symmetry

MDPI

Article

PointSCNet: Point Cloud Structure and Correlation Learning Based on Space-Filling Curve-Guided Sampling

Xingye Chen [1], Yiqi Wu [1,2,*], Wenjie Xu [1], Jin Li [1], Huaiyi Dong [1] and Yilin Chen [2,3]

[1] School of Computer Science, China University of Geosciences, Wuhan 430074, China; cxy@cug.edu.cn (X.C.); xuwenjie@cug.edu.cn (W.X.); kimli@cug.edu.cn (J.L.); dhy@cug.edu.cn (H.D.)
[2] Hubei Key Laboratory of Intelligent Robot (Wuhan Institute of Technology), Wuhan 430205, China; yilinchen@wit.edu.cn
[3] School of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan 430205, China
* Correspondence: wuyq@cug.edu.cn

Abstract: Geometrical structures and the internal local region relationship, such as symmetry, regular array, junction, etc., are essential for understanding a 3D shape. This paper proposes a point cloud feature extraction network named PointSCNet, to capture the geometrical structure information and local region correlation information of a point cloud. The PointSCNet consists of three main modules: the space-filling curve-guided sampling module, the information fusion module, and the channel-spatial attention module. The space-filling curve-guided sampling module uses Z-order curve coding to sample points that contain geometrical correlation. The information fusion module uses a correlation tensor and a set of skip connections to fuse the structure and correlation information. The channel-spatial attention module enhances the representation of key points and crucial feature channels to refine the network. The proposed PointSCNet is evaluated on shape classification and part segmentation tasks. The experimental results demonstrate that the PointSCNet outperforms or is on par with state-of-the-art methods by learning the structure and correlation of point clouds effectively.

Keywords: point cloud; space-filling curve; structure correlation; feature extraction; deep learning

check for updates

Citation: Chen, X.; Wu, Y.; Xu, W.; Li, J.; Dong, H.; Chen, Y. PointSCNet: Point Cloud Structure and Correlation Learning Based on Space-Filling Curve-Guided Sampling. *Symmetry* **2022**, *14*, 8. https://doi.org/10.3390/sym14010008

Academic Editor: Tomohiro Inagaki

Received: 4 December 2021
Accepted: 14 December 2021
Published: 22 December 2021

1. Introduction

Point cloud is an ubiquitous form of 3D shapes and is suitable for countless applications in computer graphics due to its accessibility and expressiveness for 3D representation. The points are captured from the surface of objects by equipment such as 3D scanner, Light Detection and Ranging (LiDAR) or RGB-D cameras, or sampling from other 3D representations [1]. While containing rich information about the surface, structure, and shape of 3D objects, they are unlikely to have as ordered and structured data as images that are arranged on regular pixel grids. Hence, although many classical deep neural networks have shown tremendous success in image processing, there are still a lot of challenges when it comes to deep learning methods for point cloud [2].

To coordinate these incompatibilities, an intuitive idea is transforming the point cloud into a structured representation. Earlier multi-view methods tried to project the 3D object onto multiple view-wise images to fit 2D image processing approaches [3–7]. On the other hand, volumetric methods voxelized the point cloud to a regular 3D grid representation and adopted extensions of the 2D networks, such as 3D Convolutional Neural Network (CNN), for feature extraction [8]. Moreover, some following voxel-based research introduced certain data structures (such as octree) to reorganize the input shape [9–11]. While achieving impressive performances, these methods are often considered to have some inevitable shortcomings, such as losing 3D geometric information during 2D projection or having high computational and memory costs when processing voxels.

Against this backdrop, nowadays, research on directly consuming raw point clouds via end-to-end networks is becoming increasingly popular. The well-known PointNet [12] and subsequent PointNet++ [13] are the pioneer works of direct point cloud processing based on deep learning methods. The introduction of a symmetric function reflected by networks adapts to the inherent characteristics of a 3D points set. Inspired by PointNet and Pointnet++, many following research adopted the idea for point feature extraction or encoding to achieve the permutation invariance of point clouds [14–17].

The basic methodology of these point-based network is exacting point-wise high dimensional information and then aggregating a local or global representation of the point cloud for downstream tasks. Follow-up research based on this idea have demonstrated that the hierarchical structure with the subset abstraction procedure is effective for point cloud reasoning. It has been found that sampling central subsets of input points is essential for hierarchical structures [18,19]. However, the most popular sampling and grouping methods, Farthest Point Sampling (FPS) and K-Nearest Neighbor (KNN), are based on low-dimension Euclidean distance exclusively, without sufficient consideration of the semantically high-level correlations of the points and their surrounding neighbors.

In the real world, there are inherent correlations between local regions of 3D objects, especially for Computer Aided Design (CAD) models or industrially manufactured products [20], such as the symmetric wing design of an airplane; the regular arrays of wheels for a car; or the distinct structure between the collar, sleeves and body part of a shirt. These geometric correlations of local regions play a crucial role in 3D object understanding and are significant for typical point cloud processing tasks such as shape classification and part segmentation.

Moreover, in the procedure of high dimensional information extraction, a basic and effective approach is using shared Multi-Layer Perceptron (MLP) or 1D CNN to project the input feature to a high dimensional space. Inspired by applications of attention mechanism for image processing, it can be inferred that, similar to image processing, information of critical local areas and feature channel of the point cloud has more impact on specific tasks.

Based on these issues above, this paper proposes a point cloud feature extraction network, namely PointSCNet, which captures global structure and local region correlations of the point cloud for shape classification and part segmentation tasks. As shown in Figure 1, a space-filling curve-guided sampling module is proposed to choose key points that represent geometrically significant local regions from the point cloud. Then, an information fusion module is designed to learn the structure and correlation information between those local regions. Moreover, a channel and spatial attention module is adopted for the final point cloud feature refinement.

The main contributions of this paper are summarized as follows:

- An end-to-end point cloud processing network, namely PointSCNet, is proposed to learn structure and correlation information between local regions of a point cloud for shape classification and part segmentation tasks.
- The idea of a space-filling curve is adopted for points sampling and local sub-cloud generation. Specifically, points are encoded and sorted by Z-order curve coding, which makes the points contain meaningful geometric ordering.
- An information fusion module is designed to represent the local region correlation and shape structure information. The information fusion is achieved by correlating the local and structure feature via a correlation tensor and by skipping connection operations.
- A channel-spatial attention module is adopted to learn the significant points and crucial feature channels. The channel-spatial attention weights are learned for the refinement of the point cloud feature.

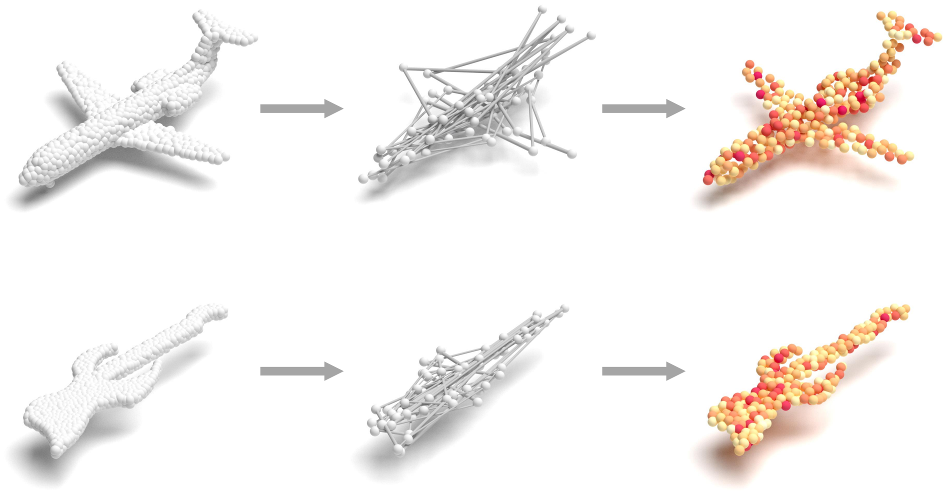

Figure 1. Learning structure and correlation on point cloud based on space-filling curve-guided sampling. Columns shown from left to right are the original input point cloud, points sampled by the Z-order space-filling curve, and the point cloud heat map based on the responses of points to the proposed network, respectively.

2. Related Work

This paper uses a deep learning method to extract point cloud features with construction and correlation information. In this section, recent research in highly related areas of our work, including traditional point cloud processing, point-wise embedding, point cloud structure reasoning, and attention in point cloud processing, are briefly summarized and analyzed.

2.1. Traditional Point Cloud Processing Methods

One of the biggest challenges in processing point clouds is dealing with unstructured point cloud data. The early methods of processing point clouds are mostly indirect representation conversion. Some methods try to convert the point cloud to structured data, such as octree and kd-tree [21] to reduce the difficulty of analysis. Another classical method converts the point cloud to voxel models. The voxel-based methods [3,22–24] use 3D convolution, which is a direct extension of image processing applications for point cloud. The advantages of the methods are that they can preserve the spatial relationship well at high voxel resolutions, but these methods are computationally very expensive. If the resolution of voxelization is reduced, the geometric information that the voxels can represent is significantly lost. FPNN [25] and Vote3 [26] proposed special methods to deal with the sparse problem, but their methods still cannot handle large-scale point cloud data well. Therefore, it is quite difficult to achieve real-time performance while considering the balance between accuracy and computational cost. Traditional methods inevitably lead to the loss of geometric information. This paper uses a point-by-point feature extraction method to overcome the high cost of voxel-based methods and is not conducive to processing low-resolution point clouds.

2.2. Point-Wise Embedding

The research of PointNet [12] is the breakthrough work for a deep learning-based direct point cloud processing method. Its groundbreaking proposal of a max-pooling symmetric function solves the problem of disordered point clouds. The MLP layer extracts the features and uses the maximum pooling aggregation to obtain the global features of the point clouds. Then, the PointNet++ [13] proposed a multi-layer sampling and grouping method to improve the PointNet. Much research on point cloud processing [27–31] later followed the idea of point-wise and hierarchical point feature extraction. However, the feature extraction in PointNet ignores geometrical structure information and the potential relationship between the local regions. Therefore, in this paper, the points are embedded based on the idea of PointNet++ first, and a space-filling curve-guided downsampling method and an information fusion method are proposed to learn the structure and correlation information of the point cloud.

2.3. Point Cloud Structure Reasoning

As an extension of point-wise feature learning, various methods have been proposed to reason the structure of points. The DGCNN [32] captures the features between point neighborhoods through graph convolution. The network extracts point cloud structure information by capturing the topological relationship between points. MortonNet [19] proposed an unsupervised way to learn the local structure of point clouds. In PCT [33], the KNN method is adopted to extract the features between the point fields. The SRN [14] uses a concatenation for structural features and position coding between local sub-clouds, and the multi-scale features extracted by the method are used for point cloud processing, which improves the PointNet++ [13]. However, these methods mainly pay attention to the relationship between local regions and ignore the relationship between the local region and the global shape. In this paper, a more effectively structure reasoning method is designed to capture the correlation between local regions and the shape structure.

2.4. Attention in Point Cloud Processing

Due to the advancement of an attention mechanism-based method in many deep learning applications [18,34–36], the attention mechanism meets the demand of dealing with unstructured data and is well applied in point cloud processing [33,37,38]. The Point Transformer [38] and Point Cloud Transformer [33] have made precedents for the application of the Transformer [36] in point cloud processing and achieved the state-of-the-art performance. The adoption of an attention mechanism for point cloud is mainly for exploring the relation between points and for enhancing the feature representation of attended points. Therefore, inspired by this idea, a channel-spacial attention module [39] is designed for feature refinement by enhancing key points and crucial feature channels.

3. Method

As shown in Figure 2, the proposed PointSCNet first uses the original point set $P \in \mathbb{R}^{N \times C}$ as the input. C is the feature channel of the point set. After a regular sampling and grouping [13] block, we obtain the sampled point set of $N^{'}$ points with the original spatial position information, denoted as $X_{position} \in \mathbb{R}^{N^{'} \times 3}$, and the embedded sampled point set with a $C^{'}$ dimension feature, in which each point represents information of the surrounding points within a certain radius, denoted as $X_{embedding} \in \mathbb{R}^{N^{'} \times C^{'}}$. Then, we send $X_{embedding}$ and $X_{position}$ to a Z-order sampling module for further sampling based on the points' geometrical relation. The sampled point set contains the shape structure and local regions correlation information, denoted as $X^{'}_{Z\text{-}order} \in \mathbb{R}^{N^{''} \times C^{'}}$ and $X_{Z\text{-}order} \in \mathbb{R}^{N^{''} \times 3}$. After that, an information fusion module is designed to establish the correlation between each local sub-cloud and the entire point cloud for the shape structure and local region correlation information learning. Moreover, after the information fusion procedure, the point cloud feature is forwarded to a channel-spatial attention module for feature refinement.

The pipeline of classification and segmentation module is similar to the PointNet++ [13]. The dimension of local point cloud features is increased to 1024 first, and then, an aggregate function pooling is adopted to obtain $X_{global} \in \mathbb{R}^{1 \times 1024}$ global features. For the shape classification task, after being fed into the fully connected layers, the dimension of the global feature is reduced to $1 \times k$ as the output of the PointSCNet, where k is the number of classes. For the part segmentation task, the output is the segmentation result $N \times k^{'}$ obtained by up-sampling the global feature X_{global}, where $k^{'}$ is the number of part classes.

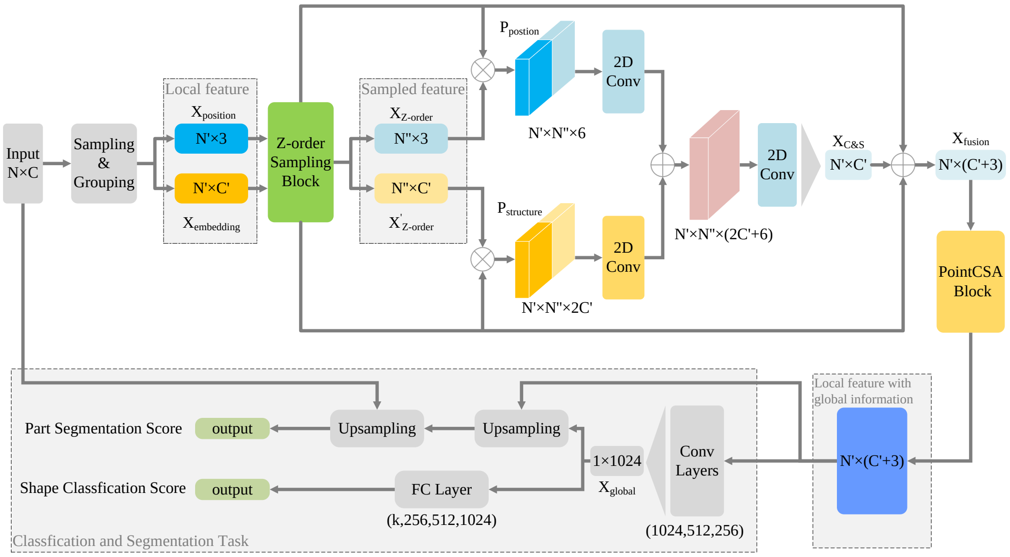

Figure 2. Model architecture of PointSCNet: The original point cloud is fed to a sampling and grouping block. Then, a Z-order sampling block is designed for further generation of local regions. After the sampled point cloud feature is extracted, the feature fusion module is designed to learn the structure and correlation information. Lastly, the point cloud feature is forwarded to the PointCSA block, which is based on a channel-spatial attention mechanism to obtain the refined feature for classification and segmentation.

3.1. Initial Sampling and Grouping

The PointSCNet first uses the original point cloud data as input. A series of points $X_{position} \in \mathbb{R}^{N' \times 3}$ in the space are sampled via FPS, and the ball query method is used to obtain all points that are within a radius to the sampled point, denoted as

$$d(X_r, X_{position}) < r, X_r \in \mathbb{R}^{N_r \times 3}, \tag{1}$$

where $X_{position} \in \mathbb{R}^{N' \times 3}$ are the points sampled by FPS, $X_r \in \mathbb{R}^{N_r \times 3}$ are the points around $X_{position}$, and $d()$ is the Euclidean distance.

These points are encoded to a high-dimensional space through MLPs and aggregated to the sampled point via the aggregation function $Pooling()$ to obtain $X_{embedding} \in \mathbb{R}^{N' \times C'}$, and the aggregation function can be denoted as

$$X_{embedding} = Pooling(Concat(Conv(X_r), X_r)), \tag{2}$$

where $X_{embedding} \in \mathbb{R}^{N' \times C'}$ is the encoded points feature, max-pooling function is used for pooling operation, the $Concat()$ function represents point feature concatenation, and the $Conv()$ function is the 1D convolution operation.

After this procedure, the feature information of neighboring points is aggregated to all sampled points $X_{embedding}$.

3.2. Z-Order Curve-Guided Sampling Module

The principle of a space-filling curve is to use a continuous curve to pass through all points in the space, and each point corresponds to a position code. After the FPS-based sampling and grouping, the Z-order curve coding function is adopted to further downsample the local sub-cloud $X_{embedding}$ to obtain local regions with semantically high-level correlations.

After the Z-order encoding, the 3D position coordinates of the local sub-cloud are mapped to the 1D feature space, as shown in Figure 3. The locality of the original point can be well preserved due to the nature of a Z-order curve, which means direct Euclidean neighbors in 1D tend to be similar to those in 3D. After the points are encoded and sorted, equally spaced points are sampled, as shown in Figure 4. Then, the point set with N'' points and C' dimension feature, denoted as $X'_{Z\text{-}order} \in \mathbb{R}^{N'' \times C'}$, and the point set with N'' points and 3D coordinate, denoted as $X_{Z\text{-}order} \in \mathbb{R}^{N'' \times 3}$, are sampled. The final sampled point set represents the global structure and local correlation of the original point set.

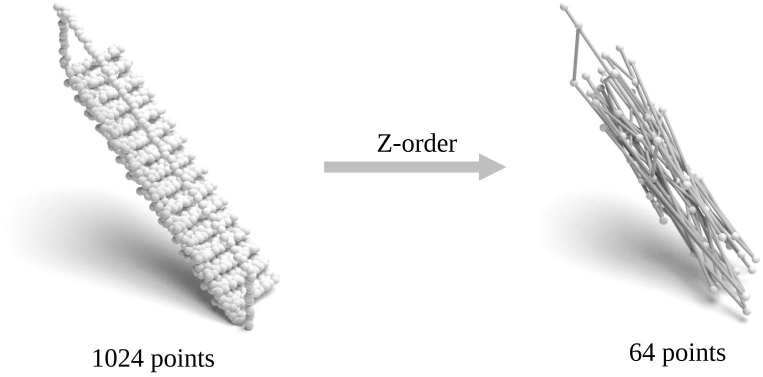

Figure 3. The point cloud structure obtained by sampling 1024 points in the original point cloud using the Z-order space-filling curve.

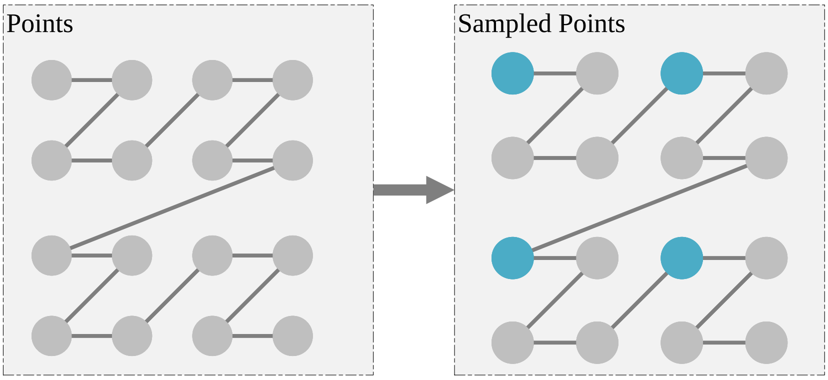

Figure 4. Sampling strategy based on Z-order curve sorting. Equally spaced points are sampled, and the spacing is set to 3 in the figure.

3.3. *Information Fusion of Local Feature and Structure Feature*

After obtaining the Z-order based sampled point cloud, the local sub-cloud feature and the structure feature are correlated to learn the shape structure and local region correlation information. As shown in Figure 5, a correlation tensor, represented as $N' \times N'' \times 2C$, is developed to evaluate the correlation between a local sub-cloud feature, represented as $N' \times C$, and a structure feature, represented as $N'' \times C$. The generation of the correlation tensor can be formalized as

$$P_{structure} = Fusion(X_{embedding}, X'_{Z\text{-}order}), P_{position} = Fusion(X_{position}, X_{Z\text{-}order}), \tag{3}$$

$$Fusion(X,Y) = \begin{bmatrix} Concat(X_1,Y_1) & Concat(X_1,Y_2) & \cdots & Concat(X_1,Y_n) \\ Concat(X_2,Y_1) & Concat(X_2,Y_2) & \cdots & Concat(X_2,Y_n) \\ \vdots & \vdots & \ddots & \vdots \\ Concat(X_m,Y_1) & Concat(X_m,Y_2) & \cdots & Concat(X_m,Y_n) \end{bmatrix}, \tag{4}$$

where $X \in \mathbb{R}^{n \times c}$ and $Y \in \mathbb{R}^{m \times c}$, and X, Y have the same numbers of feature channel. $X_i \in X$ and $Y_j \in Y$ are single points in point sets. The $Concat()$ function is proposed to concatenate the feature channels of X_i and Y_j.

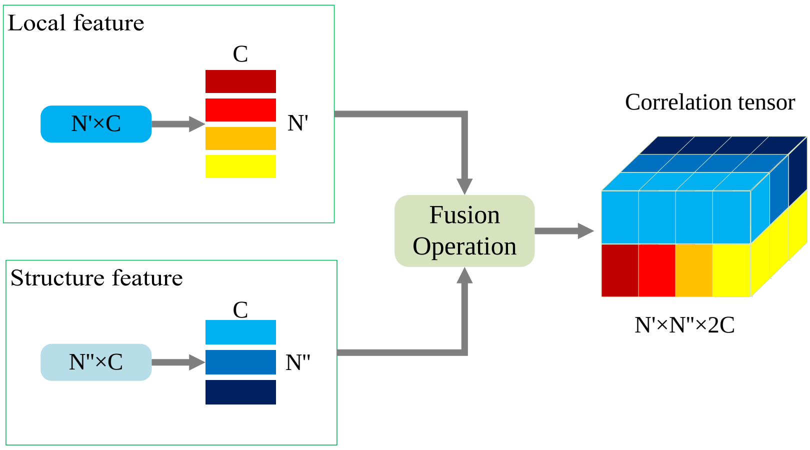

Figure 5. The correlation tensor is designed for the evaluation of the correlation between the local feature and structure feature. N' and N'' represent the number of points sampled via FPS and Z-order sampling block, and C is the feature channel of points.

Then, 2D convolution layers are designed to obtain the structure and local correlation of the point cloud, as shown in Figure 2. After the information fusion, a point cloud feature $X_{C\&S} \in \mathbb{R}^{N' \times C'}$ containing structure and correlation information is extracted. This process can be formalized as

$$X_{C\&S} = H(P_{structure}, P_{position}) = Pooling(Relu(g(Concat(P_{structure}, P_{position})))), \tag{5}$$

where $g()$ is the $Conv2d$ function and $Concat()$ is the concatenation operation.

Finally, as shown in Figure 2, $X_{C\&S}$, $X_{embedding}$, and $X_{position}$ are fused together to the fusion feature X_{fusion} via skip connections and the process can be formalized as

$$X_{fusion} = Concat(X_{embedding} + X_{C\&S}, X_{position}), \tag{6}$$

where $X_{embedding} \in \mathbb{R}^{N' \times C}$ represents local point cloud features, $X_{structure} \in \mathbb{R}^{N' \times C}$ represents skeleton point cloud features, $X_{position} \in \mathbb{R}^{N' \times 3}$ represents point cloud location features, and the Concat function represents feature dimension concatenation of points.

3.4. Points Channel-Spatial Attention Module

As shown in Figure 6, a channel-spatial attention module parallel to the channel attention block and the spatial attention block is adopted to strengthen the PointSCNet's ability by capturing the most important points and feature channels. In the channel attention module, the point cloud feature is aggregated by max-pooling and average pooling operation and then forwarded to convolution layers. The design of the convolution layer reduces the feature dimension first and then raises it for better feature extraction. The outputs of convolution layers are summed and activated to learn the weight of each feature channel. The channel attention block is formalized as

$$Channel(X) = ReLU(MLP(Max(X) + Avg(X))), \tag{7}$$

where $X \in \mathbb{R}^{N' \times C'}$,$Max()$ and $Avg()$ represent max-pooling and average-pooling functions. In the spatial attention module, the feature is fed to the MLPs with shared weights, and then, the information on each channel is aggregated through the batch normalization layer and the pooling layer to obtain the spatial position attention weight. The spatial attention block is formalized as

$$Spatial(X) = Pooling(BN(MLP(X))). \tag{8}$$

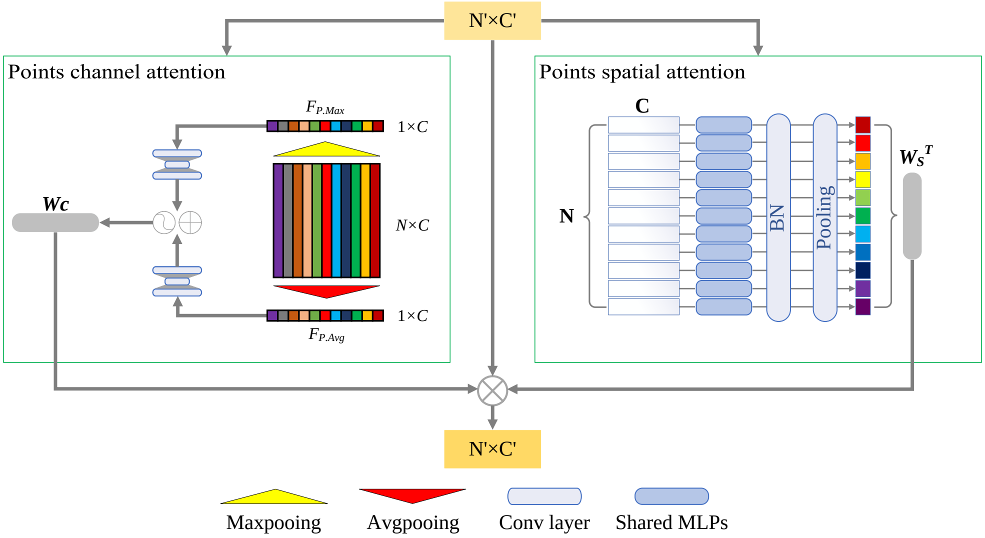

Figure 6. Points channel-spatial attention module: The points feature is fed to the channel-spatial module to capture the most important points and feature channels. In channel attention module, channel weights are obtained via the two aggregation functions and convolution layers. In the spatial attention module, spatial weights are obtained via shared MLPs.

4. Experiments

In this section, some quantitative and qualitative experiments are designed to demonstrate the performance of our proposed PointSCNet. First, the network is evaluated on shape classification and part segmentation tasks. Then, more quantitative analyses of the network are presented. Moreover, some more visualization experiments are performed to demonstrate the ability of PointSCNet quantitatively. Finally, the ablation study is designed to show the effectiveness of each module of PointSCNet. The source code of the PointSCNet is available at https://github.com/Chenguoz/PointSCNet (accessed on 1 December 2021).

4.1. Implementation Details

The development environment is Ubuntu18.04+Cuda11.1+Pytorch1.8.0, and the hardware environment includes a GPU device, an RTX3080 single discrete graphics card. In the classification task, we set the random sampling of 1024 points as the input of the PointSCNet and the random sampling of 2048 points in the segmentation task. Our training hyperparameters are set to the batch size of 24, the number of iterations is set to 200, the initial learning rate is set to 1×10^{-3}, and the learning rate decays to the original 0.9 after every 20 iterations. The optimizer is Adam, and the weight decay rate is 1×10^{-4}. In Z-order sampling, the number of sampled points is set to 64. The loss is measured by calculating the cross entropy between the real label and the predicted value. We set the number of local sub-clouds as $N^{'} = 256$, and the number of local sub-cloud feature channels as $C = 192$. The number of skeleton sub-clouds is $N^{''} = 64$.

4.2. Shape Classification on ModelNet40

The shape classification experiment was performed on the ModelNet40 [23] dataset, which is the most commonly used dataset for training point cloud classification networks. The dataset has 9843 training data and 2468 test data, belonging to 40 different shape classes.

In the shape classification experiment, we randomly sampled 1024 point features as the input, and 64 regions were sampled based on the Z-order curve coding. The PointSCNet is compared with some of the state-of-the-art methods. As shown in Table 1, the overall classification accuracy of PointSCNet on ModelNet40 reaches 93.7%, which outperforms or is on par with classical classification networks and recent state-of-the-art methods.

Table 1. Comparison with state-of-the-art methods on the ModelNet40 classification dataset. The column of "Acc" means overall accuracy(%). All results quoted are taken from the cited papers. "xyz" in the column of Input means the 3D coordinate of points and nr means normal.

Method	Input	Points	Acc
Pointnet [12]	xyz	1024	89.2
Pointnet++ [13]	xyz	1024	90.7
Kd-Net [21]	xyz	32k	91.8
DGCNN [40]	xyz	1024	92.9
SRN [14]	xyz	1024	91.5
PointGrid [11]	xyz	1024	92.0
PointCNN [41]	xyz	1024	92.2
RS-CNN [42]	xyz	1024	93.6
PCT [33]	xyz	1024	93.6
PAConv [43]	xyz	1024	93.9
CurveNet [44]	xyz	1024	93.8
RPNet-W9 [45]	xyz	1024	93.9
Pointnet++ [13]	xyz,nr	1024	91.7
PAT [16]	xyz,nr	1024	91.7
SpiderCNN [46]	xyz,nr	5k	92.4
A-CNN [47]	xyz,nr	1024	92.6
PointASNL [48]	xyz,nr	1024	93.2
SO-Net [49]	xyz,nr	1024	93.4
PointSCNet	xyz,nr	1024	**93.7**

PointNet++ [13] is the pioneer work of hierarchical point cloud feature extraction, which captures the multi-scale local structure with hierarchical layers. It aggregates local features through a simple maximum pooling operation without using their structural relationship. The DGCNN [40] and the SRN [14] are both classical methods used to learn structural relation of point cloud. The DGCNN [40] simply concatenates the feature relationships of local sub-clouds in different dimensions, and the captured structural relationship features cannot fully represent the structure of the point cloud.The SRN [14] adopts a regular FPS-based sampling and grouping method to obtain local point clouds and simply concatenates the points position and geometry feature to capture the structure relationship. The PointSCNet uses the space-filling curve to sample the points in the point cloud that can characterize the point cloud structure and, then, processes them through a specially designed feature fusion module to explore the correlations between local regions and the structure of the point cloud. The performance is significantly improved compared with these baseline methods.

4.3. Part Segmentation on ShapeNet

The ShapeNet [50] dataset covers 55 common object categories, and there are approximately 51,300 3D models. The part segmentation task is performed on the ShapeNet part segmentation dataset with 16,880 models and 16 categories. The 3D models are divided into 14,006 training point clouds and 2874 test point clouds, where each point is associated with a point-by-point label of the point cloud segmentation task. In the point cloud component segmentation task, we randomly sampled 2048 point features as the original input of PointSCNet. The quantitative results of PointSCNet and some classical state-of-the-art methods are shown in Table 2. By capturing the skeleton structure features of the point cloud, the PointSCNet significantly outperforms Pointnet [12], Pointnet++ [13], and SRN [14], and the performance of PointSCNet is particularly outstanding in some specific classes, as shown in the table. Figure 7 shows the visualization part segmentation result of PointSCNet.

Table 2. The performance of part segmentation task on ShapeNet. The metric is part-average Intersection-over-Union (IoU, %). All results quoted are taken from the cited papers.

Class	**Pointnet [12]**	**Pointnet++ [13]**	**SRN [14]**	**PCNN [51]**	**PointCNN [41]**	**PointSCNet**
Airplane	83.4	82.3	82.4	82.4	84.1	83.3
Bag	78.7	79.7	79.8	80.1	86.4	84.3
Cap	82.5	86.1	88.1	85.5	86.0	**88.1**
Car	74.9	78.2	77.9	79.5	80.8	79.2
Chair	89.6	90.5	90.7	90.8	90.6	**91.0**
Earphone	73.0	73.7	69.6	73.2	79.7	74.3
Guitar	91.5	91.5	90.9	91.3	92.3	91.2
Knife	85.9	86.2	86.3	86.0	88.4	87.4
Lamp	80.8	83.6	84.0	85.0	85.3	84.5
Laptop	95.3	95.2	95.4	95.7	96.1	95.7
Motorbike	65.2	71.0	72.2	73.2	77.2	73.4
Mug	93.0	94.5	94.9	94.8	95.3	**95.3**
Pistol	91.2	80.8	81.3	83.3	84.2	81.7
Rocket	57.9	57.7	62.1	51.0	64.2	60.7
Skateboard	72.8	74.8	75.9	75.0	80.0	75.9
Mean	83.7	85.1	85.3	85.1	86.1	85.6

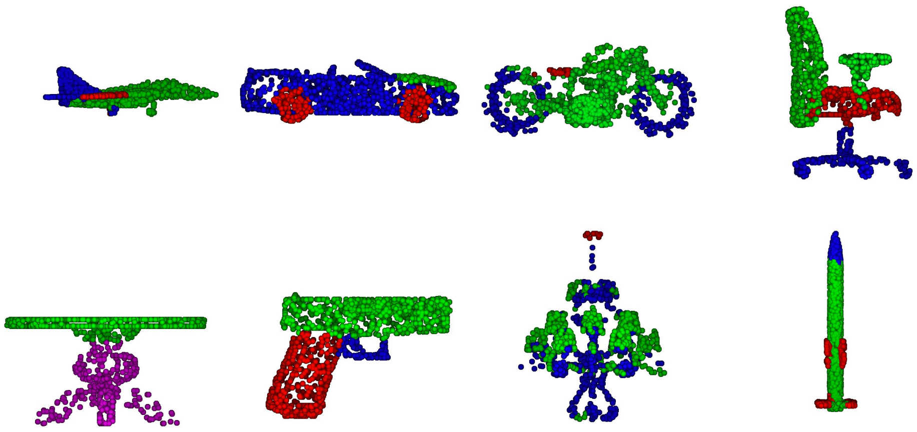

Figure 7. Results of our PointSCNet on the part segmentation.

4.4. Additional Quantitative Analyses

The number of model parameters reflect the training speed of the network indirectly. Our PointSCNet adopts the space-filling curve-guided sampling strategy to capture a few points to represent local regions and the structure of the point cloud, which reduces the number of model parameters. The PointSCNet achieves outstanding classification accuracy with relatively few model parameters, as shown in Table 3. For the PointNet [13], its multi-layer sampling structure introduces redundant information and slows down the training speed. Figure 8 shows the loss curve of PointSCNet decreases more rapidly compared with the Pointnet++. The PCT [33] network uses the Transformer structure repeatedly to capture the structural relationship characteristics of the point cloud. Hence, it has excessive parameters and its convergence speed is slow. The SRN [14] adopts a regular FPS-based sampling and grouping method to obtain sub-regions and adopts a duplicated SRN module, which leads to a large number of parameters and a slow convergence speed too.

Table 3. The performance of PointSCNet on the ModelNet40 dataset to test classification tasks.

Method	Params	Acc
Pointnet [12]	3.472 M	89.2
Pointnet++ [13]	1.748 M	91.9
SRN [14]	3.743 M	91.5
DGCNN [40]	1.811 M	92.9
NPCT [33]	1.36 M	91.0
SPCT [33]	1.36 M	92.0
PCT [33]	2.88 M	93.2
PointSCNet	**1.827 M**	**93.7**

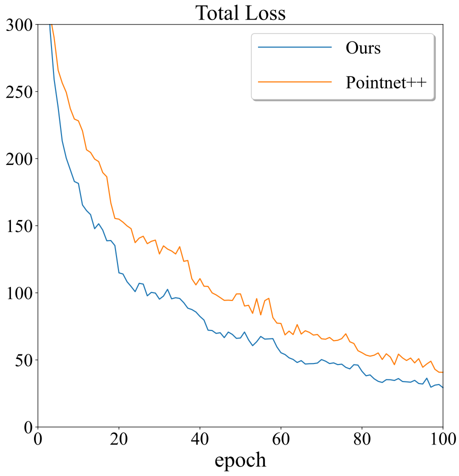

Figure 8. Experiment on the drop speed of the loss curve.

4.5. Additional Visualization Experiments

The heat map for points with a high response to PointSCNet is shown in Figure 9. The points are colored according to their response to the network and those with higher response are colored darker. The darker points in of the mug display the model structure. The darker points in the airplane mainly gather on one side of the symmetry axis, which indicates the symmetry of the airplane model. In the table model, both the model structure and a repetitive arrayed table leg are emphasized. The points with high responses appear on the rim of the bowl.

According to the visualization results, points with higher responses either present the structure of a point cloud or show the geometrical and locational interactions of local regions, which proves that the points sampled by the Z-order sampling module represents meaningful geometrical local regions and that the information fusion block extracts the structure and correlation information effectively.

Figure 10 shows the performance of our PointSCNet in feature extraction. By using t-SNE [52] to reduce the dimension of high-dimensional features to 2D, the classification ability of our network is visualized as shown in the figure. It can be seen that most of the classes are divided into separate clusters. For some clusters with similar point cloud structures, such as tables and stools being close in semantic space, the PointSCNet can still distinguish them precisely.

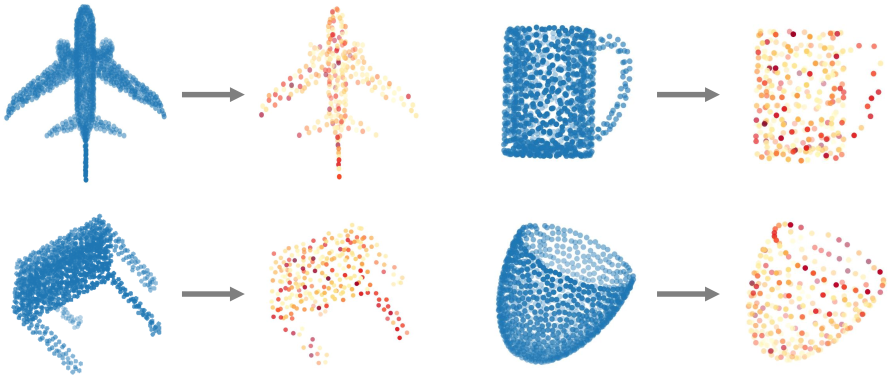

Figure 9. Heat map for points with high responses to PointSCNet.

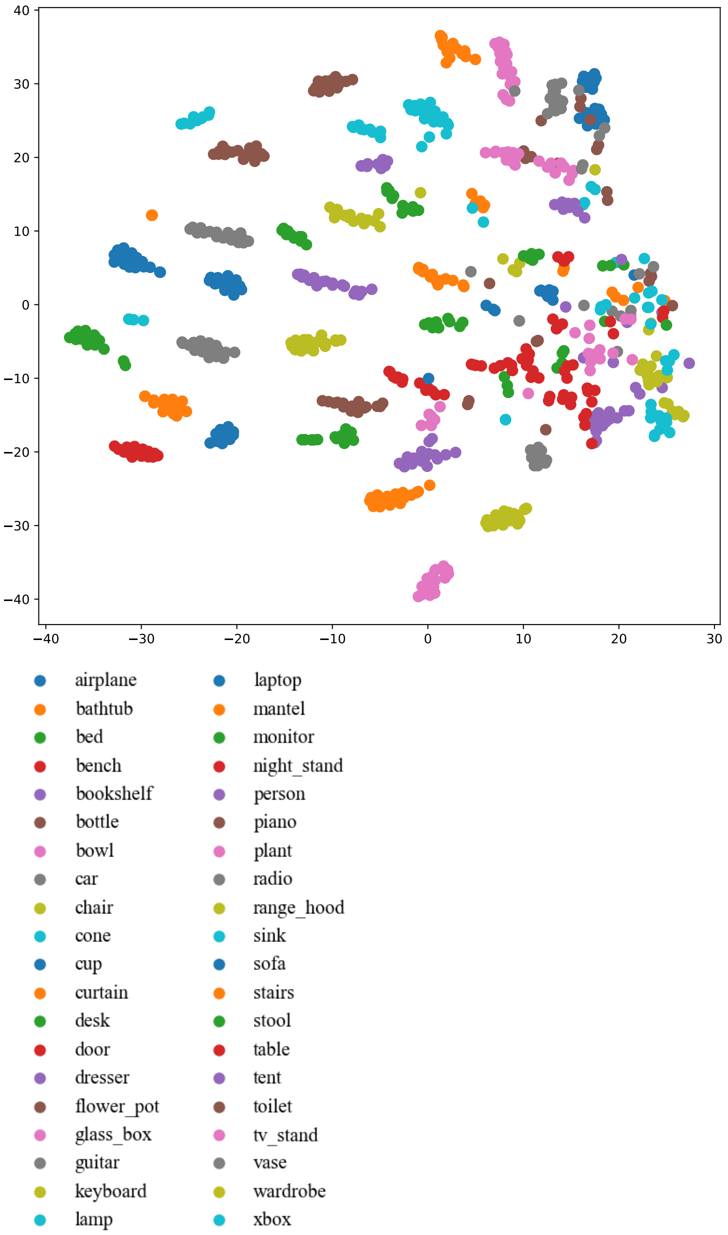

Figure 10. Visualization results of t-SNE on the ModelNet40 dataset.

4.6. Ablation Study

A set of ablation studies were designed to test the impacts of critical components of our network, including the Z-order sampling block (Section 3.2), structure and correlation information fusion module (Section 3.3) and the channel-spatial attention block (Section 3.4). The ablation strategies and results are shown in Table 4. It can be found that all of the critical components of the PointSCNet improve the network performance. The Z-order sampling block and C&S module provide obvious improvement. The convergence speed is slow while only using the information fusion module. When all these three modules are used at the same time, the model training speed is greatly improved, and the highest accuracy of the classification task is achieved, which further proves the importance of each module.

Table 4. The strategies and results of ablation studies. "ZS" represents the Z-order curve guided sampling block. "C&S" represents the structure and correlation information fusion module. "AM" is the channel-spatial attention module. "✓" represents existence, and "×" represents inexistence. "ToBestAcc" is the minimum number of epochs when the PointSCNet achieves the highest accuracy in the training phase.

Methods	ZS	C&S	AM	Acc	ToBestAcc/Epochs
A	✓	×	×	93.0	87
B	✓	✓	×	93.4	95
C	✓	×	✓	93.2	85
D	×	✓	✓	93.3	120
E	×	✓	×	93.2	148
F	×	×	✓	93.2	73
PointSCNet	✓	✓	✓	93.7	67

5. Conclusions

In this paper, a point cloud processing network named PointSCNet is proposed to learn the shape structure and local region correlation information based on space-filling curve-guided sampling. Different from most existing methods using FPS method for downsampling, which only utilizes the low-dimension Euclidean distance, our proposed space-filling curve-guided sampling module uses the Z-order curve for sampling to explore high-level correlations of points and local regions. The feature of sampled points are fused in the proposed information fusion block, in which the shape structure and local region correlation are learned. Finally, the channel-spatial module is designed to enhance the feature of key points. Quantitative and qualitative experimental results demonstrate that the proposed PointSCNet learns the point cloud structure and correlation effectively and achieves superior performance on shape classification and part segmentation tasks. The idea of structure and correlation learning can be adopted for related vision tasks other than 3D points processing. Hence, in the future, we plan to optimize our network and to apply the method to more vision scenarios [53–55].

Author Contributions: X.C. and Y.W. conceived and designed the algorithm and the experiments. X.C. wrote the manuscript. Y.W. supervised the research. W.X. assisted in the experiments. Y.W. provided suggestions for the proposed method and its evaluation and assisted in the preparation of the manuscript. W.X. and Y.C. assisted in the experiments. J.L. analyzed the data. H.D. collected and organized the literature. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the National Science Foundation of China (grant No. 61802355 and 61702350) and Hubei Key Laboratory of Intelligent Robot (HBIR 202105).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Guo, Y.; Sohel, F.; Bennamoun, M.; Lu, M.; Wan, J. Rotational projection statistics for 3D local surface description and object recognition. *Int. J. Comput. Vis.* **2013**, *105*, 63–86. [CrossRef]
2. Guo, Y.; Wang, H.; Hu, Q.; Liu, H.; Liu, L.; Bennamoun, M. Deep learning for 3d point clouds: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 4338 – 4364. [CrossRef]
3. Qi, C.R.; Su, H.; Nießner, M.; Dai, A.; Yan, M.; Guibas, L.J. Volumetric and multi-view cnns for object classification on 3d data. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 5648–5656.
4. 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, Santiago, Chile, 7–13 December 2015; pp. 945–953.
5. Chen, X.; Ma, H.; Wan, J.; Li, B.; Xia, T. Multi-view 3d object detection network for autonomous driving. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July2017; pp. 6526–6534.
6. Yu, T.; Meng, J.; Yuan, J. Multi-view harmonized bilinear network for 3d object recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 186–194.
7. Yang, Z.; Wang, L. Learning relationships for multi-view 3D object recognition. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 7504–7513.
8. Maturana, D.; Scherer, S. Voxnet: A 3d convolutional neural network for real-time object recognition. In Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Boston, MA, USA, 7–12 June 2015; pp. 3367–3375.
9. Riegler, G.; Osman Ulusoy, A.; Geiger, A. Octnet: Learning deep 3d representations at high resolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 6620–6629.
10. Wang, P.S.; Liu, Y.; Guo, Y.X.; Sun, C.Y.; Tong, X. O-cnn: Octree-based convolutional neural networks for 3d shape analysis. *ACM Trans. Graph. (TOG)* **2017**, *36*, 1–11. [CrossRef]
11. Le, T.; Duan, Y. Pointgrid: A deep network for 3d shape understanding. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 9204–9214.
12. Qi, C.R.; Su, H.; Mo, K.; Guibas, L.J. Pointnet: Deep learning on point sets for 3d classification and segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, UT, USA, 21–26 July 2017; pp. 77–85.
13. Qi, C.R.; Yi, L.; Su, H.; Guibas, L.J. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *arXiv* **2017**, arXiv:1706.02413.
14. Duan, Y.; Zheng, Y.; Lu, J.; Zhou, J.; Tian, Q. Structural relational reasoning of point clouds. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 949–958.
15. Yin, K.; Huang, H.; Cohen-Or, D.; Zhang, H. P2p-net: Bidirectional point displacement net for shape transform. *ACM Trans. Graph. (TOG)* **2018**, *37*, 1–13. [CrossRef]
16. Yang, J.; Zhang, Q.; Ni, B.; Li, L.; Liu, J.; Zhou, M.; Tian, Q. Modeling point clouds with self-attention and gumbel subset sampling. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3318–3327.
17. Sarode, V.; Li, X.; Goforth, H.; Aoki, Y.; Srivatsan, R.A.; Lucey, S.; Choset, H. PCRNet: Point cloud registration network using PointNet encoding. *arXiv* **2019**, arXiv:1908.07906.
18. Lin, Z.; Feng, M.; Santos, C.N.d.; Yu, M.; Xiang, B.; Zhou, B.; Bengio, Y. A structured self-attentive sentence embedding. *arXiv* **2017**, arXiv:1703.03130.
19. Thabet, A.; Alwassel, H.; Ghanem, B. Mortonnet: Self-supervised learning of local features in 3D point clouds. *arXiv* **2019**, arXiv:1904.00230.
20. Wu, Y.; He, F.; Yang, Y. A grid-based secure product data exchange for cloud-based collaborative design. *Int. J. Coop. Inf. Syst.* **2020**, *29*, 2040006. [CrossRef]
21. 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, Venice, Italy, 22–29 October 2017; pp. 863–872.
22. Dai, A.; Chang, A.X.; Savva, M.; Halber, M.; Funkhouser, T.; Nießner, M. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2432–2443.
23. Wu, Z.; Song, S.; Khosla, A.; Yu, F.; Zhang, L.; Tang, X.; Xiao, J. 3d shapenets: A deep representation for volumetric shapes. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1912–1920.
24. Johnson, J.; Hariharan, B.; Van Der Maaten, L.; Hoffman, J.; Li, F.-F.; Lawrence Zitnick, C.; Girshick, R. Inferring and executing programs for visual reasoning. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 3008–3017.
25. Li, Y.; Pirk, S.; Su, H.; Qi, C.R.; Guibas, L.J. Fpnn: Field probing neural networks for 3d data. *Adv. Neural Inf. Process. Syst.* **2016**, *29*, 307–315.
26. Wang, D.Z.; Posner, I. Voting for voting in online point cloud object detection. In *Robotics: Science and Systems*; Sapienza University of Rome: Rome, Italy, 2015; Volume 1, pp. 10–15.

27. Sun, X.; Lian, Z.; Xiao, J. Srinet: Learning strictly rotation-invariant representations for point cloud classification and segmentation. In Proceedings of the 27th ACM International Conference on Multimedia, Nice France, 21–25 October 2019; pp. 980–988.
28. Joseph-Rivlin, M.; Zvirin, A.; Kimmel, R. Momen (e) t: Flavor the moments in learning to classify shapes. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, Seoul, Korea, 27–28 October 2019; pp. 4085–4094.
29. Achlioptas, P.; Diamanti, O.; Mitliagkas, I.; Guibas, L. Learning representations and generative models for 3d point clouds. In Proceedings of the International Conference on Machine Learning, PMLR, Stockholmsmässan, Stockholm, Sweden, 10–15 July 2018; Volume 80, pp. 40–49.
30. Lin, H.; Xiao, Z.; Tan, Y.; Chao, H.; Ding, S. Justlookup: One millisecond deep feature extraction for point clouds by lookup tables. In Proceedings of the 2019 IEEE International Conference on Multimedia and Expo (ICME), Shanghai, China, 8–12 July 2019; pp. 326–331.
31. Zhang, D.; He, F.; Tu, Z.; Zou, L.; Chen, Y. Pointwise geometric and semantic learning network on 3D point clouds. *Integr. Comput.-Aided Eng.* **2020**, *27*, 57–75. [CrossRef]
32. Wang, Y.; Sun, Y.; Liu, Z.; Sarma, S.E.; Bronstein, M.M.; Solomon, J.M. Dynamic graph cnn for learning on point clouds. *Acm Trans. Graph. (TOG)* **2019**, *38*, 1–12. [CrossRef]
33. Guo, M.H.; Cai, J.X.; Liu, Z.N.; Mu, T.J.; Martin, R.R.; Hu, S.M. PCT: Point cloud transformer. *Comput. Vis. Media* **2021**, *7*, 187–199. [CrossRef]
34. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
35. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European conference on computer vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
36. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 5998–6008.
37. Li, X.; Yu, L.; Fu, C.W.; Cohen-Or, D.; Heng, P.A. Unsupervised detection of distinctive regions on 3D shapes. *ACM Trans. Graph. (TOG)* **2020**, *39*, 1–14. [CrossRef]
38. Zhao, H.; Jiang, L.; Jia, J.; Torr, P.H.; Koltun, V. Point transformer. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 11–17 October 2021; pp. 16259–16268.
39. Shaw, P.; Uszkoreit, J.; Vaswani, A. Self-attention with relative position representations. *arXiv* **2018**, arXiv:1803.02155.
40. Phan, A.V.; Le Nguyen, M.; Nguyen, Y.L.H.; Bui, L.T. Dgcnn: A convolutional neural network over large-scale labeled graphs. *Neural Netw.* **2018**, *108*, 533–543. [CrossRef]
41. Li, Y.; Bu, R.; Sun, M.; Wu, W.; Di, X.; Chen, B. Pointcnn: Convolution on x-transformed points. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 820–830.
42. Liu, Y.; Fan, B.; Xiang, S.; Pan, C. Relation-shape convolutional neural network for point cloud analysis. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 8887–8896.
43. Xu, M.; Ding, R.; Zhao, H.; Qi, X. PAConv: Position Adaptive Convolution with Dynamic Kernel Assembling on Point Clouds. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 3172–3181.
44. Muzahid, A.; Wan, W.; Sohel, F.; Wu, L.; Hou, L. Curvenet: Curvature-based multitask learning deep networks for 3D object recognition. *IEEE/CAA J. Autom. Sin.* **2020**, *8*, 1177–1187. [CrossRef]
45. Ran, H.; Zhuo, W.; Liu, J.; Lu, L. Learning Inner-Group Relations on Point Clouds. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 11–17 October 2021; pp. 15477–15487.
46. 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), Munich, Germany, 8–14 September 2018; pp. 87–102.
47. Komarichev, A.; Zhong, Z.; Hua, J. A-cnn: Annularly convolutional neural networks on point clouds. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 7413–7422
48. Yan, X.; Zheng, C.; Li, Z.; Wang, S.; Cui, S. Pointasnl: Robust point clouds processing using nonlocal neural networks with adaptive sampling. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 5588–5597.
49. Li, J.; Chen, B.M.; Lee, G.H. So-net: Self-organizing network for point cloud analysis. In Proceedings of the IEEE conference on computer vision and pattern recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 9397–9406.
50. Chang, A.X.; Funkhouser, T.; Guibas, L.; Hanrahan, P.; Huang, Q.; Li, Z.; Savarese, S.; Savva, M.; Song, S.; Su, H.; et al. Shapenet: An information-rich 3d model repository. *arXiv* **2015**, arXiv:1512.03012.
51. Atzmon, M.; Maron, H.; Lipman, Y. Point convolutional neural networks by extension operators. *arXiv* **2018**, arXiv:1803.10091.
52. Van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
53. Zhang, D.; Zhang, Z.; Zou, L.; Xie, Z.; He, F.; Wu, Y.; Tu, Z. Part-based visual tracking with spatially regularized correlation filters. *Vis. Comput.* **2020**, *36*, 509–527. [CrossRef]

54. Zhang, D.; Wu, Y.; Guo, M.; Chen, Y. Deep Learning Methods for 3D Human Pose Estimation under Different Supervision Paradigms: A Survey. *Electronics* **2021**, *10*, 2267. [CrossRef]
55. Wu, Y.; Ma, S.; Zhang, D.; Sun, J. 3D Capsule Hand Pose Estimation Network Based on Structural Relationship Information. *Symmetry* **2020**, *12*, 1636. [CrossRef]