Hierarchical Direction Perception via Atomic Dot-Product Operators for Rotation-Invariant Point Clouds Learning

Chenyu Hu[1], Xiaotong Li[1*], Hao Zhu[1], Biao Hou[1]

[1]Xidian University, Xi'an, China

hcy299792@stu.xidian.edu.cn, {lixiaotong, haozhu}@xidian.edu.cn, avcodec@163.com

Abstract

Point cloud processing has become a cornerstone technology in many 3D vision tasks. However, arbitrary rotations introduce variations in point cloud orientations, posing a long-standing challenge for effective representation learning. The core of this issue is the disruption of the point cloud's intrinsic directional characteristics caused by rotational perturbations. Recent methods attempt to implicitly model rotational equivariance and invariance, preserving directional information and propagating it into deep semantic spaces. Yet, they often fall short of fully exploiting the multiscale directional nature of point clouds to enhance feature representations. To address this, we propose the Direction-Perceptive Vector Network (DiPVNet). At its core is an atomic dot-product operator that simultaneously encodes directional selectivity and rotation invariance–endowing the network with both rotational symmetry modeling and adaptive directional perception. At the local level, we introduce a Learnable Local Dot-Product (L2DP) Operator, which enables interactions between a center point and its neighbors to adaptively capture the non-uniform local structures of point clouds. At the global level, we leverage generalized harmonic analysis to prove that the dot-product between point clouds and spherical sampling vectors is equivalent to a direction-aware spherical Fourier transform (DASFT). This leads to the construction of a global directional response spectrum for modeling holistic directional structures. We rigorously prove the rotation invariance of both operators. Extensive experiments on challenging scenarios involving noise and large-angle rotations demonstrate that DiPVNet achieves state-of-the-art performance on point cloud classification and segmentation tasks.

Extended version — https://arxiv.org/abs/2511.08240
Code — https://github.com/wxszreal0/DiPVNet

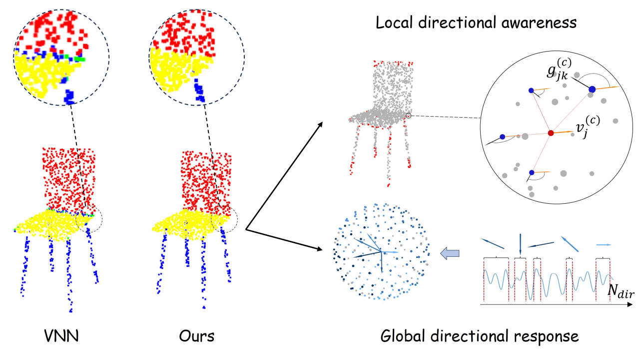

Figure 1: We extract direction-sensitive local features via the L2DP operator; concurrently, a global directional response spectrum is constructed through the DASFT module, capturing the directional characteristics of the overall structure.

1 Introduction

The widespread adoption of 3D sensing technologies has made point cloud processing a critical component in numerous 3D vision applications (Guo et al. 2020; Feng et al. 2023; Jia et al. 2025; Feng et al. 2024), including autonomous driving scene understanding and embodied AI interaction. However, conventional point cloud representation learning methods face a fundamental challenge in real-world deployments (Dym and Maron 2020): arbitrary rotations in 3D space can alter the spatial distribution of point clouds. As a result, traditional networks (Qi et al. 2017a,b; Xiang et al. 2021) may map different orientations of the same object to inconsistent feature representations, ultimately degrading the performance of downstream tasks under rotational disturbances.

At the root of this issue lies the directional nature of point cloud features across multiple scales. On the local scale, discriminative cues are embedded in features such as edge orientation and surface normals; on the global scale, high-level geometric priors arise from principal axis directions, interpart angles, and overall structural symmetries. These intrinsic directional characteristics are disrupted under arbitrary rotations, implying that robust point cloud representations must explicitly model and leverage such multiscale directional information to achieve rotation resilience.

To tackle this challenge, earlier methods focused on explicit modeling by independently extracting features along predefined directions—such as dividing local neighborhoods into directional sectors (Jiang et al. 2018) or introducing orientation density functions (ODF) (Sahin, Mertan, and Unal 2022)—thus encoding directional information in a manually structured manner. While this mitigated

*Corresponding author.

orientation-related inconsistencies to some extent, more recent implicit modeling paradigms incorporate the rotational symmetry of 3D space directly into the network design. These methods produce outputs that are either equivariant (rotate synchronously with the input)(Thomas et al. 2018; Schütt, Unke, and Gastegger 2021; Poulenard and Guibas 2021) or invariant (remain unchanged) (Zhang et al. 2019; Chou et al. 2021; Gu et al. 2021) under input rotations, thereby allowing directional information to be implicitly propagated to deeper semantic layers.

Although both paradigms have demonstrated promising results under rotational perturbations, they remain unsatisfactory in critical ways. Explicit approaches often rely on fixed directional partitions or handcrafted statistical schemes, lacking the adaptability to learn from non-uniform local point distributions, which limits the robustness and discriminability of their directional representations. Implicit approaches, while successfully preserving directional information through rotation-equivariant or invariant formulations, often fail to exploit this information efficiently at the feature representation level. For instance, in VNN (Deng et al. 2021), the non-linear layers are guided by a single learned global direction vector, which is used to gate vector neurons through inner products. This reliance on a single global direction fails to capture the complex hierarchical directional structure inherent in real-world point clouds.

To enable networks to preserve rotational symmetry while adaptively perceiving multiscale directional features, we propose DiPVNet, a framework constructed upon atomic dot-product operators. Our key insight lies in revealing the intrinsic property of the dot-product operation: it acts as a directional filter that simultaneously exhibits directional selectivity and rotation invariance. Based on this, DiPVNet avoids the need to explicitly model predefined directional features. Instead, as visualized in Figure 1, the directional selectivity of dot-product operators enables adaptive perception of local directions, while the hierarchical aggregation of global directional responses builds discriminative multiscale representations. The rotation invariance of the operator ensures robustness to arbitrary rotational transformations.

Specifically, our contributions are as follows:

- We identify and exploit the dual property of the dot-product operation—directional selectivity and rotation invariance—to design atomic dot-product operators that empower the network to perceive directional structures adaptively while maintaining rotational symmetry.
- We propose a Learnable Local Dot-Product (L2DP) operator, which performs differentiable dot-products between a center point and its neighbors, enabling adaptive learning of local directional features and improving the network's ability to handle non-uniform local structures.
- We propose the Direction-Aware Spherical Fourier Transform (DASFT). It uses dot-products to project point clouds onto spherical vectors, yielding a global directional response spectrum. This captures global directional features while reducing geometry misinterpretation risks from over-reliance on local features.

Extensive experiments under various challenging scenarios—including noise, large-angle rotations, and occlusions—demonstrate that DiPVNet outperforms existing rotation-robust methods across multiple benchmarks. Our method accurately identifies both discriminative global directional structures and critical local regions, significantly improving the robustness and accuracy of point cloud classification, segmentation, and other downstream 3D tasks.

2 Related Work

To address the challenge of rotational perturbations in 3D point cloud analysis, existing research has explored various strategies to enhance rotational robustness. These approaches can be broadly categorized into two main directions: explicit directional encoding and implicit modeling of rotational symmetry.

2.1 Explicit Directional Encoding

Explicit directional encoding focuses on designing specific mechanisms to directly capture directional information in point clouds. For example, common strategies (Xia et al. 2021; Wang et al. 2023) such as spatial orientation partitioning divide local space into fixed directional quadrants to extract orientation-aware features. ODFNet (Sahin, Mertan, and Unal 2022) models directionality by partitioning the spherical neighborhood into predefined directional cones and statistically analyzing the point distributions. Although these methods alleviate representation inconsistency caused by orientation variations through explicit directional modeling, their fixed spatial partitioning schemes are limited in adapting to the non-uniform distribution of point clouds.

2.2 Implicit Modeling of Rotational Symmetry

Implicit modeling of rotational symmetry aims to construct network architectures that are either equivariant or invariant to 3D rotations, so that the output features co-transform with (equivariance) or remain unaffected by (invariance) input rotations, thereby preserving directional information in deeper semantic spaces.

Invariant Representation Learning Earlier works relied on handcrafted descriptors (Chen and Cong 2022; Gu et al. 2021, 2022; Li et al. 2021b; Zhang, Yang, and Xiang 2024) or applied PCA-based (Li et al. 2021a; Yu et al. 2020; Xiao et al. 2019) alignment to eliminate orientation discrepancies. Modern approaches (Zhang et al. 2020a; Kim, Park, and Han 2020; Zhang, Yang, and Xiang 2024) often learn Local Reference Frames (LRFs) and model features within the predicted local coordinate systems to achieve invariance. However, their reliance on local features may lead to misinterpretations of the global geometric structure. Notably, the dot-product operator, due to its inherent rotational invariance, has been widely adopted to construct stable invariant information flows. For instance, SGMNet (Xu et al. 2021) uses dot-product operations to build local rotation-invariant features. However, its sorting mechanism disrupts the original spatial orientation relationships between point pairs, and it lacks a feature aggregation strategy to further enhance directional awareness.

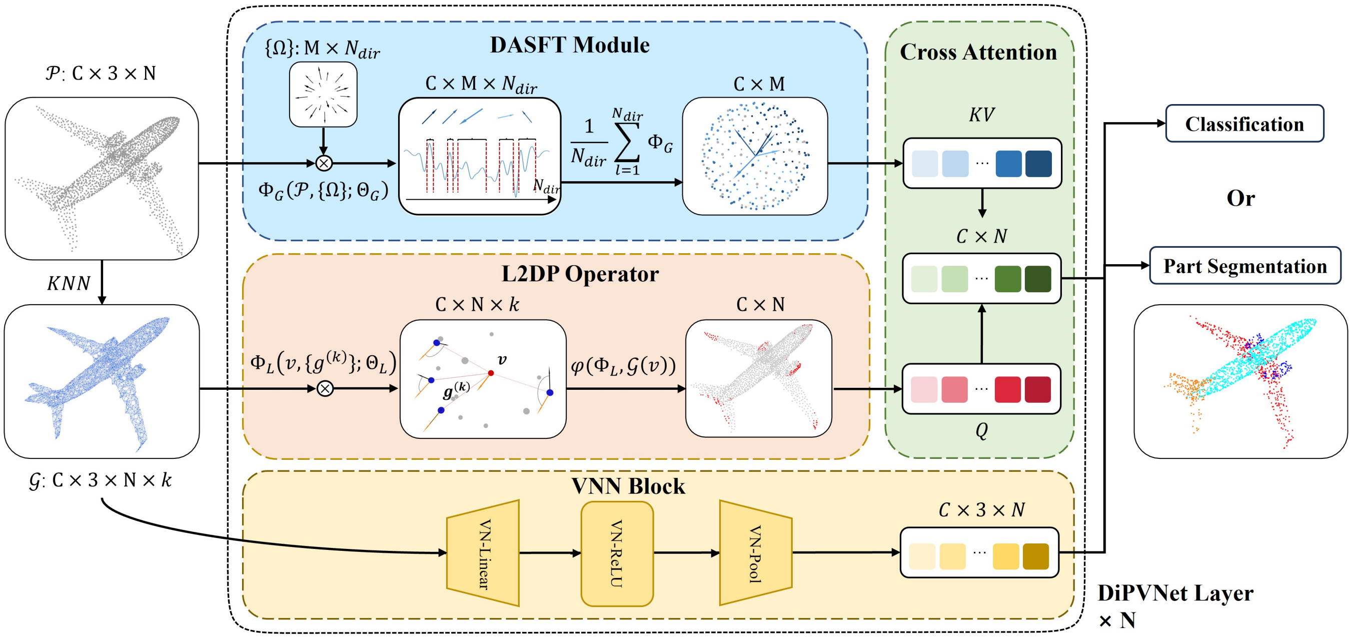

Figure 2: DiPVNet single-layer architecture. Point cloud features $\mathcal{P}$ are transformed into graph features $\mathcal{G}$ via KNN graph construction. In the DiPVNet layer, the VNN Block models rotation equivariance, while concurrently the L2DP operator processes graph features through Φ_L and extracts local directional features via aggregation mapping $\varphi(\cdot, G(v))$, and the DASFT module constructs a global directional response spectrum through the dot-product operator Φ_G between point clouds and spherical sampling vectors. Local and global features are fused via cross-attention mechanism, with the output discriminative directional features utilized for downstream tasks.

Equivariant Representation Learning Tensor Field Networks (TFN)(Thomas et al. 2018) leverage spherical harmonics to construct group-equivariant convolutional kernels, theoretically enabling the handling of arbitrary rotations. Nonetheless, the high-dimensional harmonic expansions introduce considerable computational overhead. Inspired by TFNs, Vector Neuron Networks (VNNs)(Deng et al. 2021) extend neural outputs from scalars to 3D vectors and employ vectorized operations to impose equivariant constraints implicitly at the network level. Subsequent developments (Satorras, Hoogeboom, and Welling 2021; Lin, Zhu, and Ghaffari 2023; Melnyk et al. 2024) have further strengthened the stable propagation of directional information into deeper semantic layers. However, these methods (Luo et al. 2022; Su et al. 2022; Jing et al. 2020) still struggle to efficiently exploit directional cues at the feature representation level. For example, VNN selects a single global direction vector to filter vector neurons, which fails to capture the intrinsic, fine-grained directional diversity inherent in point clouds. Other VNN-based works, such as VN-Transformer (Assaad et al. 2022), primarily utilize the dot-product to construct similarity matrices rather than to explicitly model directional information.

Recent studies have focused on joint modeling of rotational equivariance and invariance, typically through dual-branch architectures (Chen et al. 2024) or feature decoupling mechanisms (Zhang et al. 2023). These designs preserve directional information via the equivariant branch and extract rotation-robust features through the invariant branch, achieving stable representations under arbitrary rotations. However, current methods still face certain limitations: directional information is often indirectly recovered based on predefined geometric constraints, while the global distribution of orientations remains under-explored and under-modeled.

3 Method

In this section, we first present the Atomic Dot-Product Operator. Building on this, we propose DiPVNet for direction-aware point cloud representation. Further, we introduce two key components that leverage the operator to enhance multi-scale directional perception: the L2DP operator and the DASFT module.

3.1 Atomic Dot-Product Operator

We encapsulate the dot-product operation into a differentiable atomic operator:

$$\Phi\left(\mathbf{a}, \{\mathbf{b}_i\}; \boldsymbol{\Theta}\right) = \mathrm{FFN}\left(\{\langle \mathbf{a} \cdot \mathbf{b}_i \rangle\}_{i=1}^{K}; \boldsymbol{\Theta}\right), \quad (1)$$

where $\{\mathbf{b}_i\}_{i=1}^{K}$ denotes a set of associated vectors, and $\boldsymbol{\Theta}$ represents the learnable parameters of the FFN (Feed-Forward Network). This operator inherently combines directional awareness with rotational robustness, serving as the fundamental computational unit and core component for constructing DiPVNet.

3.2 DiPVNet

Building upon the atomic dot-product operator, we propose DiPVNet—a point cloud representation learning framework centered around dot-product operators. Figure 2 illustrates the DiPVNet architecture and its module designs. This framework hierarchically models cross-scale directional characteristics of point clouds while preserving rotational symmetry.

First, we propose the L2DP operator, which extends the atomic dot-product operator for local feature extraction:

$$\Phi_{\mathrm{L}}\left(\mathbf{v},\{\mathbf{g}^{(k)}\};\boldsymbol{\Theta}\right)=\mathrm{FFN}_{\mathrm{L}}\left(\langle\mathbf{v}\cdot\{\mathbf{g}^{(k)}\}\rangle;\boldsymbol{\Theta}_{\mathrm{L}}\right), \quad (2)$$

where $\{\mathbf{g}^{(k)}\}\in\mathcal{G}(\mathbf{v})$ denotes the set of k-nearest neighbors of the central point $\mathbf{v}$.

Finally, adaptive integration of directional information within the neighborhood is achieved through an invariant feature aggregation mapping $\phi(\cdot,\mathcal{G}(\mathbf{v}))$, designed with two implementations: Direct Linear Projection (DLP) and Statistic-Aware Projection (SAP), each tailored for distinct application scenarios:

$$f_{\mathrm{L2DP}}(\mathbf{v})=\phi\left[\Phi_{\mathrm{L}}\left(\mathbf{v},\{\mathbf{g}^{(k)}\};\boldsymbol{\Theta}_{\mathrm{L}}\right),\mathcal{G}(\mathbf{v})\right]. \quad (3)$$

This operator can adaptively perceive and learn local directional features, significantly enhancing the model's representation capability for non-uniform local structures.

Second, to model the global directional characteristics of point clouds, we compute the dot-product between the raw point cloud $\mathcal{P}$ and a set of sampled vectors $\{\boldsymbol{\Omega}_l\}_{l=1}^{N_{\mathrm{dir}}}$ on the sphere S^2. This operation is equivalent to the Fourier transform of the point cloud in the spherical frequency domain $\mathcal{F}(\mathcal{P},\{\boldsymbol{\Omega}_l\})$. We thus define the Global Dot-Product Operator:

$$\Phi_{\mathrm{G}}\left(\mathcal{P},\{\boldsymbol{\Omega}_l\};\boldsymbol{\Theta}\right)=\mathrm{FFN}_{\mathrm{G}}\left(E(\mathcal{P},\{\boldsymbol{\Omega}_l\});\boldsymbol{\Theta}_{\mathrm{G}}\right), \quad (4)$$

where

$$E(\mathcal{P},\{\boldsymbol{\Omega}_l\})=|\mathcal{F}(\mathcal{P},\{\boldsymbol{\Omega}_l\})|^2 \quad (5)$$

denotes the response power spectrum of point cloud P along the sampled directions $\{\boldsymbol{\Omega}_l\}_{l=1}^{N_{\mathrm{dir}}}$.

The rotation-invariant global descriptor is then obtained through spherical averaging:

$$f_{\mathrm{DASFT}}(\mathcal{P})=\frac{1}{N_{\mathrm{dir}}}\sum_{l=1}^{N_{\mathrm{dir}}}\Phi_{\mathrm{G}}\left(\mathcal{P},\boldsymbol{\Omega}_l;\boldsymbol{\Theta}_{\mathrm{G}}\right), \quad (6)$$

where N_{dir} is the number of spherical sampling directions.

This entire pipeline constitutes the DASFT module. By learning multiscale directional response spectra, it effectively models global directional characteristics of point clouds, mitigating potential misjudgments of overall geometric structures that may arise from relying solely on local features.

Finally, a cross-attention mechanism fuses the invariant features: L2DP serves as the query, while DASFT provides the key/value to adaptively guide local features. Concurrently, the equivariant features from the baseline VNN Block are projected onto a learned canonical basis to produce rotation-robust scalar tokens, which are then concatenated with the fused invariant features for downstream tasks.

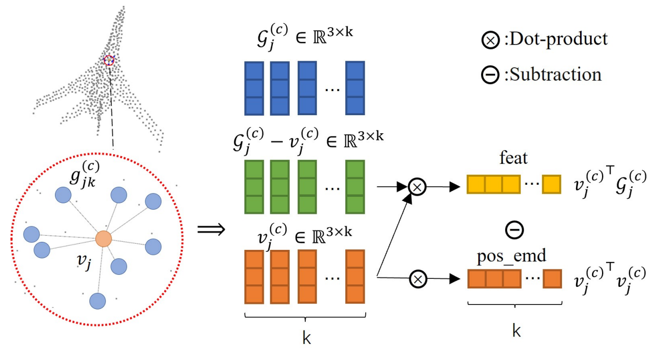

Figure 3: In the partial process of the L2DP operator acting on the j-th center point v_j, the center point feature is replicated k times and subtracted from neighboring features in the k-nearest neighbor neighborhood $\mathcal{G}_j$; a dot-product operation is performed to obtain directional information relative to the center point and its positional encoding; and the result is fed into FFN.

3.3 Learnable Local Dot-Product (L2DP)

Rotation-Invariant Local Representation Building upon the proven rotational invariance of the dot-product (proof in Appendix), we define the rotation-invariant local geometric representation for each point v_j as:

$$I_j^{(\mathcal{N})}=\{\langle v_j,\,v_k\rangle \mid v_k\in\mathcal{N}_j\}, \quad (7)$$

where $\mathcal{N}_j$ denotes the neighborhood set of point v_j comprising its K nearest neighbors. If two point clouds, $\mathcal{P}$ and $\mathcal{P}'$, are equivalent under the SO(3) group (i.e., $\exists\, R\in\mathrm{SO}(3)$ such that $\mathcal{P}'=R\cdot\mathcal{P}$), then their set of local invariants satisfies:

$$I_j^{(\mathcal{N})}=I_j^{(\mathcal{N})'} \quad \forall j. \quad (8)$$

This means that the local invariant set I_j fully characterizes the equivalence class of the point cloud under the action of the rotation group. However, the original definition $I_j^{(\mathcal{N})}$ only describes the degree of directional correlation between the center point v_j and its neighbor point v_k, but loses the characteristics of the center point itself. In order to enhance the geometric perception, the relative position coding is further introduced:

$$I_j^{(\mathcal{N},\mathrm{rel})}=\{\langle v_j,\,v_k-v_j\rangle \mid v_k\in\mathcal{N}_j\}. \quad (9)$$

The "rel" tag indicates the introduction of relative position encoding.

In practice, for each point v_j in the point cloud $\mathcal{P}=\{v_j\in\mathbb{R}^{c\times3}\}_{j=1}^{N}$, we can obtain its K nearest neighbor features to form the local group $\mathcal{G}_j=\{g_j^{(k)}\in\mathbb{R}^{c\times3}\}_{k=1}^{K}$. The aggregated local neighborhood is then given by $\mathcal{G}=\{\mathcal{G}_j\}_{j=1}^{N}$ (Yu, Zhang, and Shen 2024). Therefore, we can obtain the local dot-product invariant

$$I_j^{(\mathcal{G})}=\{\langle v_j,\,g_{jk}\rangle \mid k=1,\ldots,K\}. \quad (10)$$

In the L2DP module, we are actually dealing with:

$$I_j^{(\mathcal{G},\mathrm{rel})}=\{\langle v_j,\,g_{jk}-v_j\rangle \mid k=1,\ldots,K\}. \quad (11)$$

This means that the point product result, computed channel-wise over the 3D spatial dimensions, is $\langle v_j, g_{jk}\rangle - \langle v_j, v_j\rangle$, where $\langle v_j, g_{jk}\rangle$ contains local geometric information, and $v_j^{(c)\top} v_j^{(c)}$ can be interpreted as injecting position coding (Qiu et al. 2022), which rectifies the inherent disorder of the point cloud and improves the generalization ability of subsequent attention calculations. The process of obtaining local invariants is shown in Figure 3. We implement the feature extraction process through the dot-product operator:

$$\Phi_{\mathrm{L}}\left(I_j^{(\mathrm{G,rel})}; \mathbf{\Theta}_{\mathrm{L}}\right) = \mathrm{FFN}_{\mathrm{L}}\left(\left\langle \mathbf{v}_j, \mathbf{g}_j^{(k)} - \mathbf{v}_j\right\rangle; \mathbf{\Theta}_{\mathrm{L}}\right), \tag{12}$$

thereby adaptively enhancing neighborhood directional awareness while preserving rotation invariance, ultimately strengthening the model's local modeling capability for non-uniform structures.

Invariant Feature Aggregation To integrate local invariants into point-level features while preserving rotation invariance, we propose two learnable aggregation mappings $\phi(\cdot, \mathcal{G}(\mathbf{v}))$ with complementary characteristics.

The Direct Linear Projection (DLP) implements $\phi_{\mathrm{DLP}}(\cdot, \mathcal{G}(\mathbf{v})) : \mathbb{R}^{c\times K} \to \mathbb{R}^c$, which directly maps all K dot-product outputs to a high-dimensional semantic space:

$$f_{\mathrm{DLP}}(\mathbf{v}) = \mathrm{Layernorm}\left(\phi_{\mathrm{DLP}}\left(\Phi_{\mathrm{L}}\left(I^{(\mathrm{G,rel})}; \mathbf{\Theta}_{\mathrm{L}}\right)\right)\right). \tag{13}$$

This approach preserves full neighbor interaction information but incurs linearly growing computational complexity with K, making it optimal for small-scale neighborhoods.

Conversely, the Statistic-Aware Projection (SAP) employs feature compression through statistical pooling, which computes the maximum, variance, and average of the relative features, denoted as $I_j^{\max}$, I_j^{var}, and I_j^{avg}, respectively. This is followed by $\phi_{\mathrm{SAP}}(\cdot, \mathcal{G}(\mathbf{v})) : \mathbb{R}^{c\times 3} \to \mathbb{R}^c$:

$$f_{\mathrm{SAP}}(\mathbf{v}) = \mathrm{Dropout}\left(\phi_{\mathrm{SAP}}\left(\left[I_j^{\max} \| I_j^{\mathrm{var}} \| I_j^{\mathrm{avg}}\right]\right)\right). \tag{14}$$

SAP sacrifices geometric details for enhanced computational efficiency and noise robustness, better suited for large-scale neighborhoods requiring long-range modeling.

Operator Formalization Finally, we encapsulate the aforementioned process into the L2DP operator.

$$f_{\mathrm{L2DP}}(\mathbf{v}) = \phi\left[\Phi_{\mathrm{L}}\left(\mathbf{v}, \{\mathbf{g}^{(k)}\}; \mathbf{\Theta}_{\mathrm{L}}\right), \mathcal{G}(\mathbf{v})\right]. \tag{15}$$

This operator adaptively perceives and learns local directional characteristics while aggregating features into rotation-invariant point-level representations. The framework significantly enhances the model's capacity for representing non-uniform local structures through its direction-aware learning mechanism.

3.4 Direction-Aware Spherical Fourier Transform (DASFT)

To model the global directional characteristics of point clouds, we can regard the point cloud $\mathcal{P} = \{\mathbf{v}_j \in \mathbb{R}^3\}_{j=1}^N$ as a discrete signal distributed in 3D space. Formally, this can be represented as:

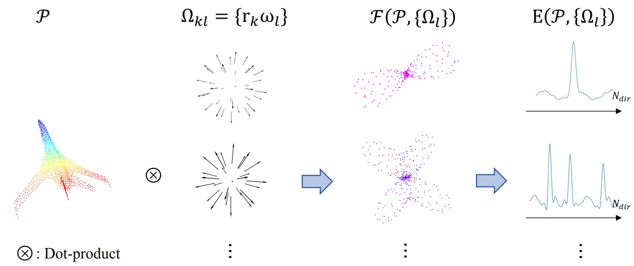

Figure 4: The point cloud $\mathcal{P}$ is projected via dot-products with spherical sampling unit vectors Ω of varying frequency amplitudes, yielding the spherical frequency-domain response $F(\mathcal{P}, \{\Omega\})$ of global features, subsequently constructing the directional response spectrum $E(\mathcal{P}, \{\Omega\})$ which characterizes the dominant directions of the point cloud's macroscopic structure.

$$\mathcal{P}(v) = \sum_{j=1}^{n} \delta(v - v_j), \quad \delta(v) = \begin{cases} +\infty & v = 0 \\ 0 & \text{otherwise} \end{cases}, \tag{16}$$

where $\delta(\cdot)$ is the Dirac delta function. This means the point cloud $\mathcal{P}$ can be seen as a superposition of pulsed signals located at $\{v_j \in \mathbb{R}^{c\times 3}\}_{j=1}^N$.

Under the framework of generalized functions (Kondor 2007), the Fourier transform of $\mathcal{P}(v)$ is:

$$\mathcal{F}(\mathcal{P}, \{\mathbf{\Omega}\}) = \int \mathcal{P}(v) \exp(-i\Omega^\top v) dv = \sum_{j=1}^{n} \exp(-i\Omega^\top v_j), \tag{17}$$

where $\Omega = r\omega$ is a frequency space vector (r: frequency modulus length, $\omega \in S^2$: unit direction vector). The dot-product between $\mathcal{P}$ and Ω corresponds to the phase term:

$$\mathcal{F}(\mathcal{P}, \{\mathbf{\Omega}\}) = \sum_{j=1}^{n} \exp(-ir\omega^\top v_j). \tag{18}$$

This indicates each point v_j contributes a phase factor $e^{-ir\omega^\top v_j}$ to frequency $r \cdot \omega$, and the spherical spectrum $\mathcal{F}(r \cdot \omega)$ contains complete global information (Xu et al. 2022; Son et al. 2024). Further, we define the energy spectrum as the modulus square of the Fourier coefficient

$$E(\mathcal{P}, \{\mathbf{\Omega}\}) = |\mathcal{F}(\mathcal{P}, \{\mathbf{\Omega}\})|^2. \tag{19}$$

This equation characterizes the structural energy distribution of the point cloud across different frequency scales r and directions $\boldsymbol{\omega}$. Figure 4 visualizes the directional response results, reflecting the multiscale global directional characteristics of the point cloud.

Based on this analysis, we construct the Global Dot-Product Operator:

$$\Phi_{\mathrm{G}}(\mathcal{P}, \{\mathbf{\Omega}\}; \mathbf{\Theta}_{\mathrm{G}}) = \mathrm{FFN}_{\mathrm{G}}(E(\mathcal{P}, \{\mathbf{\Omega}\}); \mathbf{\Theta}_{\mathrm{G}}), \tag{20}$$

where $\Omega = r \cdot \omega$. This operator adaptively fuses energy channels from different directions through learnable channel mapping.

To construct a rotation-invariant descriptor, we uniformly sample N_{dir} directions $\{\omega_l\}_{l=1}^{N_{\text{dir}}}$ on the unit sphere S^2 and perform energy integration:

$$G(r) = \frac{1}{4\pi} \int_{S^2} E(r \cdot \omega) d\omega \approx \frac{1}{N_{\text{dir}}} \sum_{l=1}^{N_{\text{dir}}} E(r \cdot \omega_l) \,. \quad (21)$$

Since $G(r)$ only depends on the modulus r, it is rotation-invariant. While its computation (Eq. 21) is a discrete approximation, we empirically verified its robustness to the sampling density N_{dir} (≥ 36) in the Appendix.

Finally, the DASFT module can be formally expressed as:

$$f_{\text{DASFT}}(\mathcal{P}) = \frac{1}{N_{\text{dir}}} \sum_{l=1}^{N_{\text{dir}}} \Phi_{\text{G}}\left(P, \boldsymbol{\omega}_l; \boldsymbol{\Theta}_{\text{G}}\right) \,. \quad (22)$$

Through hierarchical learning of multiscale directional response spectra, this module models the global directional characteristics of point clouds, effectively mitigating potential misjudgments of overall geometric structures that may occur when relying solely on local features (Zhang et al. 2020b).

4 Experiments

4.1 Datasets and Tasks

Shape Classification We tested the classification performance of our model on the synthetic dataset ModelNet40 (Wu et al. 2015) and the real-world dataset ScanObjectNN (Uy et al. 2019). ModelNet40 consists of 40 categories and a total of 12311 CAD mesh models, of which 9843 are used for training and 2468 are used for testing. ScanObjectNN contains 15000 complete objects scanned from 2902 real-world objects, and we use the OBJ_BG subset from them.

Part Segmentation For the segmentation task, we evaluated the segmentation performance of the network on the ShapeNetPart dataset (Chang et al. 2015), which contains 16880 synthetic samples divided into 14006 training samples and 2874 training samples. Including 16 object categories, each category has 2 to 6 sections of annotations, totaling 50 subdivisions.

Task Setup We follow the general practice of using universal training/testing rotation settings: z/z, z/SO(3), SO(3)/SO(3), Among them, z represents random rotation around the z-axis, and SO(3) represents three-dimensional random rotation at any angle. These rotations are generated and applied to the input point clouds.

4.2 Experimental Setup

In terms of point cloud sampling, we use farthest point sampling (FPS). In the classification task, we uniformly sample 1024 points, and in the segmentation task, we sample 2048 points. We implemented our model in PyTorch using the official implementation (Deng et al. 2021; Li et al. 2021a; Melnyk et al. 2024), while using VN-DGCNN as the baseline network and employing the same data augmentation method: random translation within the range of [-0.2, 0.2] and random scaling within the range of [2/3, 3/2] during the training process. In the classification task, we employ DLP aggregation for ModelNet40 and SAP aggregation for ScanObjectNN; in the segmentation task, we utilize SAP aggregation. When calculating the DASFT, we linearly sample the frequency within the range of [0,12] and select $N_{dir} = 36$ in the directional sampling. Consistent with the baseline, we use SGD with an initial learning rate of 0.1 and momentum equal to 0.9, as well as a cosine annealing strategy for gradual reduction with a learning rate of 0.001 and minimizing the cross-entropy loss with label smoothing. For the selection of epochs, we trained 200 epochs on ModelNet40 and 250 epochs on ShapeNet, with a batch size of 32.

Methods	**z/z**	**z/SO(3)**	**SO(3)/SO(3)**
	Rotation-sensitive		
PointNet	85.9	19.6	74.7
DGCNN	**92.2**	20.6	81.1
	Rotation-robust		
TFN	89.7	89.7	-
VN-DGCNN	89.5	89.5	90.2
Pose Selector	90.2	90.2	90.2
LGR-Net	90.9	90.9	<u>91.1</u>
OrientedMP	88.4	88.4	88.9
VN-Transformer	-	-	90.8
PaRot	90.9	<u>91.0</u>	90.8
TetraSphere	90.5	90.5	-
Ours	<u>91.4</u>	**91.4**	**91.4**

Table 1: The results of each method on ModelNet40 were compared under different training/testing conditions. The overall optimal results are shown in bold and the sub-optimal items are underlined.

4.3 Results

We categorize existing methods into rotation sensitive and rotation robust classes.

Shape Classification Tables 1 and 2 present the comparison results between our method and existing methods in classification tasks. The results demonstrate that our method maintains consistent performance across three training/testing sessions. The results under the z/SO(3) setting best reflect the model's generalization performance. For rotation-sensitive methods (Li et al. 2018; Wang et al. 2019), although their performance on z/z is decent, they fail to generalize to unknown orientations under the z/SO(3) setting, resulting in a sharp performance decline. Compared with PCR-cored framework (Yu, Zhang, and Cai 2023), Pose Selector (Li et al. 2021a) and other rotational equivariant architectures (Zhao et al. 2022; Fuchs et al. 2020), our method achieves in-depth exploration of multi-scale directional characteristics in point clouds by introducing dot-product operators into each rotational equivariant layer,

Methods	z/z	z/SO(3)	SO(3)/SO(3)
Rotation-sensitive			
PointCNN	86.1	14.6	63.7
DGCNN	82.8	17.7	71.8
Rotation-robust			
PaRINet + PCA	83.3	83.3	83.3
PCR-cored framework	-	86.6	86.3
VN-DGCNN	83.5	83.5	84.2
Pose Selector	84.3	84.3	84.3
LGR-Net	81.2	81.2	81.4
TetraSphere	<u>87.3</u>	<u>87.3</u>	<u>87.1</u>
Ours	**87.5**	**87.5**	**87.4**

Table 2: Under different training/testing conditions, the results of each method on ScanObjectNN are compared. The overall optimal results are shown in bold and the sub-optimal items are underlined.

while simultaneously modeling rotational symmetry.

Part Segmentation Table 3 shows the comparison results between our method and existing methods in segmentation tasks. We evaluated the model performance using the Mean Intersection over Union (mIoU) on each instance, and the results showed that our model achieved the best performance on the instance. Figure 5 illustrates the segmentation results through visualizations.

Methods	z/SO(3)	SO(3)/SO(3)
Rotation-sensitive		
PointNet	37.8	74.4
DGCNN	37.4	73.3
Rotation-robust		
PCR-cored framework	80.3	80.4
VN-DGCNN	81.4	81.4
Pose Selector	81.7	<u>81.7</u>
LGR-Net	80.0	80.1
OrientedMP	80.1	80.9
TetraSphere	<u>82.3</u>	-
Ours	**82.5**	**82.7**

Table 3: The results of each method are compared on ShapeNetPart dataset under different training/testing conditions, and the evaluation index is the average intersection union ratio of all instances. The overall optimal results are shown in bold and the sub-optimal items are underlined.

4.4 Ablation Studies

As shown in the Table 4, Model A incorporates only the DASFT module, and its performance shows no significant improvement over the baseline, indicating that global direction-aware features alone are insufficient to boost performance. Models B and C compare SAP and DLP aggregation strategies under the condition of removing the DASFT module. On the ModelNet40 dataset, which contains relatively low noise interference, SAP aggregation underperforms compared to DLP, confirming the advantage of DLP in preserving local geometric consistency. However, the results of Model C also demonstrate that relying solely on local features can lead to misinterpretation of global structures, resulting in lower performance compared to the optimal model. Model D replaces Cross Attention with a gating mechanism for feature fusion, and the results indicate that static weight allocation fails to achieve the dynamic feature calibration characteristic of attention mechanisms, leading to suboptimal performance.

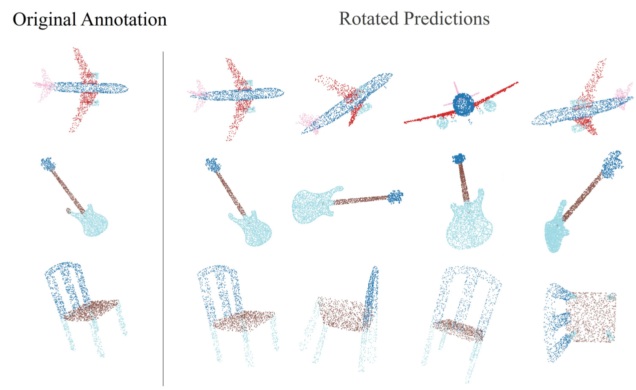

Figure 5: Visualization of segmentation results.

Model	L2DP		DASFT	Feature Fusion		Acc.
	DLP	SAP		Gate	CA	
Baseline						89.5
A			✓			89.5
B		✓				89.9
C	✓					90.6
D	✓		✓	✓		90.9
E (Ours)	✓		✓		✓	91.4

Table 4: Ablation results assessing component impacts on ModelNet40 under z/SO(3) conditions: L2DP aggregation (DLP or SAP alternatives), DASFT inclusion, and fusion method (CA or Gate).

5 Conclusion

This paper proposes an efficient direction-aware point cloud rotation-robust framework based on atomic dot-product operators, revealing a new design paradigm that goes beyond module-level combinations from a novel atomic dot-product perspective. The network architecture inspired by this enhances direction-aware capability while modeling rotational symmetry, paving the way for more robust point cloud analysis under arbitrary rotations. Future work will explore the applicability of atomic dot-product operators in other mainstream network architectures, thereby verifying their universal value for performance improvement.

References

Assaad, S.; Downey, C.; Al-Rfou, R.; Nayakanti, N.; and Sapp, B. 2022. Vn-transformer: Rotation-equivariant attention for vector neurons. *arXiv preprint arXiv:2206.04176*.

Chang, A. X.; Funkhouser, T.; Guibas, L.; Hanrahan, P.; Huang, Q.; Li, Z.; Savarese, S.; Savva, M.; Song, S.; Su, H.; et al. 2015. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*.

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.

Chou, Y.-C.; Lin, Y.-P.; Yeh, Y.-M.; and Lu, Y.-C. 2021. 3d-gfe: a three-dimensional geometric-feature extractor for point cloud data. In *2021 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2013–2017. IEEE.

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.

Dym, N.; and Maron, H. 2020. On the universality of rotation equivariant point cloud networks. *arXiv preprint arXiv:2010.02449*.

Feng, T.; Wang, W.; Ma, F.; and Yang, Y. 2024. Lsk3dnet: Towards effective and efficient 3d perception with large sparse kernels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 14916–14927.

Feng, T.; Wang, W.; Wang, X.; Yang, Y.; and Zheng, Q. 2023. Clustering based point cloud representation learning for 3d analysis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 8283–8294.

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.

Gu, R.; Wu, Q.; Xu, H.; Ng, W. W.; and Wang, Z. 2021. Learning efficient rotation representation for point cloud via local-global aggregation. In *2021 IEEE International conference on multimedia and expo (ICME)*, 1–6. IEEE Computer Society.

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.

Jia, S.; Gong, X.; Liu, F.; and Ma, L. 2025. AI-Powered LiDAR Point Cloud Understanding and Processing: An Updated Survey. *IEEE Transactions on Intelligent Transportation Systems*.

Jiang, M.; Wu, Y.; Zhao, T.; Zhao, Z.; and Lu, C. 2018. Pointsift: A sift-like network module for 3d point cloud semantic segmentation. *arXiv preprint arXiv:1807.00652*.

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.

Kondor, R. 2007. A novel set of rotationally and translationally invariant features for images based on the non-commutative bispectrum. *arXiv preprint cs/0701127*.

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.

Lin, T.-Y.; Zhu, M.; and Ghaffari, M. 2023. Lie neurons: Adjoint-equivariant neural networks for semisimple lie algebras. *arXiv preprint arXiv:2310.04521*.

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.

Qiu, Z.; Li, Y.; Wang, Y.; Pan, Y.; Yao, T.; and Mei, T. 2022. Spe-net: Boosting point cloud analysis via rotation robustness enhancement. In *European Conference on Computer Vision*, 593–609. Springer.

Sahin, Y. H.; Mertan, A.; and Unal, G. 2022. ODFNet: Using orientation distribution functions to characterize 3D point clouds. *Computers & Graphics*, 102: 610–618.

Satorras, V. G.; Hoogeboom, E.; and Welling, M. 2021. E (n) equivariant graph neural networks. In *International conference on machine learning*, 9323–9332. PMLR.

Schütt, K.; Unke, O.; and Gastegger, M. 2021. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *International conference on machine learning*, 9377–9388. PMLR.

Son, D.; Kim, J.; Son, S.; and Kim, B. 2024. An intuitive multi-frequency feature representation for SO (3)-equivariant networks. *arXiv preprint arXiv:2405.04537*.

Su, Z.; Welling, M.; Pietikäinen, M.; and Liu, L. 2022. Svnet: Where so (3) equivariance meets binarization on point cloud representation. In *2022 International Conference on 3D Vision (3DV)*, 547–556. IEEE.

Thomas, N.; Smidt, T.; Kearnes, S.; Yang, L.; Li, L.; Kohlhoff, K.; and Riley, P. 2018. Tensor field networks: Rotation-and translation-equivariant neural networks for 3d point clouds. *arXiv preprint arXiv:1802.08219*.

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, S.; Liu, Y.; Wang, L.; Sun, Y.; and Yin, B. 2023. PASIFTNet: Scale-and-directional-aware semantic segmentation of point clouds. *Computer-Aided Design*, 156: 103462.

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, 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.

Xia, Y.; Xu, Y.; Li, S.; Wang, R.; Du, J.; Cremers, D.; and Stilla, U. 2021. SOE-Net: A self-attention and orientation encoding network for point cloud based place recognition. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, 11348–11357.

Xiang, T.; Zhang, C.; Song, Y.; Yu, J.; and Cai, W. 2021. Walk in the cloud: Learning curves for point clouds shape analysis. In *Proceedings of the IEEE/CVF international conference on computer vision*, 915–924.

Xiao, Z.; Lin, H.; Li, R.; Chao, H.; and Ding, S. 2019. Endowing deep 3d models with rotation invariance based on principal component analysis. *arXiv preprint arXiv:1910.08901*.

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.

Xu, Y.; Lei, J.; Dobriban, E.; and Daniilidis, K. 2022. Unified fourier-based kernel and nonlinearity design for equivariant networks on homogeneous spaces. In *International Conference on Machine Learning*, 24596–24614. PMLR.

Yu, C.; Zhang, S.; and Shen, L.-Y. 2024. Equivariant Analysis of Point Clouds by Message Passing on Simplicial Complexes. In *Proceedings of the 2024 7th International Conference on Image and Graphics Processing*, 286–293.

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.

Yu, R.; Wei, X.; Tombari, F.; and Sun, J. 2020. Deep positional and relational feature learning for rotation-invariant point cloud analysis. In *European Conference on Computer Vision*, 217–233. Springer.

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, J.; Yu, M.-Y.; Vasudevan, R.; and Johnson-Roberson, M. 2020a. Learning rotation-invariant representations of point clouds using aligned edge convolutional neural networks. In *2020 International conference on 3D Vision (3DV)*, 200–209. IEEE.

Zhang, Z.; Hua, B.-S.; Chen, W.; Tian, Y.; and Yeung, S.-K. 2020b. Global context aware convolutions for 3d point cloud understanding. In *2020 international conference on 3D vision (3DV)*, 210–219. IEEE.

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.