



ODFNet: Using orientation distribution functions to characterize 3D point clouds

Yusuf H. Sahin^{a,*}, Alican Mertan^a, Gozde Unal^a

^aIstanbul Technical University, Computer Engineering, Istanbul, 34469, Turkey

ABSTRACT

Learning new representations of 3D point clouds is an active research area in 3D vision, as the order-invariant point cloud structure still presents challenges for the design of neural network architectures. Recent work explored learning global, local, or multi-scale features for point clouds. However, none of the earlier methods focused on capturing contextual shape information by analyzing local orientation distributions of points. In this paper, we use point orientation distributions around a point in order to obtain an expressive local neighborhood representation for point clouds. We achieve this by dividing the spherical neighborhood of a given point into predefined cone volumes, and statistics inside each volume are used as point features. In this way, a local patch can be represented not only by the selected point's nearest neighbors, but also by considering a point density distribution defined along multiple orientations around the point. We are then able to construct an orientation distribution function (ODF) neural network that makes use of an ODFBlock which relies on MLP (multi-layer perceptron) layers. The new ODFNet model achieves state-of-the-art accuracy for object classification on ModelNet40 and ScanObjectNN datasets, and segmentation on ShapeNet and S3DIS datasets.

1. Introduction

Convolutional neural networks (CNNs) are widely used in vision and pattern recognition problems like object classification, object recognition, and segmentation [1]. However, CNNs have not been applicable to point clouds until recent years. The main obstacle to this was the problem of how to interpret a point in a point cloud representation, which has a permutation-invariant structure, a property not possessed by pixels or voxels in 2D or 3D images. On a 2D or 3D image grid, the convolution operation is defined as a weighted sum in a local neighborhood, which is defined by the kernel. However, in a point cloud, a similar neighborhood structure among the points does not exist. PointNet [2] pioneered the way to utilizing neural network models for the point cloud classification problem by aiming at

constructing a global feature transformation on all the points in the point cloud while respecting their order invariance. While PointNet did not use any neighborhood information, it is argued and shown that using local features improves the performance in recent works [3, 4, 5].

Recent approaches to extracting local features from point clouds include creating spherical volumes [3] or choosing k -nearest neighbors [7] and collecting local information of each point in those defined neighborhoods. The latest studies DensePoint [8] and ShellNet [9], that obtained state-of-the-art results for classification on the ModelNet40 classification benchmark [6], create spherical regions around each point. DensePoint employs spheres of different sizes in each layer of the neural network to obtain features from multiple scales; whereas in ShellNet, coordinates of points in each shell are transformed via an MLP (multi-layer perceptron), and a max-pooling operation aggregates the features across shells. In both methods, the points in the sphere are handled in such a way as to ignore their orientations with respect to the selected point. In [10], a spherical

*Corresponding author: Yusuf H. Sahin
 e-mail: sahinyu@itu.edu.tr (Yusuf H. Sahin), mertana@itu.edu.tr (Alican Mertan), gozde.unal@itu.edu.tr (Gozde Unal)

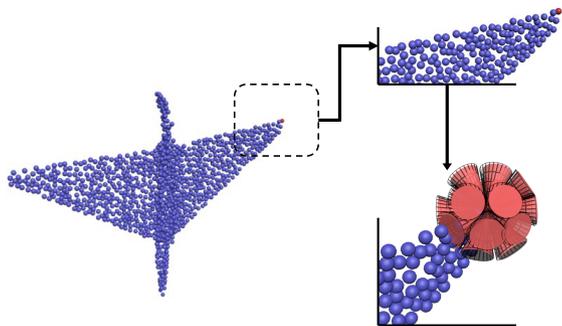


Fig. 1. A plane object from the ModelNet40 [6] dataset. For each point in a point cloud, ODFNet calculates the distribution of nearby points inside a spherical region. To do this, it utilizes cones with predefined orientations that spanning the spherical sectors. These cones can be of different scales and radii.

convolution is presented alongside octree partitioning where the sphere is divided into bins, and bin features are obtained by averaging the features of the points inside the bin, without using any point density information.

In order to increase the representative power of local features, we employ the distribution of orientations of points in a neighborhood with respect to a reference point. This leads to a new representation named point Orientation Distribution Functions (ODFs) for point clouds. ODFs can be computed by dividing each sphere around a point into a set of cones along predefined orientations, and calculating the density of points in each cone, as depicted in Figure 1. To increase the representative power of the feature, overlapping cones are also used. Some example ODFs are given in Figure 2. It can be observed that the ODF at the tip of the gun object, the ODFs on the corners of the plane or the table, and on the surface of the car, clearly capture the relative orientation of points with respect to the given center point. As such, we utilize the ODFs with their enhanced capability to compactly summarize the local neighborhood structure of a point cloud to our advantage in our point cloud analysis network model design.

The main contributions of our work can be summarized as such:

- The point ODFs, which incorporate the directional information inside a spherical neighborhood, are defined.
- A dedicated neural network architecture, the ODFNet, for classification and segmentation of point clouds, is presented.
- For different rotation-invariance scenarios, two different ways of utilizing ODFs for a point are presented. Both representations, which are either fully rotation-invariant or only rotation-invariant in the x-y plane, can be used depending on the alignment conditions of the environment.
- The ODFNet architecture is tested on popular benchmarks, and state-of-the-art (SoTA) accuracy scores on Model-

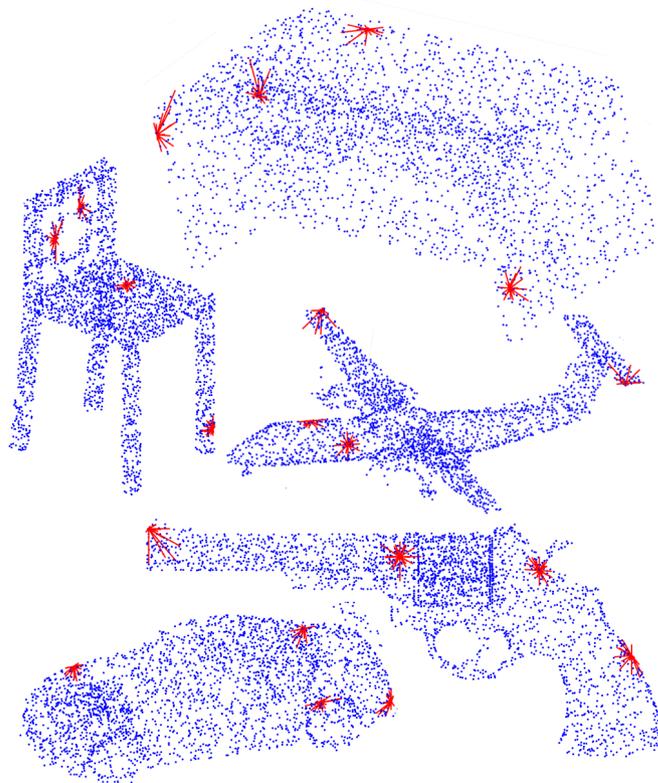


Fig. 2. ODFs for some selected points on example point clouds from the ShapeNet [11] dataset. Line length indicates strength.

Net40, Shapenet [11], S3DIS, [12] and ScanObjectNN [13] are obtained.

2. Related Works

Earlier studies like [14, 15, 16, 6] focusing on the classification of 3D objects prefer to voxelize the objects and use the voxelized occupancy map as an input to a neural network. However, this approach is not efficient for two reasons: First, the voxelization quality is highly related to the selected grid spacing and, as grid spacing dimensions get lower, distortion of a voxelized 3D object increases. Although high-resolution voxel grids are desirable, they are impractical due to computational constraints. Considering that the input to a voxel grid network is a 3D matrix, the network will consume significantly more storage and computation power when compared to 2D networks. A second drawback is that this is a very sparse representation, hence a superfluous amount of data is unnecessarily processed. [17] constructed a new representation to decrease the data amount by processing voxels, but the deformation problem remains.

Another approach is obtaining 2D views or depth maps and using them as inputs to a neural network [18, 19, 20, 21]. Feng et al. [22] construct hypergraphs extracted from view-based networks. However, view-based approaches are not deemed favorable as well since complete object or scene information is not utilized.

In order to make use of the standard grid convolution operation from the Euclidean CNN domain, Hua et al. project a grid onto every point where filter kernels are placed, and features are calculated over those grids via the convolution operation [23]. Li et al. [24] presented an architecture that learns a transformation matrix that weighs and permutes the points to be used in grid convolution.

PointNet [2] is considered as the first attempt to use point clouds as raw inputs to a neural network. As it might be expected, the main difficulty of using directly the points instead of view renders or voxelized 3D maps comes from the set representation since all permutations on a point set describes the same entry. Hence in PointNet, to classify a point cloud, all points are processed in multi-layer perceptrons in parallel (i.e. as shared weights for all points) to obtain point features and a symmetric function (e.g. a max-pooling operation) is used over these features to obtain an aggregated global feature. Although studies like [23] showed that ordered points can be used without a symmetric function, the symmetric function notion is widely used [4, 25]. Other studies using PointNet as a backbone network or in the middle steps include [26, 4].

PointNet++ [3] focuses on the fact that the original PointNet loses local features since all points are treated independently until the max-pool step. Thus, they hierarchically sample and group the point cloud and implement mini PointNets for each group. In [7], DGCNN, which handles the point cloud as a graph and uses a k-nearest neighbor approach to construct the connections, is introduced. Then, an edge convolution operation to perform a convolution centered on the selected point according to its nearest neighbors is defined. In SpiderCNN [5], a new convolution operation that benefits from Taylor series expansion is presented. In KPConv [27], kernel points with learnable weights are defined inside a local neighborhood and a linear correlation between a kernel point position and a neighbor point position is calculated and multiplied by these weights. In DensePoint[8] and ShellNet [9], a dedicated convolution is defined relying on statistics inside local neighborhoods, particularly spherical regions. Lei et al. [10] design a procedure where spherical regions are divided into bins and point features for each bin are collected according to bin weights. Then mean of the point features of each bin is used as bin features. In ConvPoint [28], convolution operation differentiates for spatial and feature operations where the spatial operations are done on randomly selected parts. For further reading on point clouds, a detailed survey on this topic can be investigated [29].

The works we examine so far are not invariant to rotation changes. Recently, developing rotation-invariant point features started to attract more attention in the point cloud community. In [30], the points are mapped on an icosahedral lattice and a new convolution operation to perform on this structure is defined. In [31], a rotation-invariant convolution, the RICConv operator is presented to obtain features from distances and angles in a rotation-invariant manner. Kim et al. presented RI-GCN which uses graph convolutions hierarchically [32]. In CG-Conv [33], local reference frames are created for each point neighborhood to obtain the local features and the global features are calculated via anchors.

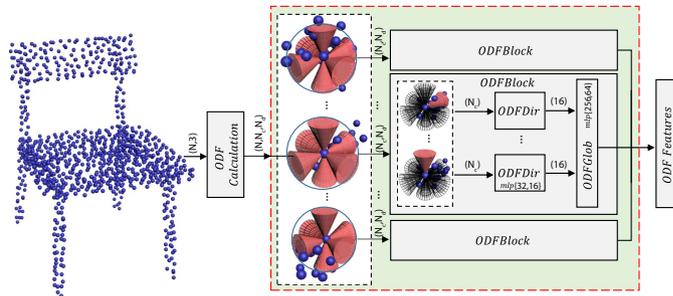


Fig. 3. The ODFBlock and its usage to obtain ODF features. Here N_c and N_d represent the number of different cones and directions, respectively. For every point, ODF values are computed along different cones. Then, ODF values are processed in ODFBlocks. An ODFBlock consists of two mpls: ODFDir and ODFGlob.

In the point cloud processing literature, the lack of any orientation-specific local feature representation motivated us to propose the ODFNet in this work. The inspiration for ODFs comes from the orientational probability distribution functions in the Diffusion MRI field [34, 35] that characterize the water diffusion in the brain. Those ODFs model the heterogeneous local tissue micro-structure in order to extract underlying multiple axonal fiber populations. For point clouds, the indirect analogy relies on constructing orientation distributions of the local point cloud mass that can reveal and help resolve the local geometry of the 3D shape along several directions. This is the main motivation in proposing ODFs for characterizing local structure in point clouds. Theoretically, for ODF estimation, one could use mathematical techniques such as spherical harmonics decomposition [36, 37], or for instance fitting von Mises-Fisher distributions [38], in which the latter involves constructing a multivariate normal distribution on the sphere. Alternatively, one could take a purely discrete approach to estimate an ODF through a histogram computation, by binning the local sphere around a given point into conic volumetric sections, as we perform in this work. We present numerical evidence that shows adopting ODFs in point clouds provides mostly competitive and in some cases superior performance among existing point cloud representations in problems of classification and segmentation of point clouds. This supports our conjecture that the inclusion of directional statistics of the local point cloud density leads to an improved localized structural representation and hence provides a performance upgrade.

3. Method

Our approach relies upon Orientation Distribution Functions due to their capability to express directional properties of local point cloud structure. In this section, we first describe the ODFs, the dedicated ODFBlock which is depicted in Figure 3, and then we present the ODFNet.

3.1. ODF Representation for Point Clouds

ODFs that we propose in this work rely on the number of points in a local conic neighborhood at multiple orientations.

Furthermore, to capture the local point density information in a hierarchy of scales, we utilize multi-scale cones with different apex angles and heights. There are three important components in defining ODFs; namely defining cones, their alignment given a point, and calculation of ODF values.

Defining cones: We divide the local sphere around a given point into conical volumes. The motivation for the usage of cones to parcellate the sphere comes from ODF-based methods for the medical imaging field [39]. Although there are some studies dividing the sphere using cylinders [35], a more efficient parcellation in terms of conical volumes is preferred in this work. Each of these cones can be characterized by their direction v_l , apex angle α_k , and length d_n as illustrated in Figure 4. We define 42 conic neighborhood directions v_l for each point, where directions are obtained after the first tessellation of an icosahedron. This is selected empirically by observing that further increasing the amount of tessellation causes a high rate of intersection between the cones and reducing it decreases the representation power.

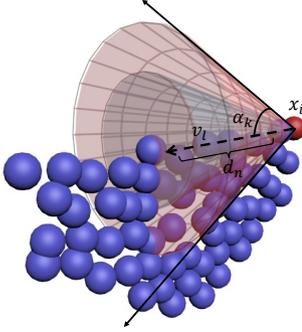


Fig. 4. An example ODF cone is placed on x_i along direction v_l . Cones at multiple scales (of heights and apex angles) are used to capture point density features at a hierarchy of neighborhoods.

For each neighborhood direction v_l , we utilize multiple cones to capture features from a hierarchy of neighborhoods. To this end, we use 2 different apex angles α_k , 31.71 degrees, which is the smallest angle that covers the whole sphere, and 60 degrees creating some intersection between the cones; and 4 different distances d_n , where d_n is the distance of the n^{th} neighbor of the given point, and n is selected from the collection of [8, 16, 24, 32]. Thus, for each point, 336 different cones are obtained (42 cone directions v_l , 4 distances d_n , and 2 apex angles α_k).

Aligning neighborhoods for a given point: In order to make the proposed ODF representation more adaptive to orientation changes of the objects, in the calculation of the ODFs, pivot directions need to be selected. Aligning cone directions v_l according to pivot directions calculated for each point as in Figure 5 effectively makes the ODF representation robust to orientation changes of the objects.

Particularly, two orthogonal directions have to be specified to achieve rotation-invariant ODF representations. We devise two methods for selecting these directions, namely RI-XY and RI-XYZ.

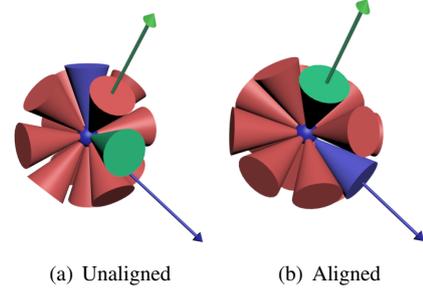


Fig. 5. An example alignment of the neighborhoods for a given point. The same rotation that aligns the two neighborhood directions, directions of the blue and the green cone, with the two pivot directions, the blue and green arrows, is applied to all neighborhood directions.

RI-XY achieves rotation-invariance in the x-y plane. Thus, it can be used when the object is aligned according to the z-axis. For RI-XY, for every point, projections of the selected point's 32 nearest neighbors on the x-y plane are calculated and the densest direction is selected as the pivot. Using the pivot direction and the z-axis, the ODF directions are aligned. In Figure 7, some pivot directions are shown for the RI-XY method. In addition to this alignment's contribution to obtain a rotation-invariant representation for rotations in the x-y plane (Figure 7.a), it also leads to symmetric representations for symmetric points (Figure 7.b). Unless otherwise stated, the RI-XY method is used for the experiments in this paper.

RI-XYZ allows us to define fully rotation-invariant representations. For RI-XYZ, the direction from the selected point x to the object center c_{object} is chosen as the first pivot direction. The second pivot direction is the cross product of the first pivot and the vector from point x to the center of the 32 nearest neighbors c_{local} as depicted in Fig 6. This pivot selection technique can be applied when no information about the object's alignment is available. Thus, RI-XYZ could be used for scenarios with totally unaligned objects. However a decrease in performance is expected since the relative coordinates of the points are vastly changed.

Calculation of ODF values: To compute the ODF value at a point $x_i \in S$, where S denotes the set of all points, and for a specific cone with an apex angle $2\alpha_k$, height d_n , and center direction vector v_l ,

$$ODF(x_i, \alpha_k, d_n, v_l) = \sum_{x_j \in S, i \neq j} \mathbb{1}(\|x_i - x_j\|_2 < d_n) \cdot \mathbb{1}(\text{acos}\left(\frac{(x_i - x_j) \cdot v_l}{\|x_i - x_j\| \|v_l\|}\right) < \alpha_k) \quad (1)$$

which gives us the point count inside the selected cone. Here, $\mathbb{1}(\cdot)$ refers to the indicator function. Since the cone heights d_n are selected according to the n^{th} -neighbor distance, this value is then normalized by n . In our experiments on a single NVIDIA Titan RTX graphics card, calculation of the cone values takes under ~ 76 msecs for a point cloud of 1024 points despite its complex information. It is also advantageous that the representation is calculated only once in the network.

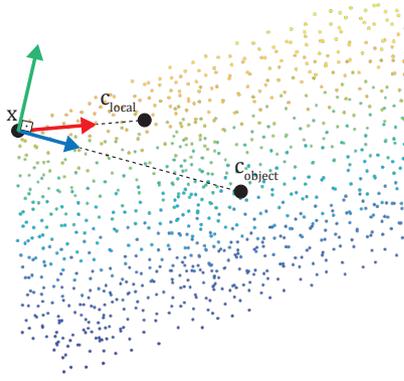


Fig. 6. An example of rotation-invariant pivot calculation in RI-XYZ. The blue arrow shows the direction to the object center which is the first pivot, the red arrow shows the direction to the center of the neighbors, and the green arrow is the second pivot direction which is the cross product of the other two vectors.

3.2. ODFBlock

In a point cloud, those differently sized and oriented cones that we construct help capture local variations in terms of point density, and encode them into the ODF features, which is provided into the dedicated neural network block in Figure 3. After calculating the ODFs at each point, we define the ODFBlock which operates on the ODFs as follows:

$$ODFBlock(x_i, \theta_d, \theta_g) = ODFGlob(\theta_d, ODFDir(\theta_g, ODF(x_i))) \quad (2)$$

where $ODF(x_i)$ in short denotes a tensor of point density values $ODF(x_i, \alpha_k, d_n, v_l)$ for different cones along a collection of direction vectors in \mathbb{S}^2 . Parameters $\theta_d = [\theta_d^1, \theta_d^2, \dots, \theta_d^n]$, and $\theta_g = [\theta_g^1, \theta_g^2, \dots, \theta_g^p]$ are learnable parameters of the ODFBlock. $ODFDir$ is an mlp that embeds the ODF tensor through aggregating features by collecting features of each direction, and $ODFGlob$ is another mlp which aggregates the output embedding over all directions to obtain the aggregate ODFBlock output features.

3.3. ODFNet Network Models

3.3.1. ODFNet

To exploit our point ODF's representation capability over point clouds, we design different ODFNets, which benefit from ODFs as well as point locations to capture both local and global features of point clouds and, which can be used for classification and segmentation tasks. As depicted in Figure 8, ODFNet first calculates the ODFs for each point as given by Equation 1 for N_c cones rotated around N_d direction vectors. Then for each point, it benefits from two different mlps inside ODFBlock: $ODFDir$ and $ODFGlob$. For differently scaled cones placed along the same direction, $ODFDir$ calculates features capturing the point density along that direction. $ODFGlob$ then operates on the features captured by $ODFDir$ in order to fuse those features and the result is later concatenated with the point coordinates.

In the blocks shown as $ODFBlock$ in Figure 8, in a similar manner with DGCNNs [7], each point's features are combined

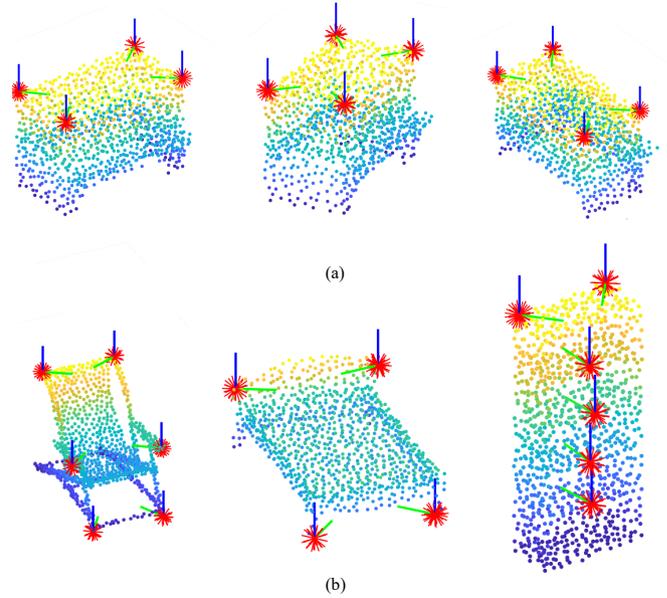


Fig. 7. For the RI-XY method, ODF directions according to nearest neighbors are depicted. Green arrows indicate pivot and blue arrows indicate the z-axis. (a) By rotating the objects in the x-y plane, ODF values do not change since the directions are aligned with respect to rotation-invariant pivots. (b) For symmetric points of the same object, nearly symmetric pivot directions are obtained.

with its nearest neighbors' features and the difference vector features, where the latter represents $\mathbf{x} - \mathbf{x}_i$ for \mathbf{x} indicating a point location and \mathbf{x}_i indicating one of its 32 neighbors.

For the classification task, the last layers of the architecture include a max-pooling operation to produce a global feature vector describing the shape. After three fully connected layers, class scores are obtained at the output. For part segmentation in ShapeNet, the ODFNet architecture makes use of the categorical vector, which indicates the one-hot-coded object class. It is fed to the segmentation part of the ODFNet, as in [7] and [2]. The output size of the final output layer in both tasks depends on the number of object classes and object parts. For semantic segmentation in S3DIS, nearly the same architecture with part classification is used. However, since the dataset also includes RGB color information, colors and color differences are also used to obtain difference features and location features.

3.3.2. ODFNet-xyz

To ensure the global invariance, we used RI-XYZ pivot selection method and slightly changed the architecture as shown in Figure 8.e to obtain ODFNet-xyz, where instead of difference vectors we used a vector set of magnitude of difference vectors $|\mathbf{x} - \mathbf{x}^i|$, point's distances to object center $|\mathbf{x} - \mathbf{c}_{object}|$, point neighbors' distances to object center $|\mathbf{x}^i - \mathbf{c}_{object}|$, angles between points and their neighbors $\angle(\mathbf{x}, \mathbf{x}^i)$, and angles between points and object center $\angle(\mathbf{x}, \mathbf{c}_{object})$. Also $\angle(\mathbf{x}, \mathbf{c}_{object})$ and $|\mathbf{x} - \mathbf{c}_{object}|$ are concatenated to the features from the last block of the network.

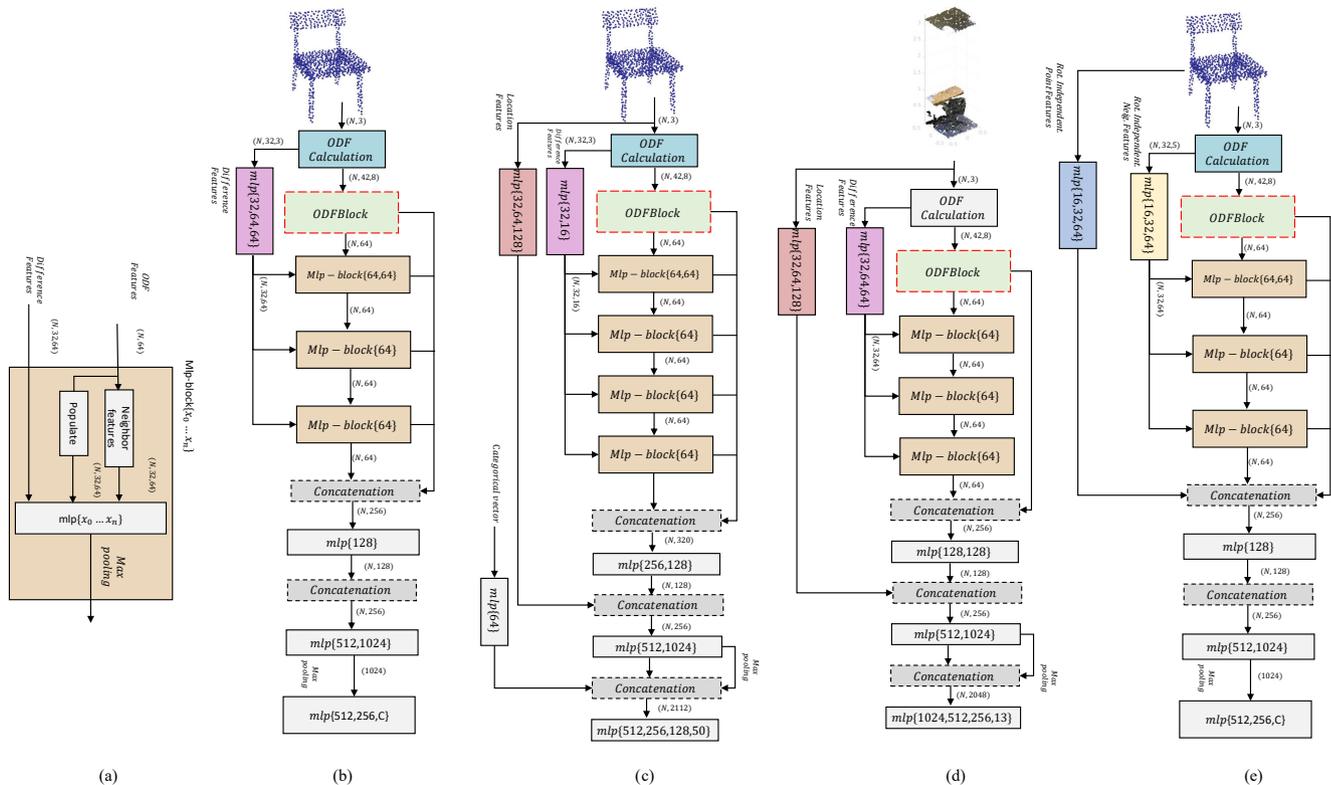


Fig. 8. (a) Mlp-block common structure in the ODFNet that is employed for tasks of (b) classification, (c) part segmentation (d) Scene segmentation. (e) represents the ODFNet-XYZ which is fully rotation-invariant. c represents class count.

4. Experimental Results

In this section, we present the details of the experiments and performance evaluation results for ODFNet on classification and segmentation tasks¹. We select widely used benchmark datasets and evaluate the ODFNet model over those, in order to provide a comparison with the existing methods. Particularly, we experiment and report results on ModelNet40 and ScanObjectNN for classification, ShapeNet for part segmentation, and S3DIS for scene segmentation.

4.1. Shape Classification

For shape classification, we evaluated our model on ModelNet40 [6] which consists of mesh models for 40 different categories, and ScanObjectNN [13] which contains 3D real-life scans for 13 different object categories which also contains noise and missing parts.

ModelNet: To have a fair comparison, we use the preprocessed data from [2] and use 1024 points for each object. Following the previous studies [7, 2], the objects are fit into a unit sphere. Recent works prefer to do scaling and translation [8], scaling and perturbing [7], and only perturbing [9] for data augmentation during training. Instead, nonuniform scaling, flip-

ping in x and y directions, and rotation by multiples of 90 degrees are applied. Also, random sampling is used as in [24] by deleting half of the points before the last classification block before max-pooling. Since the classification features are obtained by a max-pooling layer, the deletion operation does not affect the network structure. The ODFs are calculated regardless of this operation to avoid shape inconsistencies. To make a fair comparison with the previous work, we compared our study with the methods that use only 1024 points for each object. The results in Table 1 show that our implementation outperforms other methods by an overall accuracy score of 93.4% via a single prediction. Also, by applying a voting mechanism with random scaling and averaging the predictions, we obtained state-of-the-art results among the studies that follow a similar voting procedure.

Using the RI-XYZ pivot to calculate the ODFs, and ODFNet-xyz network architecture, we also evaluated our ODF features in three different scenarios focusing on rotation invariance: train and test with z rotations (z/z), train and test with SO3 rotations (SO3/SO3), train with z rotations and test with SO3 rotations (z/SO3).

In Table 2, we examined the results for these experiments in three groups of works (separated by horizontal lines in the table). The first group of networks [16, 47, 48, 19, 2, 3, 24, 44] which indicates the rotation-variant networks, for the z/z scenario, generally achieves scores that are comparable to their scores for the default setup. However, they lack robustness

¹The source code for our ODFNet model will be provided at the time of publication.

Table 1. Overall Accuracies (OA) for point cloud classification results on ModelNet40 dataset. (p: points, n: normals)

Method	input	voting	OA
PointNet [2]	p		89.2
PointNet++ [3]	p+n		90.7
SpiderCNN [5]	p+n		92.4
Point2Seq [40]	p		92.6
InterpCNN [41]	p		93.0
PointwiseCNN [23]	p		86.1
ShapeContextNet [42]	p		90.0
KCNet [43]	p		91.0
PointCNN [24]	p		92.2
RS-CNN [44]	p		92.4
ShellNet [9]	p		93.1
DGCNN [7]	p		92.9
ODFNet	p		93.4
Kd-network [45]	p	✓	91.8
GDANet [46]	p	✓	93.8
DensePoint [8]	p	✓	93.2
RS-CNN [44]	p	✓	93.6
ODFNet	p	✓	94.2

and their performance drops drastically for SO3/SO3 and z/SO3 scenarios. For the second group of approximately rotation-invariant networks, small standard deviations of accuracy are obtained for different scenarios. The third group, which also includes the ODFNet-xyz, is fully rotation-invariant as verified by the zero standard deviation values. The performances of rotation-invariant methods are consistent across the three scenarios. According to the scores, the ODFNet achieves the second-best results for the more challenging SO3/SO3 and z/SO3 scenarios among totally rotation-invariant methods.

Despite ODFNet’s success for the z/z scenario and the original scenario (Table 1), its accuracy is drastically decreased for the z/SO3 scenario. It is a natural result of depending on the object’s alignment to the XY-plane for both the training procedure and the architecture.

ScanObjectNN: We use the original dataset that has 2048 points for each object to have a fair comparison. To evaluate the ODFNet on ScanObjectNN, an augmentation procedure similar to the ModelNet experiments is used. However, because the object point counts are larger here, 1024 points are randomly selected at train time. There are five different tasks: OBJ_ONLY, OBJ_BG, PB_T25, PB_T25R, PB_T50R, and PB_T50RS. In OBJ_BG, objects with background noise are classified. OBJ_ONLY consists of objects having no background noise. The other sets are augmented versions of OBJ_BG. Comparing our results with the scores given in [13], the ODFNet model produces SoTA accuracy scores as can be observed in Table 3.

Further Experiments: To further investigate our network’s capacity for point cloud object abstraction, for each test subject, we extract the output of the last classification layer for the ODFNet, as well as for DensePoint and ShellNet, where the latter two are the previous SoTA point cloud architectures. Then, using two widely utilized dimensionality reduction tech-

Table 2. Comparisons of the classification accuracy under different rotation settings. Best results are bolded and second bests are underlined. std. represents the standard deviation of accuracy between different settings. The ODFNet-xyz architecture is only used for these experiments.

Method	z/z	SO3/SO3	z/SO3	std.
VoxNet [16]	83.0	87.3	-	3.0
SubVolSup [47]	88.5	82.7	36.6	28.4
SphericalCNN [48]	88.9	86.9	78.6	5.5
MVCNN 80x [19]	90.2	86.0	81.5	4.3
PointNet [2]	87.0	80.3	21.6	41.0
PointNet++ [3]	89.3	85.0	28.6	33.8
PointCNN [24]	<u>91.3</u>	84.5	41.2	27.2
RS-CNN [44]	<u>90.3</u>	82.6	48.7	22.1
ODFNet	<u>91.3</u>	88.0	17.4	34.0
RIConv [31]	86.5	86.4	86.4	<u>0.1</u>
SPHNet [49]	87.0	87.6	86.6	<u>0.5</u>
SFCNN [30]	92.3	91.0	85.3	3.5
ClusterNet [50]	87.1	87.1	87.1	0.0
RI-GCN [32]	<u>91.0</u>	91.0	91.0	0.0
GCANet [33]	89.0	89.2	89.1	0.0
ODFNet-xyz	90.2	<u>90.2</u>	<u>90.2</u>	0.0

niques, UMAP [53] and t-SNE [54], we project those vectors onto 2D space and assess this mapping by the Silhouette score [55], which evaluates whether each object is well matched to its own cluster. The scores are given in Table 5, and the projections are visualized in Figure 9. As can be seen from the results, although quantitative scores on the classification of these methods are close to each other, ODFNet’s features appear more distinctive and produce a relatively more separated layout than the other two methods, which is also observed in the Silhouette scores.

Furthermore, to understand which points generate global features at the classification step, features before the last max-pooling of the network are examined. Heat maps according to the contribution of each point to the final result are obtained for ODFNet and DGCNN as given in Figure 10. It can be observed that the ODFNet selects more diverse and seemingly important feature points from the point clouds when compared to the DGCNN². According to the visual results, we can conclude that for objects with relatively more complex geometric shapes, visually representative points, which are mostly corners and end-points, are selected by the ODFNet. The ODFs for those points that are around corners and edges have anisotropic distributions while the ODFs on the flat areas have isotropic distributions with similar strengths over many directions.

We also further investigate our method to analyze the effects of two important decisions we make when calculating ODF values: number of different conic neighborhood directions and pivot selection. Table 6 shows the accuracies for different settings. The results for Experiment A, B, and C indicate that 42 directions which are obtained by the second tessellation of the icosahedron give the best performance compared to the first (12

²Pointwise heat maps cannot be created for DensePoint and ShellNet since they do not use directly the points but their distribution.

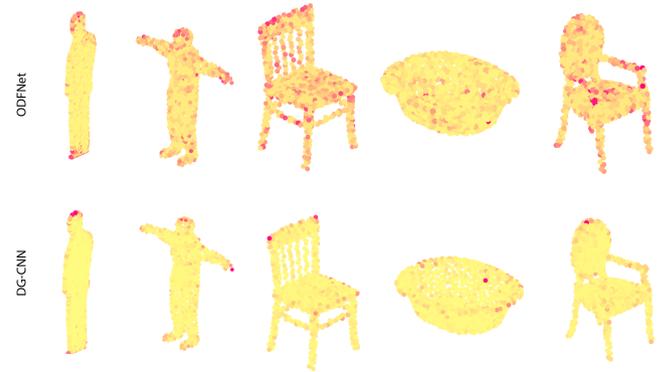
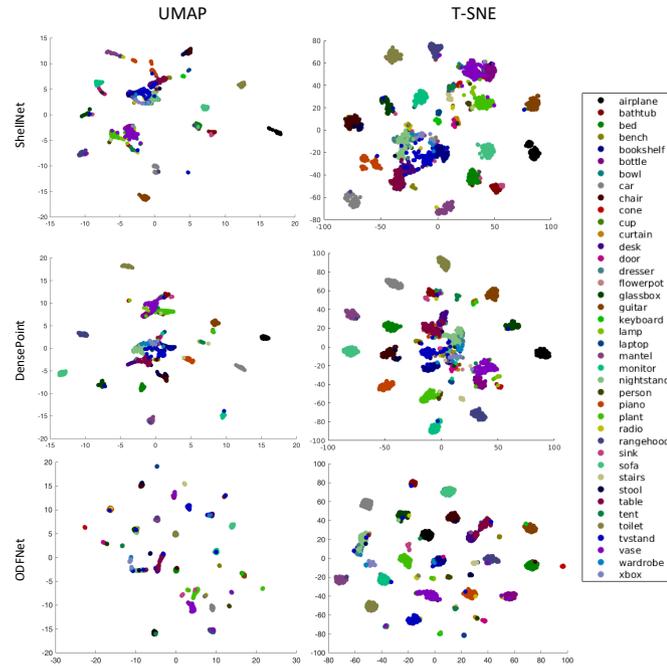
Table 3. Classification Accuracies for different tasks in ScanObjectNN [13] dataset.

Model	OBJ_BG	PB_T25	PB_T25R	PB_T50R	PB_T50RS	OBJ_ONLY
3DmFV [51]	68.2	67.1	67.4	63.5	63.0	73.8
PointNet [2]	73.3	73.5	72.7	68.2	68.2	79.2
SpiderCNN [5]	77.1	78.1	77.7	73.8	73.7	79.5
PointNet++ [3]	82.3	82.7	81.4	79.1	77.9	84.3
DGCNN [7]	82.8	83.3	81.5	80.0	78.1	86.2
PointCNN [24]	86.1	83.6	82.5	78.5	78.5	85.5
ODFNet	87.2	88.9	86.7	88.8	85.1	89.3

Method	input	mpIoU	mIoU	a.plane	bag	cap	car	chair	e.phone	guitar	knife	lamp	laptop	m.bike	mug	pistol	rocket	s.board	table
PointNet [2]	p	80.4	83.7	83.4	78.7	82.5	74.9	89.6	73.0	91.5	85.9	80.8	95.3	65.2	93.0	81.2	57.9	72.8	80.6
PointNet++ [3]	p+n	81.9	85.1	82.4	79.0	87.7	77.3	90.8	71.8	91.0	85.9	83.7	95.3	71.6	94.1	81.3	58.7	76.4	82.6
DGCNN [7]	p	84.6	85.2	84.0	83.4	86.7	77.8	90.6	74.7	91.2	87.5	82.8	95.7	66.3	94.9	81.1	63.5	74.5	82.6
PCNN [52]	p	81.8	85.1	82.4	80.1	85.5	79.5	90.8	73.2	91.3	86.0	85.0	95.7	73.2	94.8	83.3	51.0	75.0	81.8
DensePoint [8]	p	84.2	86.4	84.0	85.4	90.0	79.2	91.1	81.6	91.5	87.5	84.7	95.9	74.3	94.6	82.9	64.6	76.8	83.7
Point2Sequence [40]	p	82.2	85.2	82.6	81.8	87.5	77.3	90.8	77.1	91.1	86.9	83.9	95.7	70.8	94.6	79.3	58.1	75.2	82.8
ODFNet	p	83.3	86.5	85.1	85.0	89.5	78.7	91.9	73.6	92.2	88.2	85.9	96.1	74.9	95.3	82.2	53.7	77.7	83.3

Table 4. ShapeNet Part segmentation results for different architectures. Input column indicates whether points (p) and normals (n) are used.**Table 5. Silhouette scores for last layers of ODFNet, DGCNN and ShellNet after UMAP [53] and t-SNE [54] projection. {worst:best}:{-1:1}**

Method	tSNE S. Score	UMAP S. Score
ODFNet	0.623	0.660
DensePoint	0.453	0.474
ShellNet	0.472	0.466

**Fig. 10. Different heat maps for DGCNN and ODFNet.****Fig. 9. The projections from the last layer outputs of ODFNet, DensePoint and ShellNet architectures for ModelNet40 dataset onto a 2D space using the UMAP [53] and t-SNE [54] methods.****Table 6. An experiment on selection of hyperparameters: direction count and pivot selection. OA refers to overall accuracy.**

Experiment	Dir. Count	Pivot Sel.	OA
A	12	RI-XY	92.9
B	42	RI-XY	93.4
C	162	RI-XY	93.1
D	12	RI-XYZ	91.6
E	42	RI-XYZ	91.8
F	162	RI-XYZ	91.4

directions) and third (162 directions) tessellations. The results for the remaining experiments show that choosing pivots that are rotation-invariant in the x-y plane works better compared to the pivots that are fully rotation-invariant. Because all the training data is aligned according to the z-direction, using this direction as a pivot improves the performance.

4.2. Part Segmentation

We evaluated our segmentation network on the ShapeNet part segmentation benchmark [11], which contains 14007 training

and 1874 test samples, 16 object categories each of them partitioned into {2 – 6} parts, making a total of 50 parts. Following the practice of the previous state-of-the-art [8], which used ensembling, during the testing phase, for every test object, we also obtain scaled versions of test objects by +0.3%, -0.3% in each direction and averaged their scores. We compare our performance with those studies using a point cloud structure. Our experimental results for ShapeNet part segmentation are reported for ODFNet along with previous methods in Table 4.

4.3. Scene Segmentation

For scene segmentation, the commonly used S3DIS dataset [12] is utilized. The dataset contains point clouds sampled from six different challenging scenes. General practice in experimentation with this dataset involves training with a leave-one-out cross-validation (6-fold) strategy. State-of-the-art results are obtained for S3DIS for the overall accuracy measure as shown in Table 7.

Table 7. Overall accuracies and mIOU values for point cloud scene segmentation results on S3DIS dataset.

Method	OA	mIOU
PointNet [2]	78.6	47.6
PointNet++ [3]	81.0	54.5
PointSIFT [56]	88.7	70.2
Engelmann [57]	84.0	58.3
3DContextNet [58]	84.9	55.6
PointWeb [59]	87.3	66.7
ShellNet [9]	87.1	66.8
PointCNN [24]	88.1	65.4
InterpCNN [41]	88.7	66.7
DGCNN [7]	84.1	56.1
Liu et al. [60]	88.5	64.1
RandLA-Net [61]	88.0	70.0
HEPIN [62]	88.2	67.8
PointWeb [59]	87.3	66.7
CF-SIS [63]	88.0	74.0
ODFNet	90.8	72.2

4.4. Ablation Studies

We perform ablation studies on the ModelNet40 dataset for the classification task to further analyze our decisions for ODFNet and ODF-xyz.

To quantify the functionality of different modules of the ODFNet, we experimented with removing ODF-Dir and ODF-Glob blocks. For these experiments, we use 42 conic neighborhood directions and the RI-XY pivot selection method. The performance results in Table 8 show that both ODF-Dir and ODF-Glob blocks contribute to the performance.

For the ODFNet-xyz, another ablation study is carried out in order to analyze the two factors contributing to the rotation invariance: (i) using rotation invariant feature vectors instead of difference vectors; and (ii) using RI-XYZ, that is selection of pivot directions that are also rotation invariant, instead of RI-XY. The results given in Table 9 show that by using both factors, the best accuracy is obtained.

Table 8. Ablation study results on ODFNet. OA refers to overall accuracy.

Experiment	ODF-Dir	ODF-Glob	OA
A	✓	✓	93.4
B	✓		92.3
C		✓	91.9
D			91.2

Table 9. Ablation study results on ODFNet-xyz. OA refers to overall accuracy.

Exp.	Rot.Inv. Features	RI-XYZ	z/z	SO3/SO3	z/SO3	std.
E	✓		90.4	85.0	42.9	26.0
F		✓	91.6	89.5	24.9	37.9
G	✓	✓	90.2	90.2	90.2	0

5. Discussions and Conclusion

Our experimental results demonstrate that ODFNet achieves the SoTA performance in both the classification task and particularly for the more challenging segmentation task in ShapeNet and S3DIS. This provides evidence to our hypothesis that the ODF representation, which exploits the idea of incorporation of local point orientation distribution characteristics into the point cloud neural network models, is highly beneficial.

The point orientation distribution features help the neural network models capture further informative characteristics of the point cloud. This is revealed by the attention that the ODFNet model pays to the unique identifying points on an object such as corners, tips, and borders between planar regions. Features generated by the ODFNet correlate with an increased representation power for point clouds.

Moreover, we investigate the rotation invariance properties of ODF representations through a variant of the proposed ODFNet: ODFNet-xyz, which is fully rotation-invariant. Our experimental results show that ODF features can be effectively calculated in a rotation-invariant manner, and ODFNet-xyz performs comparably against state-of-the-art rotation-invariant point cloud analysis models.

From the results in Table 1 and 2, we can conclude that the original point locations and other location-based properties are highly discriminative. However, there is a trade-off between exploiting rotation-invariant features versus location-based rotation-variant point features, as location-based features are naturally not robust to rotation changes. Indeed, when the objects are rotated, the performances of rotation-variant networks drastically decrease. Moreover, the established benchmarks and procedures for evaluations on those benchmarks do not reward the rotation-invariant representations, as indicated by the inferior performance of rotation-invariant methods on those benchmarks.

As for future work, in geometric representation learning of point clouds, capturing essential defining characteristics of an object, whether through better augmented definitions of local patches around points as the ODFNet does or in other similarly effective ways of local and global encoding schemes, could provide further improvements in supervised and unsupervised

learning tasks on point clouds.

References

- [1] LeCun, Y, Bengio, Y, Hinton, G. Deep learning. *nature* 2015;521(7553):436–444.
- [2] Qi, CR, Su, H, Mo, K, Guibas, LJ. Pointnet: Deep learning on point sets for 3d classification and segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, p. 652–660.
- [3] Qi, CR, Yi, L, Su, H, Guibas, LJ. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In: *Advances in neural information processing systems*. 2017, p. 5099–5108.
- [4] Deng, H, Birdal, T, Ilic, S. Ppfnnet: Global context aware local features for robust 3d point matching. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, p. 195–205.
- [5] Xu, Y, Fan, T, Xu, M, Zeng, L, Qiao, Y. Spidercnn: Deep learning on point sets with parameterized convolutional filters. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018, p. 87–102.
- [6] Wu, Z, Song, S, Khosla, A, Yu, F, Zhang, L, Tang, X, et al. 3d shapenets: A deep representation for volumetric shapes. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, p. 1912–1920.
- [7] Wang, Y, Sun, Y, Liu, Z, Sarma, SE, Bronstein, MM, Solomon, JM. Dynamic graph cnn for learning on point clouds. *ACM Transactions on Graphics (TOG)* 2019;38(5):146.
- [8] Liu, Y, Fan, B, Meng, G, Lu, J, Xiang, S, Pan, C. Densepoint: Learning densely contextual representation for efficient point cloud processing. In: *Proceedings of the IEEE International Conference on Computer Vision*. 2019, p. 5239–5248.
- [9] Zhang, Z, Hua, BS, Yeung, SK. Shellnet: Efficient point cloud convolutional neural networks using concentric shells statistics. In: *Proceedings of the IEEE International Conference on Computer Vision*. 2019, p. 1607–1616.
- [10] Lei, H, Akhtar, N, Mian, A. Octree guided cnn with spherical kernels for 3d point clouds. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019, p. 9631–9640.
- [11] Yi, L, Kim, VG, Ceylan, D, Shen, I, Yan, M, Su, H, et al. A scalable active framework for region annotation in 3d shape collections. *ACM Transactions on Graphics (TOG)* 2016;35(6):210.
- [12] Armeni, I, Sener, O, Zamir, AR, Jiang, H, Brilakis, I, Fischer, M, et al. 3d semantic parsing of large-scale indoor spaces. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016, p. 1534–1543.
- [13] Uy, MA, Pham, QH, Hua, BS, Nguyen, T, Yeung, SK. Revisiting point cloud classification: A new benchmark dataset and classification model on real-world data. In: *Proceedings of the IEEE International Conference on Computer Vision*. 2019, p. 1588–1597.
- [14] Sedaghat, N, Zolfaghari, M, Amiri, E, Brox, T. Orientation-boosted voxel nets for 3d object recognition. *arXiv preprint arXiv:160403351* 2016;.
- [15] Li, Y, Pirk, S, Su, H, Qi, CR, Guibas, LJ. Fpnn: Field probing neural networks for 3d data. In: *Advances in Neural Information Processing Systems*. 2016, p. 307–315.
- [16] Maturana, D, Scherer, S. Voxnet: A 3d convolutional neural network for real-time object recognition. In: *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*. IEEE; 2015, p. 922–928.
- [17] Riegler, G, Ulusoy, AO, Geiger, A. Octnet: Learning deep 3d representations at high resolutions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, p. 3577–3586.
- [18] Sfikas, K, Pratikakis, I, Theoharis, T. Ensemble of panorama-based convolutional neural networks for 3d model classification and retrieval. *Computers & Graphics* 2018;71:208–218.
- [19] Su, H, Maji, S, Kalogerakis, E, Learned-Miller, E. Multi-view convolutional neural networks for 3d shape recognition. In: *Proceedings of the IEEE international conference on computer vision*. 2015, p. 945–953.
- [20] Zanuttigh, P, Minto, L. Deep learning for 3d shape classification from multiple depth maps. In: *Proceedings of IEEE International Conference on Image Processing (ICIP)*. 2017;.
- [21] Guo, H, Wang, J, Gao, Y, Li, J, Lu, H. Multi-view 3d object retrieval with deep embedding network. *IEEE Transactions on Image Processing* 2016;25(12):5526–5537.
- [22] Feng, Y, You, H, Zhang, Z, Ji, R, Gao, Y. Hypergraph neural networks. In: *Proceedings of the AAAI Conference on Artificial Intelligence*; vol. 33. 2019, p. 3558–3565.
- [23] Hua, BS, Tran, MK, Yeung, SK. Pointwise convolutional neural networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, p. 984–993.
- [24] Li, Y, Bu, R, Sun, M, Wu, W, Di, X, Chen, B. Pointcnn: Convolution on x-transformed points. In: *Advances in Neural Information Processing Systems*. 2018, p. 820–830.
- [25] Yang, Y, Feng, C, Shen, Y, Tian, D. Foldingnet: Point cloud auto-encoder via deep grid deformation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, p. 206–215.
- [26] Zamorski, M, Zieba, M, Klukowski, P, Nowak, R, Kurach, K, Stokowiec, W, et al. Adversarial autoencoders for compact representations of 3d point clouds. *Computer Vision and Image Understanding* 2020;193:102921.
- [27] Thomas, H, Qi, CR, Deschaud, JE, Marcotegui, B, Goulette, F, Guibas, LJ. Kpconv: Flexible and deformable convolution for point clouds. *arXiv preprint arXiv:190408889* 2019;.
- [28] Boulch, A. Convpoint: Continuous convolutions for point cloud processing. *Computers & Graphics* 2020;88:24–34.
- [29] Guo, Y, Wang, H, Hu, Q, Liu, H, Liu, L, Bennamoun, M. Deep learning for 3d point clouds: A survey. *arXiv preprint arXiv:191212033* 2019;.
- [30] Rao, Y, Lu, J, Zhou, J. Spherical fractal convolutional neural networks for point cloud recognition. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, p. 452–460.
- [31] Zhang, Z, Hua, BS, Rosen, DW, Yeung, SK. Rotation invariant convolutions for 3d point clouds deep learning. In: *2019 International Conference on 3D Vision (3DV)*. IEEE; 2019, p. 204–213.
- [32] Kim, S, Park, J, Han, B. Rotation-invariant local-to-global representation learning for 3d point cloud. *arXiv preprint arXiv:201003318* 2020;.
- [33] Zhang, Z, Hua, BS, Chen, W, Tian, Y, Yeung, SK. Global context aware convolutions for 3d point cloud understanding. *arXiv preprint arXiv:200802986* 2020;.
- [34] Tuch, DS. Q-ball imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 2004;52(6):1358–1372.
- [35] Cetin, S, Unal, G. A higher-order tensor vessel tractography for segmentation of vascular structures. *IEEE transactions on medical imaging* 2015;34(10):2172–2185.
- [36] Bloy, L, Verma, R. On computing the underlying fiber directions from the diffusion orientation distribution function. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer; 2008, p. 1–8.
- [37] Descoteaux, M, Angelino, E, Fitzgibbons, S, Deriche, R. Regularized, fast, and robust analytical q-ball imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 2007;58(3):497–510.
- [38] McGraw, T, Vemuri, BC, Yezierski, B, Mareci, T. Von mises-fisher mixture model of the diffusion odf. In: *3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro*. IEEE; 2006, p. 65–68.
- [39] Ehrlicke, HH, Otto, KM, Klose, U. Regularization of bending and crossing white matter fibers in mri q-ball fields. *Magnetic resonance imaging* 2011;29(7):916–926.
- [40] Liu, X, Han, Z, Liu, YS, Zwicker, M. Point2sequence: Learning the shape representation of 3d point clouds with an attention-based sequence to sequence network. In: *Proceedings of the AAAI Conference on Artificial Intelligence*; vol. 33. 2019, p. 8778–8785.
- [41] Mao, J, Wang, X, Li, H. Interpolated convolutional networks for 3d point cloud understanding. In: *Proceedings of the IEEE International Conference on Computer Vision*. 2019, p. 1578–1587.
- [42] Xie, S, Liu, S, Chen, Z, Tu, Z. Attentional shapecontextnet for point cloud recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, p. 4606–4615.
- [43] Shen, Y, Feng, C, Yang, Y, Tian, D. Mining point cloud local structures by kernel correlation and graph pooling. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, p. 4548–4557.
- [44] Liu, Y, Fan, B, Xiang, S, Pan, C. Relation-shape convolutional neural network for point cloud analysis. In: *Proceedings of the IEEE Conference*

- on Computer Vision and Pattern Recognition. 2019, p. 8895–8904.
- [45] Klokov, R, Lempitsky, V. Escape from cells: Deep kd-networks for the recognition of 3d point cloud models. In: Proceedings of the IEEE International Conference on Computer Vision. 2017, p. 863–872.
 - [46] Xu, M, Zhang, J, Zhou, Z, Xu, M, Qi, X, Qiao, Y. Learning geometry-disentangled representation for complementary understanding of 3d object point cloud. arXiv preprint arXiv:201210921 2020;.
 - [47] Qi, CR, Su, H, Nießner, M, Dai, A, Yan, M, Guibas, LJ. Volumetric and multi-view cnns for object classification on 3d data. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016, p. 5648–5656.
 - [48] Esteves, C, Allen-Blanchette, C, Makadia, A, Daniilidis, K. Learning so(3) equivariant representations with spherical cnns. In: Proceedings of the European Conference on Computer Vision (ECCV). 2018, p. 52–68.
 - [49] Poulencard, A, Rakotosaona, MJ, Ponty, Y, Ovsjanikov, M. Effective rotation-invariant point cnn with spherical harmonics kernels. In: 2019 International Conference on 3D Vision (3DV). IEEE; 2019, p. 47–56.
 - [50] Chen, C, Li, G, Xu, R, Chen, T, Wang, M, Lin, L. 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. 2019, p. 4994–5002.
 - [51] Ben-Shabat, Y, Lindenbaum, M, Fischer, A. 3dmfv: Three-dimensional point cloud classification in real-time using convolutional neural networks. IEEE Robotics and Automation Letters 2018;3(4):3145–3152.
 - [52] Atzmon, M, Maron, H, Lipman, Y. Point convolutional neural networks by extension operators. arXiv preprint arXiv:180310091 2018;.
 - [53] McInnes, L, Healy, J, Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:180203426 2018;.
 - [54] Maaten, Lvd, Hinton, G. Visualizing data using t-sne. Journal of machine learning research 2008;9(Nov):2579–2605.
 - [55] Rousseeuw, PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of computational and applied mathematics 1987;20:53–65.
 - [56] Jiang, M, Wu, Y, Zhao, T, Zhao, Z, Lu, C. Pointsift: A sift-like network module for 3d point cloud semantic segmentation. arXiv preprint arXiv:180700652 2018;.
 - [57] Engelmann, F, Kontogianni, T, Schult, J, Leibe, B. Know what your neighbors do: 3d semantic segmentation of point clouds. In: Proceedings of the European Conference on Computer Vision (ECCV). 2018, p. 0–0.
 - [58] Zeng, W, Gevers, T. 3dcontextnet: Kd tree guided hierarchical learning of point clouds using local and global contextual cues. In: Proceedings of the European Conference on Computer Vision (ECCV). 2018, p. 0–0.
 - [59] Zhao, H, Jiang, L, Fu, CW, Jia, J. Pointweb: Enhancing local neighborhood features for point cloud processing. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019, p. 5565–5573.
 - [60] Liu, J, Yu, M, Ni, B, Chen, Y. Self-prediction for joint instance and semantic segmentation of point clouds. In: European Conference on Computer Vision. Springer; 2020, p. 187–204.
 - [61] Hu, Q, Yang, B, Xie, L, Rosa, S, Guo, Y, Wang, Z, et al. Randlanet: Efficient semantic segmentation of large-scale point clouds. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020, p. 11108–11117.
 - [62] Jiang, L, Zhao, H, Liu, S, Shen, X, Fu, CW, Jia, J. Hierarchical point-edge interaction network for point cloud semantic segmentation. In: Proceedings of the IEEE International Conference on Computer Vision. 2019, p. 10433–10441.
 - [63] Wen, X, Han, Z, Youk, G, Liu, YS. Cf-sis: Semantic-instance segmentation of 3d point clouds by context fusion with self-attention. In: Proceedings of the 28th ACM International Conference on Multimedia. 2020, p. 1661–1669.