Zelin Ye, Yang You, Nianjuan Jiang, Jiangbo Lu, Weiming Wang, Lizhuang Ma, and Cewu Lu. Crin:

rotation-invariant point cloud analysis and rotation estimation via centrifugal reference frame. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 1817–1825, 2023. 7

[24] Shitong Luo, Jiahan Li, Jiaqi Guan, Yufeng Su, Chaoran Cheng, Jian Peng, and Jianzhu Ma. Equivariant point cloud analysis via learning orientations for message passing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18932–18941, 2022. 6, 7

[25] Pavlo Melnyk, Michael Felsberg, and Mårten Wadenbäck. Embed me if you can: A geometric perceptron. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1276–1284, 2021. 4

[26] Pavlo Melnyk, Michael Felsberg, and Mårten Wadenbäck. Steerable 3d spherical neurons. In *International Conference on Machine Learning*, pages 15330–15339. PMLR, 2022. 2, 4

[27] Pavlo Melnyk, Andreas Robinson, Michael Felsberg, and Mårten Wadenbäck. Tetrasphere: A neural descriptor for o (3)-invariant point cloud analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5620–5630, 2024. 2, 6, 7

[28] Christian Perwass, Vladimir Banarer, and Gerald Sommer. Spherical decision surfaces using conformal modelling. In *Joint Pattern Recognition Symposium*, pages 9–16. Springer, 2003. 4

[29] Adrien Poulenard and Leonidas J Guibas. 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*, pages 13174–13183, 2021. 3

[30] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 652–660, 2017. 2, 4, 6, 7

[31] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, 30, 2017. 1, 2, 6, 7

[32] Vıctor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E (n) equivariant graph neural networks. In *International conference on machine learning*, pages 9323–9332. PMLR, 2021. 3

[33] Corentin Sautier, Gilles Puy, Spyros Gidaris, Alexandre Boulch, Andrei Bursuc, and Renaud Marlet. Image-to-lidar self-supervised distillation for autonomous driving data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9891–9901, 2022. 1

[34] Zhuo Su, Max Welling, Matti Pietikäinen, and Li Liu. Svnet: Where so (3) equivariance meets binarization on point cloud representation. In *2022 International Conference on 3D Vision (3DV)*, pages 547–556. IEEE, 2022. 3

[35] Mikaela Angelina Uy, Quang-Hieu Pham, Binh-Son Hua, Thanh Nguyen, and Sai-Kit Yeung. 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*, pages 1588–1597, 2019. 6

[36] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E Sarma, Michael M Bronstein, and Justin M Solomon. Dynamic graph cnn for learning on point clouds. *ACM Transactions on Graphics (tog)*, 38(5):1–12, 2019. 1, 2, 3, 4, 6, 7, 8

[37] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1912–1920, 2015. 6

[38] Tiange Xiang, Chaoyi Zhang, Yang Song, Jianhui Yu, and Weidong Cai. Walk in the cloud: Learning curves for point clouds shape analysis. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 915–924, 2021. 2

[39] Jianyun Xu, Xin Tang, Yushi Zhu, Jie Sun, and Shiliang Pu. Sgmnet: Learning rotation-invariant point cloud representations via sorted gram matrix. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10468–10477, 2021. 1, 2

[40] Li Yi, Vladimir G Kim, Duygu Ceylan, I-Chao Shen, Mengyan Yan, Hao Su, Cewu Lu, Qixing Huang, Alla Sheffer, and Leonidas Guibas. A scalable active framework for region annotation in 3d shape collections. *ACM Transactions on Graphics (ToG)*, 35(6):1–12, 2016. 6, 7

[41] Ruixuan Yu, Xin Wei, Federico Tombari, and Jian Sun. Deep positional and relational feature learning for rotation-invariant point cloud analysis. In *European Conference on Computer Vision*, pages 217–233. Springer, 2020. 1, 2

[42] Dingxin Zhang, Jianhui Yu, Chaoyi Zhang, and Weidong Cai. Parot: Patch-wise rotation-invariant network via feature disentanglement and pose restoration. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 3418–3426, 2023. 6, 7

[43] Zhiyuan Zhang, Binh-Son Hua, David W Rosen, and Sai-Kit Yeung. Rotation invariant convolutions for 3d point clouds deep learning. In *2019 International conference on 3d vision (3DV)*, pages 204–213. IEEE, 2019. 6, 7

[44] Zhiyuan Zhang, Licheng Yang, and Zhiyu Xiang. Risurconv: Rotation invariant surface attention-augmented convolutions for 3d point cloud classification and segmentation. In *European Conference on Computer Vision*, pages 93–109. Springer, 2024. 2, 4, 7

[45] Haoran Zhou, Yidan Feng, Mingsheng Fang, Mingqiang Wei, Jing Qin, and Tong Lu. Adaptive graph convolution for point cloud analysis. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4965–4974, 2021. 6

[46] Yi Zhou, Connelly Barnes, Jingwan Lu, Jimei Yang, and Hao Li. On the continuity of rotation representations in neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5745–5753, 2019. 2