Computer Aided Geometric Design 79 (2020) 101859

Contents lists available at ScienceDirect

Computer Aided Geometric Design

www.elsevier.com/locate/cagd

LRC-Net: Learning discriminative features on point clouds by encoding local region contexts

Xinhai Liu [a], Zhizhong Han [b], Fangzhou Hong [a], Yu-Shen Liu [a,*], Matthias Zwicker [b]

[a] School of Software, BNRist, Tsinghua University, Beijing, China
[b] Department of Computer Science, University of Maryland, College Park, USA

ARTICLE INFO

Article history:
Available online 21 April 2020

Keywords:
Point clouds
Local region
Geometric information
Context

ABSTRACT

Learning discriminative feature directly on point clouds is still challenging in the understanding of 3D shapes. Recent methods usually partition point clouds into local region sets, and then extract the local region features with fixed-size CNN or MLP, and finally aggregate all individual local features into a global feature using simple max pooling. However, due to the irregularity and sparsity in sampled point clouds, it is hard to encode the fine-grained geometry of local regions and their spatial relationships when only using the fixed-size filters and individual local feature integration, which limit the ability to learn discriminative features. To address this issue, we present a novel Local-Region-Context Network (LRC-Net), to learn discriminative features on point clouds by encoding the fine-grained contexts inside and among local regions simultaneously. LRC-Net consists of two main modules. The first module, named *intra-region context encoding*, is designed for capturing the geometric correlation inside each local region by novel variable-size convolution filter. The second module, named *inter-region context encoding*, is proposed for integrating the spatial relationships among local regions based on spatial similarity measures. Experimental results show that LRC-Net is competitive with state-of-the-art methods in shape classification and shape segmentation applications.

1. Introduction

As an important type of 3D data which can be acquired conveniently by various 3D sensors, point cloud has been increasingly used in diverse real word applications including autonomous driving (Qi et al., 2018; Yi et al., 2019), 3D modeling (Golovinskiy et al., 2009; Gao et al., 2017; Han et al., 2017a, 2019b; Zhong et al., 2019; Skrodzki et al., 2018; Zheng et al., 2018; Gao et al., 2015), indoor navigation (Zhu et al., 2017) and robotics (Rusu et al., 2008). Therefore, there is an emerging demand to learn discriminative features with deep neural networks for 3D shape understanding.

Unlike images, point cloud is not suitable for the traditional convolutional neural network (CNN) which often requires some fixed spatial distribution in the neighborhood of each pixel. To alleviate this issue, an alternative way is to rasterize the point cloud into regular voxel representations and then apply 3D CNNs (Zhou and Tuzel, 2017). However, the performance of plain 3D CNNs is largely limited by the serious resolution loss and the fast-growing computational cost, due to the inherent

* Corresponding author.
E-mail addresses: lxh17@mails.tsinghua.edu.cn (X. Liu), h312h@umd.edu (Z. Han), hongfz16@mails.tsinghua.edu.cn (F. Hong), liuyushen@tsinghua.edu.cn (Y.-S. Liu), zwicker@cs.umd.edu (M. Zwicker).

https://doi.org/10.1016/j.cagd.2020.101859
0167-8396/

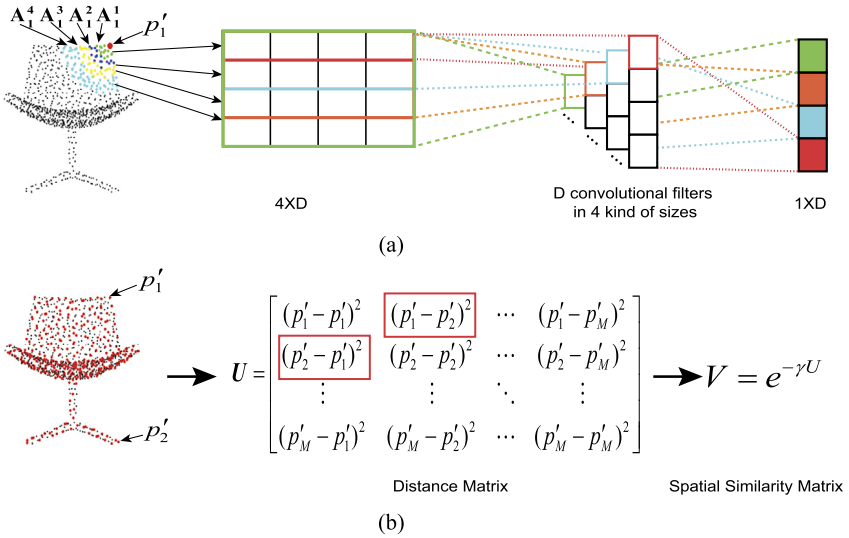

Fig. 1. (a) *Intra-region context encoding*. It shows the process of capturing the geometric correlation inside the local region around the point p'_1. (b) *Inter-region context encoding*. The spatial similarities of local regions are measured with the matrix $\boldsymbol{V}$, where two central points p'_1 and p'_2 of local regions are highlighted.

sparsity of 3D shapes. To overcome the shortcoming of 3D CNNs, PointNet (Qi et al., 2017a) was proposed as a pioneering work which directly learns global features for 3D shapes from point sets. However, PointNet learns the feature of each point individually, while omitting the important contextual information among points.

To solve above-mentioned problems, recent studies have attempted to encode the local region contexts of point clouds with various designed manners. Specifically, there are two kinds of local region contexts, including the intra-region geometric context and the inter-region spatial context. On the one hand, some methods concentrate on capturing the context of geometric correlations inside each local region. For example, PointNet++ (Qi et al., 2017b) uses a sampling and grouping strategy to hierarchically extract features for local regions. More recently, Point2Sequence (Liu et al., 2019a) learns the contextual information inside a local region with an attention-based sequence to sequence network. On the other hand, several studies attempt to utilize the context of spatial distribution information among local regions. For example, KD-Net (Klokov and Lempitsky, 2017) builds a kd-tree to divide the point cloud into small leaf bins and then hierarchically extracts the point cloud feature from the leaves to root according to a fixed spatial partition. KC-Net (Shen et al., 2018) uses a graph pooling operation which can partially utilize the spatial distribution information among local regions. However, it is still hard for these methods to encode the fine-grained contexts inside and among local regions simultaneously, especially for the geometric correlation between different scale areas inside a local region and the spatial relationships among local regions. This motivates us to employ variable-size filters inside each local region and spatial similarity measures among local regions for capturing intra-region context information and inter-region context information, respectively. Our method reliefs the limitation of the traditional CNNs in encoding the geometric context information on point clouds, which usually implements a convolution layer with fixed-size filters, while the concrete filter size is a hyper-parameter. To address above-mentioned problems, we propose LRC-Net to learn discriminative features from point clouds.

Our key contributions are summarized as follows.

- LRC-Net is presented for learning discriminative features directly from point clouds by simultaneously encoding the geometric correlation inside each local region and the spatial relationships among local regions.
- *Intra-region context encoding* module is designed for capturing the geometric correlation inside each local region by novel variable-size convolution filters, which learns the intrinsic structure and correction of multi-scale areas from their feature maps, rather than simple feature concatenating or max pooling as usually used in previous methods such as Qi et al. (2017a).
- *Inter-region context encoding* module is proposed for integrating the spatial relationships among local regions based on spatial similarity measures, which encodes the spatial distribution of local regions in their metric space.

The above two modules are illustrated in Fig. 1.

2. Related work

Feature learning from regularized 3D data. Traditional methods (Liu and Ramani, 2009; Liu et al., 2009, 2011; Gao et al., 2015; Fehr et al., 2016; Zou et al., 2018; Srivastava and Lall, 2019; Beksi and Papanikolopoulos, 2019; Zhao et al., 2020) focus on capturing the geometric information of 3D shapes, which are usually limited by the hand-crafted manner

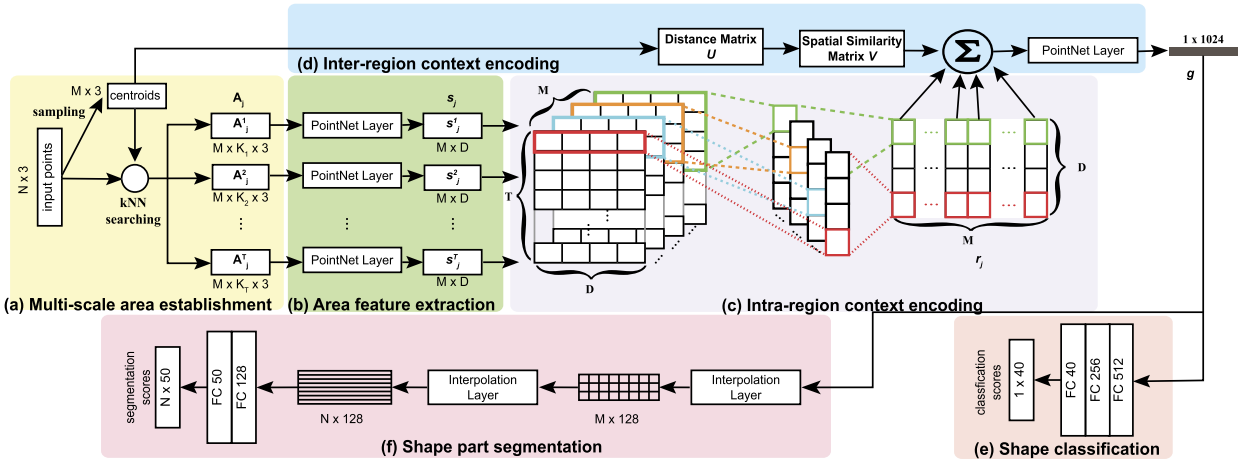

Fig. 2. Our LRC-Net architecture. In (a), LRC-Net first establishes multi-scale areas inside each local region by sampling and searching layers. Then, PointNet layer is employed to extract the feature of each scale area in (b). Subsequently, the feature of each local region is extracted for intra-region context encoding in (c), which captures the geometric correlation inside the local region. Simultaneously, the spatial similarity measure is calculated for inter-region context encoding in (d), which enhances the spatial relationships among local regions. Finally, the global feature of the point cloud is obtained by aggregating the features of local regions. The learned global feature can be used in shape classification and shape segmentation applications as shown in (e)(f).

in specific application. Benefit from the success of CNNs on large-scale image repositories such as ImageNet (Krizhevsky et al., 2012), deep neural networks are being applied to process the 3D format data. As an irregular format of 3D data, point clouds can be transformed into other kinds of regularized format, such as the 3D voxel (Han et al., 2016, 2017b, 2018a) or the rendered view (Han et al., 2018b, 2019c,e,d,g,a). The voxelization of point cloud is a feasible choice, which first converts the point cloud into voxels, and then applies 3D CNNs. 3D ShapeNets (Wu et al., 2015) and VoxNet (Maturana and Scherer, 2015) represent each voxel with a binary value which indicates the occupied of the location in the space. However, the performance is largely limited by the resolution loss and the rapid growth of computational complexity. The inherent sparsity of 3D shapes makes it hard to make full use of the storage of input data, where the hollow inside 3D shapes is often meaningless. Some improvements (Li et al., 2016) have been proposed to alleviate the data sparsity of the volumetric representation. However, it is still nontrivial to deal with large point clouds with high resolution.

Feature learning from point clouds. PointNet (Qi et al., 2017a) is a pioneering work which directly adopts point sets as input and obtains convincing performances. A concise strategy is adopted in PointNet by computing the feature for each point individually and then aggregating these features into a global representation with max-pooling. However, PointNet is largely limited in capturing the contextual information of local regions. To address this problem, many recent studies attempt to capture local region contexts. Specifically, local region contexts can be divided into two categories, which are *intra-region context* and *inter-region context*, respectively. On the one hand, some studies capture the intra-region context by building graph inside multi-scale local regions. PointNet++ (Qi et al., 2017b) uses sampling and grouping operations for extracting features from several clusters hierarchically to capture the context of each cluster. Point2Sequence (Liu et al., 2019a) extracts the feature of local regions by a sequence to sequence model with an attention mechanism. On the other hand, some studies (Li et al., 2018b; Wang et al., 2018; Xu et al., 2018; Wang et al., 2017; Komarichev et al., 2019; Hu et al., 2019; Wen et al., 2020) investigate CNN-like operations to aggregate neighbors of a given point by building kNN graph inside the single-scale local region. On the other hand, some studies encode the inter-region context with indexing structures. KC-Net (Shen et al., 2018) employs a kernel correlation layer and a graph pooling layer for capturing the local structure of point clouds. ShapeContextNet (Xie et al., 2018) extends the 2D Shape context (Belongie et al., 2001) to 3D, which divides the local region of a given point into bins and updates the point feature with the aggregation of these bin features. KD-Net (Klokov and Lempitsky, 2017) and OctNet (Riegler et al., 2017) first divide the input point cloud into leaves, and then hierarchically extracts features from leaves to the root. Point2SpatialCapsule (Wen et al., 2019) integrates capsules to explore the local structures of point clouds, which employs a multi-scale shuffling to increase the diversity of local region features and applies a clustering operation to capture the spatial information of local regions in the feature space. This complicated procedure significantly differentiates Point2SpatialCapsule from ours. In general, it is hard for current methods to simultaneously capture the contextual information inside and among multi-scale local regions, which limits the expressiveness of learned representations of point clouds.

3. The LRC-Net model

Fig. 2 shows the architecture of LRC-Net, which is composed of six parts: multi-scale area establishment, area feature extraction, intra-region context encoding, inter-region context encoding, shape classification and shape segmentation, respectively. LRC-Net adopts a point cloud $\mathbf{P} = \{p_i \in \mathbb{R}^3, i = 1, 2, \cdots, N\}$ as input which is composed of 3D point coordinates x, y and z. Firstly, a subset with M points, denoted by $\mathbf{P}' = \{p'_j \in \mathbb{R}^3, j = 1, 2, \cdots, M\}$, is selected from the input point

cloud $\mathbf{P}$ to act as the centroids of local regions $\{\mathbf{R}_j, j = 1, 2, \cdots, M\}$. Based on the selected centeroids $\mathbf{P}'$, T different scale areas $\mathbf{A}_j = \{\mathbf{A}_j^t, t = 1, 2, \cdots, T\}$ are established in each local region $\mathbf{R}_j$ centered at p'_j, where $\{K_t, t = 1, 2, \cdots, T\}$ points are contained in each scale area, respectively. Then, a D-dimensional feature $\boldsymbol{s}_j^t$ is extracted from each scale area $\mathbf{A}_j^t$. By stacking $\boldsymbol{s}_j^t$, a $T \times D$ feature matrix $\boldsymbol{s}_j = \{\boldsymbol{s}_j^t, t = 1, 2, \cdots, T\}$ is formed for each local region $\mathbf{R}_j$, and further aggregated into a D-dimensional feature $\boldsymbol{r}_j$ by the intra-region encoding module (see Fig. 2(c)). Meanwhile, another module of calculating the spatial similarity in the 3D space is applied to capture the inter-region context among local regions (see Fig. 2(d)). Finally, a 1024-dimensional feature $\boldsymbol{g}$ of the whole input point cloud $\mathbf{P}$ is aggregated from the feature of M local regions, which integrates the extracted intra-region and inter-region context features. The learned global feature $\boldsymbol{g}$ can be applied to shape classification and shape segmentation applications.

3.1. Multi-scale area establishment

Three key layers are engaged in our structure to establish the multi-scale areas around each sampled point, including sampling layer, searching layer and grouping layer. The sampling layer uniformly selects M points from the input point cloud $\mathbf{P}$ as the centroids of local regions. Around each sampled centroid, the searching layer continuously finds $[K_1, \cdots, K_t, \cdots, K_T]$ nearest points to build the indexing relationship between points. According to the indexes in the searching layer, the grouping layer groups multi-scale areas $\{\mathbf{A}_j^t, t = 1, 2, \cdots, T\}$ inside each local region $\mathbf{R}_j$.

In the sampling layer, farthest point sampling (FPS) is adopted to select $M (M < N)$ points $\mathbf{P}'$ which defines the centroids of local regions. In the sampling process, the new sampled point p'_j is always the farthest one from previously selected points $\{p'_1, p'_2, \cdots, p'_{j-1}\}$. Compared with other sampling methods, such as random sampling, FPS can achieve a more uniform coverage of the entire point cloud with the same number of sampled points.

To build the multi-scale areas $\mathbf{A}_j$, the k-nearest neighbors (kNN) algorithm is applied to search the neighbors of a given point based on the Euclidean distance between points. Another alternative method is the ball query (Qi et al., 2017b) which selects all points within a given radius around a point. Compared with the ball query, kNN can guarantee the information inside local regions and is robust to the input point cloud with different sparsity.

3.2. Area feature extraction

As shown in Fig. 2, a concise and effective PointNet layer is employed in LRC-Net to extract the feature for each scale area. The PointNet layer is composed of two key parts: a Multi-Layer-Perceptron (MLP) layer and a max-pooling layer, respectively. The MLP layer individually abstracts the coordinates of points in each area $\mathbf{A}_j^t$ into the feature space, and then these features are aggregated into a D-dimensional feature $\boldsymbol{s}_j^t$ by the max pooling layer. So far, a feature map of T different scale areas $\{\mathbf{A}_j, j = 1, 2, \cdots, M\}$ with the size of $M \times T \times D$ is acquired after the PointNet layer.

Following previous studies (Li et al., 2018a; Qi et al., 2017b), the relative coordinates are adopted in LRC-Net. Before feeding points inside each local region $\mathbf{R}_j$ into the PointNet layer, a relative coordinate system of the centroid p'_j is built by a simple operation: $p_l = p_l - p'_j$, where l is the index of points in the local region $\mathbf{R}_j$. Different from absolute coordinates, the relative coordinates are determined by the relative positional relationship between points. Therefore, by using relative coordinates, the learned feature of local regions can be invariant to transformations such as rotation and translation.

3.3. Intra-region context encoding

In order to capture the fine-grained contextual information between multi-scale areas inside local regions, variable-size convolution filters are employed in the architecture. Inspired by capturing the correlation of different words in the natural language processing tasks (Kim, 2014), the intra-region correlation of multi-scale areas is also important in the feature learning of point clouds. Different from most existing methods that only encode the correlation of fixed scale of areas, we consider capturing the correlation among multiple scales from 1 to T. As depicted in Fig. 2, given the features $\{\boldsymbol{s}_j^t, t = 1, 2, \cdots, T\}$ of multi-scale areas in a local region $\mathbf{R}_j$ from the area feature extraction module, we first represent these features in a $T \times D$ feature map by

$$\boldsymbol{S}_j^{1:T} = \boldsymbol{s}_j^1 \oplus \boldsymbol{s}_j^2 \oplus \cdots \oplus \boldsymbol{s}_j^T, \tag{1}$$

where $\oplus$ is the concatenation operator. In general, let $\boldsymbol{S}_j^{a:a+b}$ refer to the concatenation of features $\boldsymbol{s}_j^a, \boldsymbol{s}_j^{a+1}, \cdots, \boldsymbol{s}_j^{a+b-1}$. A convolution operation involves a filter $\boldsymbol{w} \in \mathbb{R}^{hD}$, which is applied to a window of h scale features to produce a new feature. For example, a feature c_k is generated from a window of features $\boldsymbol{S}_j^{a:a+h-1}$ by

$$c_k = f(\boldsymbol{w} \cdot \boldsymbol{S}_j^{a:a+h-1} + b). \tag{2}$$

Here $b \in \mathbb{R}$ is a bias term and f is a non-linear function such as ReLU (Nair and Hinton, 2010). As the intermediate step shown in Fig. 2, this filter is applied to each possible window of scales in the features $\boldsymbol{S}_j^{1:h}, \boldsymbol{S}_j^{2:h+1}, \cdots, \boldsymbol{S}_j^{T-h+1:T}$ to produce a feature vector

$$\boldsymbol{c} = [c_1, c_2, \cdots, c_{T-h+1}], \tag{3}$$

with length of $T-h+1$. Then, we apply max pooling operation over the feature vector, which extracts the maximum value from feature vector $\boldsymbol{c}$ by

$$\hat{\boldsymbol{c}} = max\{\boldsymbol{c}\}. \tag{4}$$

Here $\hat{\boldsymbol{c}}$ is one element of local region feature $\boldsymbol{r}_j$ corresponding to this particular filter. So far, we have shown the process of getting one element in $\boldsymbol{r}_j$ by a convolution filter with window size $h \times D$. In general, there are T kinds of convolution filters in different sizes and $\frac{D}{T}$ filters for each kind of convolution filter. Therefore, the output of each input local region $\mathbf{R}_j$ is a D-dimensional feature vector $\boldsymbol{r}_j$.

3.4. Inter-region context encoding

To obtain the global feature of point clouds, most existing methods adopt simple pooling layers to aggregate local region features. However, the inter-region context is largely lost in the pooling process, especially for the spatial distribution information among local regions. To capture the inter-region spatial context, a greedy strategy is proposed by aggregating the spatial distribution information among local regions in an explicit manner. Following the intra-region context encoding module, the feature map with the size $M \times D$ of M local regions is obtained. As shown in Fig. 2, to encode the spatial information of local regions, we explicitly calculate the spatial similarity among local regions based on the coordinates of local region centroids. Given the coordinate of centroids $\{p'_j, j = 1, 2, \cdots, M\}$, the $M \times M$ distance matrix $\boldsymbol{U}$ is build by

$$\boldsymbol{U} = \begin{bmatrix} (p'_1 - p'_1)^2 & (p'_1 - p'_2)^2 & \cdots & (p'_1 - p'_M)^2 \\ (p'_2 - p'_1)^2 & (p'_2 - p'_2)^2 & \cdots & (p'_2 - p'_M)^2 \\ \vdots & \vdots & \ddots & \vdots \\ (p'_M - p'_1)^2 & (p'_M - p'_2)^2 & \cdots & (p'_M - p'_M)^2 \end{bmatrix}. \tag{5}$$

To convert the distance matrix to the similarity space, the spatial similarity matrix $\boldsymbol{V}$ is calculated by

$$\boldsymbol{V} = e^{-\gamma \boldsymbol{U}}. \tag{6}$$

Here γ is a parameter which can regulate the effect of the spatial similarity. Thus, we obtain the spatial similarity among local regions. To enhance the feature $\boldsymbol{r}_j$ of each local region, a greedy weighting strategy is adopted as

$$\boldsymbol{r}'_j = \sum_{b=1}^{M} \boldsymbol{V}_{j,b} \cdot \boldsymbol{r}_b, \tag{7}$$

where $\boldsymbol{r}'_j$ is the enhanced feature vector of $\boldsymbol{r}_j$ and b is the index of the column. In addition, a normalization operation is applied to the enhanced features by

$$\boldsymbol{r}''_j = \frac{\boldsymbol{r}'_j}{\sum_{b=1}^{M} \boldsymbol{V}_{j,b}}. \tag{8}$$

Here $\boldsymbol{r}''_j$ is the final features of local regions after the regularization, which contains the information of spatial distribution among local regions. In general, it is a greedy strategy to compute the spatial similarity between any two local regions. The greedy strategy aims to enhance the correlation of local regions, which can promote the learning of the global features. In the subsequent network, a 1024-dimensional global feature $\boldsymbol{g}$ of the input point cloud is extracted by another PointNet layer. The learned global feature $\boldsymbol{g}$ can be applied to various applications, such as shape classification and shape segmentation.

3.5. Expansion for shape segmentation

The target of shape segmentation is to predict a semantic label for each point in the point cloud. With the obtained global feature $\boldsymbol{g}$, the key is how to acquire the feature for each point. There are two options, one is to duplicate the global feature with N times as in Wang et al. (2018), the other is to perform upsampling by the interpolation layer (Qi et al., 2017b). In the shape segmentation module, two interpolation layers are equipped in our network, which propagate the features from shape level to point level by upsampling. The feature propagation ϕ between different levels is guided by the inverse distance between k-nearest points. In the interpolation layer, we search k ($k = 3$) nearest points for each point in current level from points in previous level. Therefore, the feature of point $\phi(p)$ in current level is interpolated by the positional relationship of points between two levels, denoted by

Table 1
The shape classification results (%) on ModelNet10 and ModelNet40 benchmarks.

Method	Scales	Points	MN10	MN40
PointNet (Qi et al., 2017a)	single	1k	-	89.2
O-CNN (Wang et al., 2017)	single	-	-	90.6
MAP-VAE (Han et al., 2019f)	single	1k	94.82	90.15
Kd-Net (Klokov and Lempitsky, 2017)	single	1k	94.0	91.8
KC-Net (Shen et al., 2018)	single	1k	94.4	91.0
PointCNN (Li et al., 2018b)	single	1k	-	91.7
DGCNN (Wang et al., 2018)	single	1k	-	92.2
SO-Net (Li et al., 2018a)	single	2k	94.1	90.9
A-CNN (Komarichev et al., 2019)	single	1k	95.5	92.6
InterpCNN (Mao et al., 2019)	single	1k	-	93.0
RS-CNN (Liu et al., 2019c)	single	1k	-	93.6
PointNet++ (Qi et al., 2017b)	multi	1k	-	90.7
L2G-AE (Liu et al., 2019b)	multi	1k	95.37	90.64
ShapeContextNet (Xie et al., 2018)	multi	1k	-	90.0
Point2Sequence (Liu et al., 2019a)	multi	1k	95.3	92.6
Point2SpatialCapsule (Wen et al., 2019)	multi	1k	**95.8**	93.4
LRC-Net (ours)	multi	10k	-	**94.2**
LRC-Net (ours)	multi	1k	**95.8**	93.1

Table 2
The shape segmentation results (%) on ShapeNet part segmentation dataset.

Method	Scale	Mean	Intersection over Union (IoU)															
			air.	bag	cap	car	cha.	ear.	gui.	kni.	lam.	lap.	mot.	mug	pis.	roc.	ska.	tab.
# SHAPES			2690	76	55	898	3758	69	787	392	1547	451	202	184	283	66	152	5271
PointNet (Qi et al., 2017a)	single	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
Kd-Net (Klokov and Lempitsky, 2017)	single	82.3	80.1	74.6	74.3	70.3	88.6	73.5	90.2	87.2	81.0	94.9	57.4	86.7	78.1	51.8	69.9	80.3
KCNet (Shen et al., 2018)	single	84.7	82.8	81.5	86.4	77.6	90.3	76.8	91.0	87.2	84.5	95.5	69.2	94.4	81.6	60.1	75.2	81.3
DGCNN (Wang et al., 2018)	single	85.1	**84.2**	83.7	84.4	77.1	90.9	78.5	91.5	87.3	82.9	96.0	67.8	93.3	82.6	59.7	75.5	82.0
SO-Net (Li et al., 2018a)	single	84.9	82.8	77.8	88.0	77.3	90.6	73.5	90.7	83.9	82.8	94.8	69.1	94.2	80.9	53.1	72.9	83.0
A-CNN (Komarichev et al., 2019)	single	86.1	84.2	84.0	88.0	79.6	**91.3**	75.2	91.6	87.1	85.5	95.4	75.3	94.9	82.5	**67.8**	77.5	83.3
PointCNN (Li et al., 2018b)	single	86.1	84.1	**86.5**	86.0	**80.8**	90.6	79.7	**92.3**	**88.4**	85.3	**96.1**	**77.2**	**95.3**	**84.2**	64.2	**80.0**	83.0
RS-CNN (Liu et al., 2019c)	single	**86.2**	83.5	84.8	**88.8**	79.6	91.2	**81.1**	91.6	**88.4**	**86.0**	96.0	73.7	94.1	83.4	60.5	77.7	**83.6**
PointNet++ (Qi et al., 2017b)	multi	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
ShapeContextNet (Xie et al., 2018)	multi	84.6	**83.8**	80.8	83.5	**79.3**	90.5	69.8	**91.7**	86.5	82.9	**96.0**	69.2	93.8	82.5	**62.9**	74.4	80.8
Point2Sequence (Liu et al., 2019a)	multi	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**
Point2SpatialCapsule (Wen et al., 2019)	multi	**85.3**	83.5	83.4	**88.5**	77.6	**90.8**	79.4	90.9	**86.9**	84.3	95.4	**71.7**	**95.3**	**82.6**	60.6	75.3	82.5
LRC-Net (ours)	multi	**85.3**	82.6	**85.2**	87.4	79.0	90.7	**80.2**	91.3	**86.9**	**84.5**	95.5	71.4	93.8	79.4	51.7	75.5	82.6

$$\phi(p) = \frac{\sum_{i=1}^{k} w(p_i)\phi(p_i)}{\sum_{i=1}^{k} w(p_i)}, \tag{9}$$

where $w(p_i) = \frac{1}{(p-p_i)^2}$ is the inverse square Euclidean distance between two points, $\phi(p_i)$ is the point feature of p_i and $\{p_i, i = 1, 2, \cdots, k\}$ are the k nearest points of p in the previous level. The points in each level are already obtained from the multi-scale area establishment module and the interpolation step can be regard as a reverse process of the abstraction step.

4. Experiments

In this section, shape classification and shape segmentation applications are adopted to evaluate the performances of the LRC-Net. In the ablation study, we first investigate how the two main modules affect the performances of LRC-Net in the shape classification task on ModelNet40 (Wu et al., 2015). Then, we compare our model with several state-of-the-art methods in shape classification on ModelNet10/40 and shape part segmentation on ShapeNet part dataset (Savva et al., 2016). Finally, some visualizations of the shape segmentation results are also reported.

4.1. Network configuration

In LRC-Net, some network configurations need to be initialized. According to the input point cloud, we first initialize parameters of the number of sampled points $M = 384$, the number of scales $T = 4$, the number of points in multi-scale areas $K_1 = 16, K_2 = 32, K_3 = 64$ and $K_4 = 128$, the feature dimension $\boldsymbol{r}_j$ of each local region $D = 128$. The rest settings of our model are same as in Fig. 2. In addition, ReLU is used after each fully connected layer with Batch-normalization, and Dropout is also applied with drop ratio 0.4 in the fully connected layers. In the experiment, we train our network on a NVIDIA GTX 1,080Ti GPU using ADAM optimizer with initial learning rate 0.001, batch size of 16 and batch normalization rate 0.5. The learning rate and batch normalization rate are decreased by 0.3 and 0.5 for every 20 epochs, respectively.

5. Parameters setting

All the experiments in this section are evaluated on ModelNet40, which contains 40 categories and 12,311 CAD shapes with 9,843 shapes for training and 2,468 shapes for testing. And the results listed in tables are the instance accuracies. For each 3D shape, we adopt the point cloud with 1,024 points which are uniformly sampled from the corresponding mesh faces as input.

Table 3
The effect of the convergence factor γ on ModelNet40.

γ	0	1	10^2	10^4	10^5
Accuracy (%)	91.33	91.61	92.02	**93.07**	92.54

In the module of spatial distribution information encoding, γ is an important parameter which influences the performance of the whole model. The results of several settings of γ are shown in the Table 3. The best instance accuracy 93.07% is reached at $\gamma = 10^4$ which maximizes the effect of the spatial information encoding module. In particular, $\gamma = 0$ represents a simple summation of the local region features, which will result in the discriminative ability loss of local region features. From the results, we can see that the spatial information encoding module can promote the global representation learning of point cloud.

Table 4
The effect of the sampled points M on ModelNet40.

M	128	256	384	512
Accuracy (%)	92.22	92.34	**93.07**	92.42

To explore the effect of the sampled points M, we keep the setting $\gamma = 10^4$ and vary M from 128 to 512 as shown in Table 4. The number of the sample points influences the local regions which are visible to the network in the training process. $M = 384$ can obtain a better coverage of all training point clouds, where the input information is balanced between insufficiency and redundancy.

Table 5
The effect of the number of scale areas S in each local region on ModelNet40.

T	1	2	3	4	5
Accuracy (%)	92.26	92.42	92.54	**93.07**	92.18

Moreover, we also discuss the impact of the number of scale areas T in each local region. In the implementation, we keep the number of points 128 in each local region and range T from 1 to 5. The number of points in each scale is a power of 2 and varies in $[8, 16, 32, 64, 128]$. Specifically, $T = 3$ indicates that there are $[32, 64, 128]$ points in the scale areas. And similarly, $T = 2$ represents there are $[64, 128]$ points in the two scale areas respectively. In terms of results in Table 5, LRC-Net reaches the best performance when the scale areas number is 4. In practice, the number of scale areas is largely determined by the properties of the input point cloud, especially the sparsity of points. Therefore, when the number of scale areas in each local region is 4, it is more suitable for our model.

Table 6
The effect of the kind of filers h on ModelNet40 in the variable-size convolution module.

h	1	2	3	4
Accuracy (%)	92.38	92.42	92.67	**93.07**

With the number of scale areas to be 4, we change the kind of filters from 1 to 4 in the variable-size convolution module. In Table 6, $h = 1$ represents only one type of convolutional filter $1 \times D$, and similarly, $h = 2$ represents two kinds of filters $1 \times D, 2 \times D$. The experiment results show that the module of variable-size convolution is effective in aggregating the multi-scale area features.

5.1. Ablation study

In the following, we show the effects of the two main modules: the intra-region context encoding and the inter-region context encoding, respectively. In Table 7, we show the performances of LRC-Net with and without the intra-region context encoding module. Specifically, when we remove the intra-region context encoding, there are three widely used ways to aggregate the features of multi-scale areas by mean pooling (Mean), max pooling (Max) and concatenating (Con), respectively.

Table 7
The effect of intra-region context encoding module in LRC-Net on ModelNet40.

Metric	All	Mean	Max	Con
Accuracy (%)	**93.07**	92.50	92.38	92.30

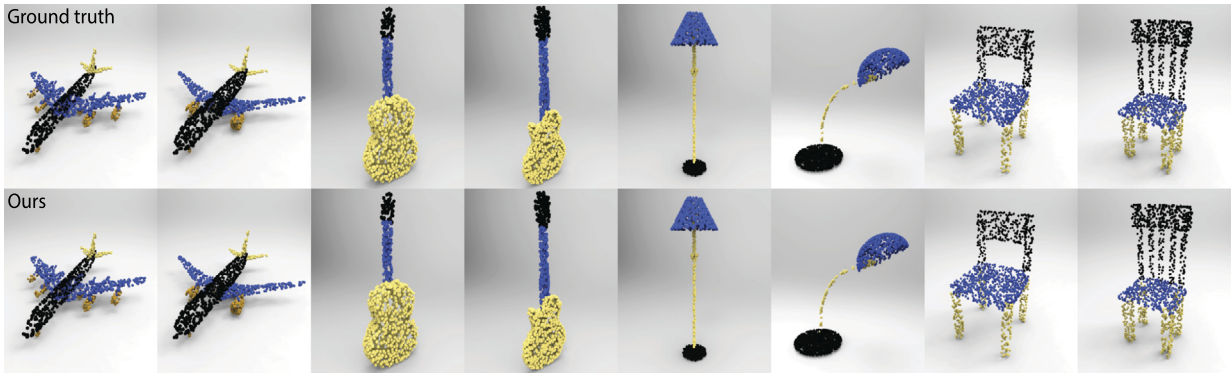

Fig. 3. Visualization of some shape segmentation results. The top row the is ground truth point clouds, and the bottom row is our predicted results, where parts with the same color belong to the same class. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

These results show that the intra-region context encoding module can promote the discriminative ability of the learned point cloud features. Similarly, we also evaluate the role of the inter-region context encoding module. As depicted in Table 8, we then list the results with (Y) and without (N) the inter-region context encoding module. In addition, we also show the influence of the max pooling operation (max) or the mean pooling operation (mean) in the PointNet layer which extracts the global representation $\boldsymbol{g}$ as shown in Fig. 2. Therefore, there are four alternative combinations, Y(max), N(max), Y(mean) and N(mean), respectively. The results suggest that the inter-region context encoding module is effective in improving the learning of global features by capturing the spatial context among local regions. According to above comparisons, the two modules in LRC-Net are effective in encoding local region contexts.

Table 8
The effect of inter-region context encoding module in LRC-Net on ModelNet40.

Metric	Y(max)	N(max)	Y(mean)	N(mean)	N(sum)
Accuracy (%)	**93.07**	92.26	91.57	91.37	91.41

5.2. Shape classification

The performances of LRC-Net are evaluated on both ModelNet10 (MN10) and ModelNet40 (MN40) 3D shape classification benchmarks. In detail, MN40 contains 40 categories and 12,311 CAD shapes with 9,843 shapes for training and 2,468 shapes for testing. And MN10 is a subset of MN40 with 4,899 CAD shapes, including 3,991 shapes for training and 908 shapes for testing. Table 1 compares LRC-Net with several state-of-the-art methods in terms of instance accuracy on MN10 and MN40, respectively. As shown in the Table 1, all methods can be divided into two categories: single-scale based methods (Li et al., 2018b; Komarichev et al., 2019) and multi-scale based methods (Qi et al., 2017b; Liu et al., 2019a). LRC-Net has greatly improved the baseline of PointNet++ (Qi et al., 2017b) on both ModelNet10 and ModelNet40. And LRC-Net achieves the same results with Point2SpatialCapsule (Wen et al., 2019) on ModelNet10 and reaches comparable results with Point2SpatialCapsule on ModelNet40. Point2SpatialCapsule benefits from its network structures such as dynamic routing for clustering and point cloud reconstruction, which aims to increase the network capability. However, the two newly added modules (i.e. clustering and point cloud reconstruction) increase both the model size and the computational cost of Point2SpatialCapsule during network learning. This makes Point2SpatialCapsule more complicated than LRC-Net in term of the network architecture. The best accuracy 94.2% is achieved with 10,000 points as input, where the higher resolution point cloud can provide more local details than the sparse input with 1,024 points. Experimental results show that LRC-Net can effectively enhance the representation learning of point clouds from multi-scale local regions by capturing the contextual information inside and among local regions.

In addition, to show the network complexity of LRC-Net intuitively, we make a statistics of model size and space cost of some point cloud based methods. We follow PointNet++ to evaluate the time and space cost of several point cloud based methods as shown in Table 9. We record forward time under the same conditions with a batch size 8 using TensorFlow 1.0 with a single GTX 1080 Ti. Table 9 shows LRC-Net can achieve tradeoff between the model complexity (number of

Table 9
Complexity, forward time, and accuracy on ModelNet40 of different models.

Method	Model size (MB)	Time (MS)	Accuracy (%)
PointNet (vanilla) (Qi et al., 2017a)	9.4	6.8	87.1
PointNet (Qi et al., 2017a)	40	16.6	89.2
PointNet++ (SSG) (Qi et al., 2017b)	8.7	82.4	-
PointNet++ (MSG) (Qi et al., 2017b)	12	163.2	90.7
PointCNN (Li et al., 2018b)	94	117.0	92.3
LRC-Net (ours)	18	115.8	93.1

parameters) and computational complexity (forward pass time). However, influenced by the setting of multi-scale grouping (MSG), LRC-Net and PointNet++ take longer than other single-scale grouping (SSG) based methods.

5.3. Shape segmentation

To further verify the validity of our model, we also evaluate the performance of LRC-Net in the shape segmentation task. The shape segmentation branch is implemented as depicted in Fig. 2. In this task, ShapeNet Part dataset is adopted as the benchmark which contains 16,881 models from 16 categories and is spit into train set, validation set and test set as PointNet++. There are 2,048 points for each point cloud, where each point belongs to a certain one of 50 part classes. And the kind of semantic parts in each shape varies from 2 to 5. There is no overlap of the part classes between shapes in different shape categories.

We employ the mean Intersection over Union (IoU) proposed in Qi et al. (2017a) as the evaluation metric for shape segmentation. For each shape, the IoU is computed between ground-truth and the prediction for each part class in the shape category. And the average IoUs are calculated in each shape category and overall shapes. In Table 2, we report the performance of LRC-Net in each category and the mean IoU of all testing shapes.

From Table 2, the performance of LRC-Net is not as good as three latest proposed single-scale based methods including PointCNN (Li et al., 2018b), A-CNN (Komarichev et al., 2019) and RS-CNN (Liu et al., 2019c) that adopt some special strategies in the training process. For example, A-CNN states “We concatenate the one-hot encoding of the object label to the last feature layer” and PointCNN states “we perturb point locations with the point shuffling for better generalization”, which are different from mainstream approaches like PointNet (Qi et al., 2017a). For fair comparison with most of other methods, we do not apply these strategies in our method.

Moreover, compared with other multi-scale based methods (Qi et al., 2017b; Liu et al., 2019a), LRC-Net achieves the best mean instance IoU of 85.3% and comparable performances on many shape categories, which shows the effective of enhancing the contextual information inside and among local regions. In addition, some visualizations of the shape segmentation results are shown in Fig. 3, where our predictions are highly consistent with the ground-truths. The shape segmentation results qualitatively show the effectiveness of LRC-Net in capturing the contextual information for each point.

Table 10
The performance of LRC-Net in the semantic segmentation on S3DIS.

Method	Mean IoU	Overall accuracy
PointNet (baseline) (Qi et al., 2017a)	20.1	53.2
PointNet (Qi et al., 2017a)	47.6	78.5
MS + CU (2) (Engelmann et al., 2017)	47.8	49.7
G + RCU (Engelmann et al., 2017)	49.7	81.1
ShapeContextNet (Xie et al., 2018)	52.7	81.6
LRC-Net (ours)	52.0	81.3

5.4. Indoor scene segmentation

We evaluate our model on Standford Large-Scale 3D Indoor Spaces Dataset (S3DIS) (Armeni et al., 2016) for the semantic scene segmentation task. There are 6 indoor areas including 272 rooms of the 3D scan point clouds in the S3DIS dataset. Each point in one point cloud belongs to one of the 13 categories, e.g. chair, board, ceiling and beam. We follow the same setting as in PointNet (Qi et al., 2017a), where each room is split into blocks and 4,096 points are sampled from each block in the training process. In the testing process, all the points are used. We also apply the 6-fold cross validation over the 6 areas and report the average evaluation results.

Similar to shape part segmentation task, the probability distribution over the semantic object classes is generated for each input point. The quantified comparison results with some existing methods are reported in Table 10. LRC-Net outperforms PointNet (Qi et al., 2017a) and achieves comparable results with ShapeContextNet (Xie et al., 2018).

6. Conclusion

In this paper, we propose a novel feature learning framework for the understanding of point cloud in the shape classification and shape segmentation. With the intra-region context encoding module, the LRC-Net effectively learns the correlation between multi-scale areas inside each local region. To enhance the aggregation of local region features, a greedy strategy enables to encode the inter-region context of point clouds. We justify that both of these two modules are vital to encode local region contexts, which promote learning discriminative feature for point clouds.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by National Key R&D Program of China (2018YFB0505400).

References

Armeni, I., Sener, O., Zamir, A.R., Jiang, H., Brilakis, I., Fischer, M., Savarese, S., 2016. 3D semantic parsing of large-scale indoor spaces. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1534–1543.

Beksi, W.J., Papanikolopoulos, N., 2019. A topology-based descriptor for 3D point cloud modeling: theory and experiments. Image Vis. Comput. 88, 84–95.

Belongie, S., Malik, J., Puzicha, J., 2001. Shape context: a new descriptor for shape matching and object recognition. In: The Conference on Neural Information Processing Systems (NeurIPS), pp. 831–837.

Engelmann, F., Kontogianni, T., Hermans, A., Leibe, B., 2017. Exploring spatial context for 3D semantic segmentation of point clouds. In: The IEEE International Conference on Computer Vision Workshops, pp. 716–724.

Fehr, D., Beksi, W.J., Zermas, D., Papanikolopoulos, N., 2016. Covariance based point cloud descriptors for object detection and recognition. Comput. Vis. Image Underst. 142, 80–93.

Gao, G., Liu, Y.S., Lin, P., Wang, M., Gu, M., Yong, J.H., 2017. BIMTag: concept-based automatic semantic annotation of online BIM product resources. Adv. Eng. Inform. 31, 48–61.

Gao, G., Liu, Y.S., Wang, M., Gu, M., Yong, J.H., 2015. A query expansion method for retrieving online BIM resources based on industry foundation classes. Autom. Constr. 56, 14–25.

Golovinskiy, A., Kim, V.G., Funkhouser, T., 2009. Shape-based recognition of 3D point clouds in urban environments. In: The IEEE International Conference on Computer Vision (ICCV), pp. 2154–2161.

Han, Z., Liu, X., Liu, Y.S., Zwicker, M., 2019a. Parts4Feature: learning 3D global features from generally semantic parts in multiple views. In: The International Joint Conference on Artificial Intelligence (IJCAI), pp. 766–773.

Han, Z., Liu, Z., Han, J., Vong, C.M., Bu, S., Chen, C., 2017a. Mesh convolutional restricted Boltzmann machines for unsupervised learning of features with structure preservation on 3D meshes. IEEE Trans. Neural Netw. Learn. Syst. 28, 2268–2281.

Han, Z., Liu, Z., Han, J., Vong, C.M., Bu, S., Chen, C., 2019b. Unsupervised learning of 3D local features from raw voxels based on a novel permutation voxelization strategy. IEEE Trans. Cybern. 49, 481–494.

Han, Z., Liu, Z., Han, J., Vong, C.M., Bu, S., Li, X., 2016. Unsupervised 3D local feature learning by circle convolutional restricted Boltzmann machine. IEEE Trans. Image Process. 25, 5331–5344.

Han, Z., Liu, Z., Vong, C.M., Liu, Y.S., Bu, S., Han, J., Chen, C.P., 2017b. BoSCC: bag of spatial context correlations for spatially enhanced 3D shape representation. IEEE Trans. Image Process. 26, 3707–3720.

Han, Z., Liu, Z., Vong, C.M., Liu, Y.S., Bu, S., Han, J., Chen, C.P., 2018a. Deep spatiality: unsupervised learning of spatially-enhanced global and local 3D features by deep neural network with coupled softmax. IEEE Trans. Image Process. 27, 3049–3063.

Han, Z., Lu, H., Liu, Z., Vong, C.M., Liu, Y.S., Zwicker, M., Han, J., Chen, C.P., 2019c. 3D2SeqViews: aggregating sequential views for 3D global feature learning by CNN with hierarchical attention aggregation. IEEE Trans. Image Process. 28, 3986–3999.

Han, Z., Shang, M., Liu, Y.S., Zwicker, M., 2019d. View Inter-Prediction GAN: unsupervised representation learning for 3D shapes by learning global shape memories to support local view predictions. In: The AAAI Conference on Artificial Intelligence (AAAI), pp. 8376–8384.

Han, Z., Shang, M., Liu, Z., Vong, C.M., Liu, Y.S., Zwicker, M., Han, J., Chen, C.P., 2018b. SeqViews2SeqLabels: learning 3D global features via aggregating sequential views by RNN with attention. IEEE Trans. Image Process. 28, 658–672.

Han, Z., Shang, M., Wang, X., Liu, Y.S., Zwicker, M., 2019e. Y2Seq2Seq: cross-modal representation learning for 3D shape and text by joint reconstruction and prediction of view and word sequences. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 126–133.

Han, Z., Wang, X., Liu, Y.S., Zwicker, M., 2019f. Multi-Angle Point Cloud-VAE: unsupervised feature learning for 3D point clouds from multiple angles by joint self-reconstruction and half-to-half prediction. In: The IEEE International Conference on Computer Vision (ICCV).

Han, Z., Wang, X., Vong, C.M., Liu, Y.S., Zwicker, M., Chen, C., 2019g. 3DViewGraph: learning global features for 3D shapes from a graph of unordered views with attention. In: The International Joint Conference on Artificial Intelligence (IJCAI), pp. 758–765.

Hu, T., Han, Z., Shrivastava, A., Zwicker, M., 2019. Render4Completion: synthesizing multi-view depth maps for 3D shape completion. In: The IEEE International Conference on Computer Vision Workshops.

Kim, Y., 2014. Convolutional neural networks for sentence classification. In: The Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1746–1751.

Klokov, R., Lempitsky, V., 2017. Escape from cells: deep kd-networks for the recognition of 3D point cloud models. In: The IEEE International Conference on Computer Vision (ICCV), IEEE, pp. 863–872.

Komarichev, A., Zhong, Z., Hua, J., 2019. A-CNN: annularly convolutional neural networks on point clouds. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7421–7430.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. In: The Conference on Neural Information Processing Systems (NeurIPS), pp. 1097–1105.

Li, J., Chen, B.M., Hee Lee, G., 2018a. SO-Net: self-organizing network for point cloud analysis. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 9397–9406.

Li, Y., Bu, R., Sun, M., Wu, W., Di, X., Chen, B., 2018b. PointCNN: convolution on X-transformed points. In: The Conference on Neural Information Processing Systems (NeurIPS), pp. 820–830.
Li, Y., Pirk, S., Su, H., Qi, C.R., Guibas, L.J., 2016. FPNN: field probing neural networks for 3D data. In: The Conference on Neural Information Processing Systems (NeurIPS), pp. 307–315.
Liu, M., Liu, Y.S., Ramani, K., 2009. Computing global visibility maps for regions on the boundaries of polyhedra using Minkowski sums. Comput. Aided Des. 41, 668–680.
Liu, X., Han, Z., Liu, Y.S., Zwicker, M., 2019a. Point2Sequence: learning the shape representation of 3D point clouds with an attention-based sequence to sequence network. In: The AAAI Conference on Artificial Intelligence (AAAI), pp. 8778–8785.
Liu, X., Han, Z., Wen, X., Liu, Y.S., Zwicker, M., 2019b. L2G auto-encoder: understanding point clouds by local-to-global reconstruction with hierarchical self-attention. In: The ACM International Conference on Multimedia (ACM MM), pp. 989–997.
Liu, Y., Fan, B., Xiang, S., Pan, C., 2019c. Relation-shape convolutional neural network for point cloud analysis. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 8895–8904.
Liu, Y.S., Ramani, K., 2009. Robust principal axes determination for point-based shapes using least median of squares. Comput. Aided Des. 41, 293–305.
Liu, Y.S., Ramani, K., Liu, M., 2011. Computing the inner distances of volumetric models for articulated shape description with a visibility graph. IEEE Trans. Pattern Anal. Mach. Intell. 33, 2538–2544.
Mao, J., Wang, X., Li, H., 2019. Interpolated convolutional networks for 3D point cloud understanding. In: The IEEE International Conference on Computer Vision (ICCV).
Maturana, D., Scherer, S., 2015. VoxNet: a 3D convolutional neural network for real-time object recognition. In: The IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, pp. 922–928.
Nair, V., Hinton, G.E., 2010. Rectified linear units improve restricted Boltzmann machines. In: The International Conference on Machine Learning (ICML), pp. 807–814.
Qi, C.R., Liu, W., Wu, C., Su, H., Guibas, L.J., 2018. Frustum pointnets for 3D object detection from rgb-d data. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 918–927.
Qi, C.R., Su, H., Mo, K., Guibas, L.J., 2017a. PointNet: deep learning on point sets for 3D classification and segmentation. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 652–660.
Qi, C.R., Yi, L., Su, H., Guibas, L.J., 2017b. PointNet++: deep hierarchical feature learning on point sets in a metric space. In: The Conference on Neural Information Processing Systems (NeurIPS), pp. 5099–5108.
Riegler, G., Ulusoy, A.O., Geiger, A., 2017. OctNet: learning deep 3D representations at high resolutions. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3577–3586.
Rusu, R.B., Marton, Z.C., Blodow, N., Dolha, M., Beetz, M., 2008. Towards 3D point cloud based object maps for household environments. In: Robotics and Autonomous Systems, pp. 927–941.
Savva, M., Yu, F., Su, H., Aono, M., Chen, B., Cohen-Or, D., Deng, W., Su, H., Bai, S., Bai, X., et al., 2016. SHREC'16 track large-scale 3D shape retrieval from ShapeNet core55. In: The Eurographics Workshop on 3D Object Retrieval, pp. 89–98.
Shen, Y., Feng, C., Yang, Y., Tian, D., 2018. Mining point cloud local structures by kernel correlation and graph pooling. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
Skrodzki, M., Jansen, J., Polthier, K., 2018. Directional density measure to intrinsically estimate and counteract non-uniformity in point clouds. Comput. Aided Geom. Des. 64, 73–89.
Srivastava, S., Lall, B., 2019. DeepPoint3D: learning discriminative local descriptors using deep metric learning on 3D point clouds. Pattern Recognit. Lett. 127, 27–36.
Wang, P.S., Liu, Y., Guo, Y.X., Sun, C.Y., Tong, X., 2017. O-CNN: octree-based convolutional neural networks for 3D shape analysis. ACM Trans. Graph. (TOG) 36, 72.
Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M., Solomon, J.M., 2018. Dynamic graph CNN for learning on point clouds. ACM Trans. Graph. (TOG).
Wen, X., Han, Z., Liu, X., Liu, Y.S., 2019. Point2SpatialCapsule: aggregating features and spatial relationships of local regions on point clouds using spatial-aware capsules. arXiv preprint. arXiv:1908.11026.
Wen, X., Li, T., Han, Z., Yu-Shen, L., 2020. Point cloud completion by skip-attention network with hierarchical folding. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
Wu, Z., Song, S., Aditya, K., Yu, F., Zhang, L., Tang, X., Xiao, J., 2015. 3D ShapeNets: a deep representation for volumetric shapes. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1912–1920.
Xie, S., Liu, S., Chen, Z., Tu, Z., 2018. Attentional ShapeContextNet for point cloud recognition. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4606–4615.
Xu, Y., Fan, T., Xu, M., Zeng, L., Qiao, Y., 2018. SpiderCNN: Deep Learning on point sets with parameterized convolutional filters. In: ECCV.
Yi, C., Lu, D., Xie, Q., Liu, S., Li, H., Wei, M., Wang, J., 2019. Hierarchical tunnel modeling from 3D raw LiDAR point cloud. Comput. Aided Des. 114, 143–154.
Zhao, H., Tang, M., Ding, H., 2020. HoPPF: a novel local surface descriptor for 3D object recognition. Pattern Recognit., 107272.
Zheng, Y., Li, G., Xu, X., Wu, S., Nie, Y., 2018. Rolling normal filtering for point clouds. Comput. Aided Geom. Des. 62, 16–28.
Zhong, S., Zhong, Z., Hua, J., 2019. Surface reconstruction by parallel and unified particle-based resampling from point clouds. Comput. Aided Geom. Des. 71, 43–62.
Zhou, Y., Tuzel, O., 2017. VoxelNet: end-to-end learning for point cloud based 3D object detection. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4490–4499.
Zhu, Y., Mottaghi, R., Kolve, E., Lim, J.J., Gupta, A., Fei-Fei, L., Farhadi, A., 2017. Target-driven visual navigation in indoor scenes using deep reinforcement learning. In: The IEEE International Conference on Robotics and Automation (ICRA), pp. 3357–3364.
Zou, Y., Wang, X., Zhang, T., Liang, B., Song, J., Liu, H., 2018. BRoPH: an efficient and compact binary descriptor for 3D point clouds. Pattern Recognit. 76, 522–536.