

# 融合门控自校准机制和图卷积网络的点云分析

徐嘉利<sup>1</sup>, 方志军<sup>1\*</sup>, 伍世虔<sup>2</sup>

<sup>1</sup>上海工程技术大学电子电气工程学院, 上海 201620;

<sup>2</sup>武汉科技大学信息科学与工程学院, 湖北 武汉 430081

**摘要** 与密集网格表示的图像不同,点云自身具有不规则和无序性的特点,因而如何准确地推理出点云数据中的形状特征是一项具有挑战性的工作。为解决当前研究存在的不足,提出了点集内-外形状卷积(IE-Conv)。该卷积利用高效的双边结构将点集内的局部形状与点集外的全局形状分开处理,在点集内部,基于门控的方式有选择地学习丰富的点间关系,同时利用自校准功能优化逐点与局部特征;在点集外部,通过图卷积构建全局图形,并聚焦于点集之间的远程依赖关系;最后将双边输出有机融合起来。将 IE-Conv 分层嵌入形状推理卷积网络中(SR-Net),并在标准 ModelNet40 和 ShapeNet 数据集上进行分类与分割实验。实验结果表明,分类任务精度达到 93.9%,分割任务平均交并比达到 86.4%,验证了 SR-Net 在点云分析中的良好性能。

**关键词** 图像处理; 点云分类与分割; 深度学习; 图卷积; 多特征融合

中图分类号 TP391

文献标志码 A

DOI: 10.3788/LOP202259.1210017

## Point Cloud Analysis Combining Gated Self-Calibration Mechanism and Graphical Convolutional Network

Xu Jiali<sup>1</sup>, Fang Zhijun<sup>1\*</sup>, Wu Shiqian<sup>2</sup>

<sup>1</sup>*School of Electrical and Electronic Engineering, Shanghai University of Engineering Science, Shanghai 201620, China;*

<sup>2</sup>*School of Information Science and Engineering, Wuhan University of Science and Technology, Wuhan 430081, Hubei, China*

**Abstract** Point clouds, unlike images represented by dense grids, are characterized by irregularity and disorder, making it difficult to precisely reason out the shape features in point cloud data. The internal-external shape convolution for point sets (IE-Conv) is proposed to address the limitations of current research. The local shape inside the point set is treated separately from the global shape outside the point set using an efficient bilateral design. Rich inter-point relationships are selectively studied in a gate-based manner within the point set, while point-by-point and local features are optimized by self-calibration functions; outside the point set, global shapes are constructed using graph convolution and focus on long-range dependencies between point sets. Finally, the organic fusion of the bilateral outputs is performed. This paper performs classification and segmentation experiments on the standard ModelNet40 and ShapeNet datasets by hierarchically embedding IE-Conv into the shape-reasoning convolutional network (SR-Net). The experimental results show that the classification task achieves an accuracy of 93.9% and the segmentation task achieves the mean intersection over the union of 86.4%, which verifies the good performance of SR-Net in point cloud analysis.

**Key words** image processing; point cloud classification and segmentation; deep learning; graph convolution; multi-feature fusion

收稿日期: 2021-07-29; 修回日期: 2021-09-12; 录用日期: 2021-09-23

基金项目: 国家自然科学基金(61772328,U2033218)

通信作者: \*zjfang@sues.edu.cn

# 1 引言

随着自动驾驶<sup>[1]</sup>、机器人<sup>[2]</sup>和遥感<sup>[3]</sup>等领域中的三维技术的快速发展,引发了人们了解和分析三维数据的兴趣。三维数据通常是以点云来表示的,它是一组无序且排列多样的点。在不同的任务需求下,可以在每个点上添加额外信息(RGB、法线)。由于点云自身非结构化的特点,传统的卷积神经网络很难扩展到三维数据。因此,需要开发一种面向形状的点云学习方法,该方法不仅要能够直接处理三维点云,也要能对隐含的形状特征进行抽象描述,并理解不同区域下,每个局部形状的差异信息。这项富有挑战性的任务吸引了越来越多的人投入相关的研究中<sup>[4-7]</sup>。

PointNet<sup>[8]</sup>是第一种基于深度学习的方法,它能够端到端地处理三维点云,但是它侧重于全局特征,却忽略了重要的局部特征。PointNet++<sup>[9]</sup>在分区的点集上应用PointNet<sup>[8]</sup>弥补了局部特征提取的不足,但却忽视了点与点之间的相关性。Dynamic graph convolutional neural network(DGCNN)<sup>[10]</sup>使用共享 multilayer perceptron (MLPs)处理点集内质心点与其邻点之间的边,但未利用点集之间存在的联系。Annularly convolutional neural networks(A-CNN)<sup>[11]</sup>由质心的法向量构造环卷积,建模不同范围下的形状特征。Point2Sequence<sup>[12]</sup>和 PointANSL<sup>[13]</sup>基于注意力思路学习不同区域的相关性,从而提高模型训练的鲁棒性。Relation-shape convolutional neural network(RS-CNN)<sup>[14]</sup>利用十维几何先验建模点间关系,并将其编码至高维空间,但过分强调了点的几何关系,省略了点集内部必要的法线特征。

为了解决上述研究存在的问题,本文提出了一

种基于双边架构的点集内-外形状卷积(IE-Conv),从点集内部与点集外部两个分支捕捉形状特征。在内-点集模块中,利用丰富的几何和法线表达式对质心点与其邻点之间的相关性进行建模,通过门控操作分配权重值,自适应地感知点间关系,保留了局部范围内的关键信息,强化了对单一点集形状的挖掘。同时,由于点云存在平移或旋转的可能性,为了保证模型的鲁棒性,利用自校准机制修正逐点和局部特征。在外-点集模块中,引入图卷积构建全局图形,建模原本孤立点集之间的远程依赖联系,构建富有空间结构的全局图,增强了在全局视角下对数据中形状信息的学习能力。最后将上述模块的输出进行有机的多特征融合,以IE-Conv为基础构建端到端的、鲁棒的形状推理卷积网络(SR-Net)。实验结果表明,该网络达到了对点云数据进行准确分类与分割的目的。

## 2 点集内-外形状卷积(IE-Conv)

局部特征和全局特征在本质上是互补的关系,局部特征包含的更多的是点信息,但缺乏全局空间布局关系,无法识别大范围点云边缘差异性,全局特征侧重于交互多个点集之间的信息流,传播点集之间的异同性,进而获得更大的感受野,但对点云细节特征的分辨能力较差。所提IE-Conv是一种可在局部和全局之间进行良好折衷的双边结构,如图1所示,它由内-点集模块、外-点集模块构成,分别推理单个点集内部细微的形状以及多个点集之间的全局结构联系,并将它们密集地穿插在一起,以此迭代更新形状特征。

由于在三维空间中不规则分布的特点,点云数据无法类似于二维图像直接应用卷积运算。数据

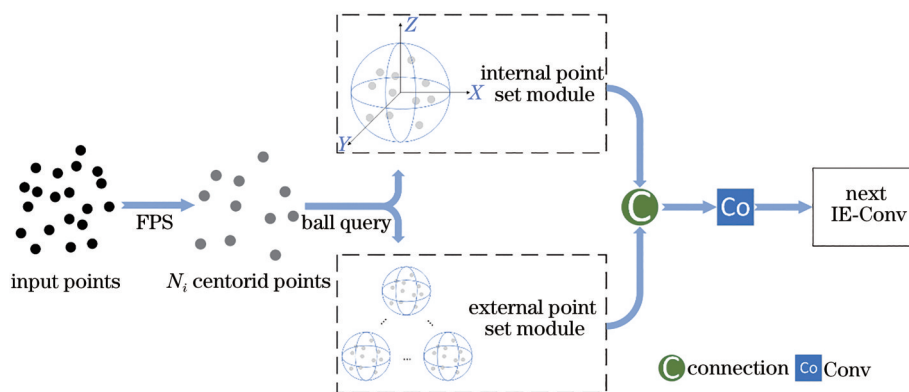


图1 IE-Conv结构图

Fig. 1 Structure of IE-Conv

中每一点并不是孤立的,不同点之间相互作用,在特定范围内的点集中反应了相似的形状信息。构建不同范围区域的点集对点云数据处理是必要的。记输入的  $N$  个无序的点云  $\mathbf{P} = \{p_1, p_2, p_3, \dots, p_N\}$ , 选择最远点采样(FPS)得到  $h$  个质心点。随后利用球查询在半径  $r$  内得到  $t$  个邻点。在此基础上构建球形点集  $\mathbf{B}_i = \{p_i, p_{i,1}, p_{i,2}, p_{i,3}, \dots, p_{i,t}\}, d_{i,j} < 2r, \mathbf{B}_i \in \mathbf{R}^{N_h \times K_{t+1} \times \Omega}$ , 其中  $d_{i,j}$  是质心点与其邻点之间的欧几里得距离,  $\Omega$  是特征通道维度。

### 2.1 内-点集模块

在单一点集中,内-点集模块原理图如[图2(a)]所示,局部形状取决于合理的逐点特征,而逐点特征反映了质心点与其邻点之间的密集作用关系。近期一些研究<sup>[10,14,15]</sup>的成果也证实了该思路的正确性。然而上述方法仅关注两点在位置坐标上的空间差异,再将其固定为逐点特征的基准,使网络在面对不规则点云数据时,无法灵活地感知点间关系,导致关键信息的丢失,影响了局部形状的分析精度。本实验组在几何先验表达式  $e_{i,j}^{geo} \in \mathbf{R}^{N_h \times K_{t+1} \times 7}$  的基础上添加法线先验表达式  $e_{i,j}^{nor} \in \mathbf{R}^{N_h \times K_{t+1} \times 7}$  来补充点间各自的交互作用,如图3所示。

$$\begin{cases} e_{i,j}^{geo} = [p_i^{coord}, p_{i,j}^{coord}, \|p_i - p_{i,j}\|_{ed}] \\ e_{i,j}^{nor} = [p_i^{nor}, p_{i,j}^{nor}, \|p_i - p_{i,j}\|_{cosd}] \end{cases}, \quad (1)$$

$$\begin{cases} F_{geo} = [MLP_1(e_{i,j}^{geo})] \cdot B_i \\ F_{nor} = [MLP_1(e_{i,j}^{nor})] \cdot B_i \end{cases}, \quad (2)$$

式中:  $[\cdot, \cdot]$  是连接函数;  $p_i^{coord} \in \mathbf{R}^{N_h \times K_{t+1} \times 3}, p_{i,j}^{coord} \in \mathbf{R}^{N_h \times K_{t+1} \times 3}, \|p_i - p_{i,j}\|_{ed} \in \mathbf{R}^{N_h \times K_{t+1} \times 1}$  分别表示的是质心点、邻点的坐标和两者之间的欧氏距离;  $p_i^{nor} \in$

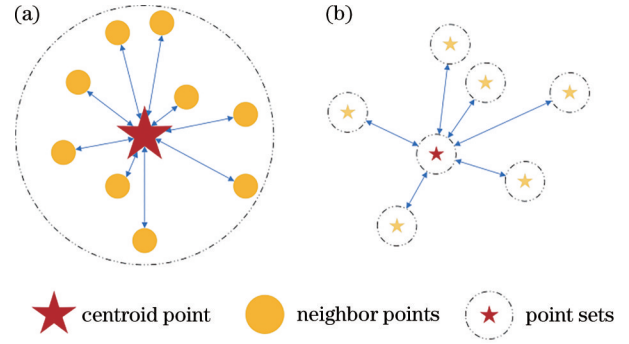


图2 内、外-点集模块原理示意图。(a)内-点集模块;  
(b)外-点集模块

Fig. 2 Schematic diagrams of inner and outer point set module. (a) Internal point set module; (b) external point set module

$\mathbf{R}^{N_h \times K_{t+1} \times 3}, p_{i,j}^{nor} \in \mathbf{R}^{N_h \times K_{t+1} \times 3}, \|p_i - p_{i,j}\|_{cosd} \in \mathbf{R}^{N_h \times K_{t+1} \times 1}$  分别为质心点、邻点的法线及两者的余弦距离。随后将它们分别级联起来,使用多层感知机(MLP<sub>1</sub>)映射出两点间的抽象关系,并编码至特定维度( $\mathbf{R}^{N_h \times K_{t+1} \times 7} \rightarrow \mathbf{R}^{N_h \times K_{t+1} \times \Omega}$ ),使它便于与点集原始特征  $\mathbf{B}_i$  点乘。最后得到初始的几何特征  $F_{geo} \in \mathbf{R}^{N_h \times K_{t+1} \times \Omega}$  与法线特征  $F_{nor} \in \mathbf{R}^{N_h \times K_{t+1} \times \Omega}$ 。

当某一采样点在空间布局上产生旋转或平移方差时,它的邻点位置会在局部区域中相应做出调整。因而添加了校准机制,选择两点间的坐标差  $(p_i^{coord} - p_{i,j}^{coord}) \in \mathbf{R}^{N_h \times K_{t+1} \times 3}$  作为结构特征,通过 MLP<sub>2</sub> 去预测修正量(维度变化为  $N_h \times K_{t+1} \times 3 \rightarrow N_h \times K_{t+1} \times 1$ ),其值可能是正的也可能是负的,避免了上述情况对逐点特征  $F_{p-w} \in \mathbf{R}^{N_h \times K_{t+1} \times 2\Omega}$  的鲁棒性产生影响。具体操作可描述为

$$F_{p-w} = [MLP_2(p_i^{coord} - p_{i,j}^{coord})] \otimes [F_{geo}, F_{nor}]. \quad (3)$$

从[图2(a)]可以看出,质心点与其邻点之间,

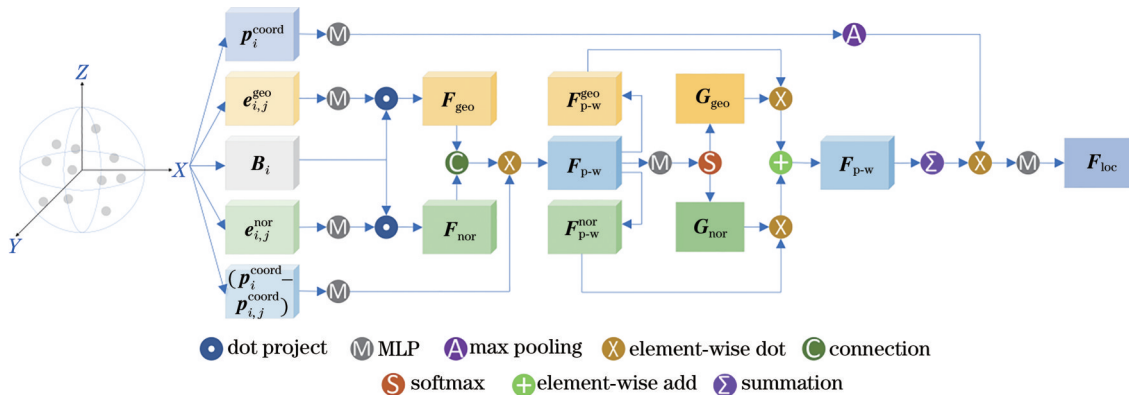


图3 内-点集模块结构图

Fig. 3 Structure of internal point set module

几何和法线特征两者对点间关系的贡献是不同的。例如两点在空间位置上距离较近,那么逐点特征更依赖于法线特征,反之选择几何特征更为有效。为了实现这一点,使得内-点集模块获得跨态特征的代表能力,受注意力机制<sup>[16-18]</sup>的启发,利用门控的方式自适应地挑选合适的表达式。因此几何门  $\mathbf{G}_{\text{geo}} \in \mathbb{R}^{N_h \times K_{t+1} \times 1}$  和法线门  $\mathbf{G}_{\text{nor}} \in \mathbb{R}^{N_h \times K_{t+1} \times 1}$  以及后续过程可以表示为

$$[\mathbf{G}_{\text{geo}}, \mathbf{G}_{\text{nor}}] = \sigma[\mathbf{g}_{\text{geo}}, \mathbf{g}_{\text{nor}}] = \sigma[\text{MLP}_3(\mathbf{F}_{\text{p-w}})], \quad (4)$$

$$[\mathbf{F}_{\text{p-w}}^{\text{geo}}, \mathbf{F}_{\text{p-w}}^{\text{nor}}] = \mathbf{F}_{\text{p-w}}, \quad (5)$$

$$\begin{cases} \mathbf{G}_{\text{geo}} = \frac{e^{\mathbf{g}_{\text{geo}}}}{e^{\mathbf{g}_{\text{geo}}} + e^{\mathbf{g}_{\text{nor}}}} \\ \mathbf{G}_{\text{nor}} = \frac{e^{\mathbf{g}_{\text{nor}}}}{e^{\mathbf{g}_{\text{geo}}} + e^{\mathbf{g}_{\text{nor}}}} \end{cases}, \quad (6)$$

$$\mathbf{F}_{\text{p-w}} = (\mathbf{G}_{\text{geo}} \otimes \mathbf{F}_{\text{p-w}}^{\text{geo}}) \oplus (\mathbf{G}_{\text{nor}} \otimes \mathbf{F}_{\text{p-w}}^{\text{nor}}), \quad (7)$$

式中:  $\oplus$  是逐元素相加;  $\otimes$  是逐元素相乘;  $\sigma$  是激活函数 Softmax;  $\mathbf{F}_{\text{p-w}}^{\text{geo}} \in \mathbb{R}^{N_h \times K_{t+1} \times \Omega}$ 、 $\mathbf{F}_{\text{p-w}}^{\text{nor}} \in \mathbb{R}^{N_h \times K_{t+1} \times \Omega}$  分别是修正后的几何特征与法线特征。由于在同一个点集中,几何特征和法线特征有很好的互补关系,为了使门更精确,所提模块将几何与法线特征通过式(3)校准机制先行聚合  $[\mathbf{F}_{\text{geo}}, \mathbf{F}_{\text{nor}}]$ , 并利用  $\text{MLP}_3$  在某两点之间组合它们在不同关系下的特征联系(维度变化为  $N_h \times K_{t+1} \times 2\Omega \rightarrow N_h \times K_{t+1} \times 2$ )。本实验组利用 Softmax 激活函数归一化得到各点对应于法线和几何的注意力权重,即门。选择门控制点间关系的信息流,避免了模型沉迷于无效关系,减少了对不重要特征的关注度,使其倾向于在交互中挑选所需的特征进行学习。随后,通过加权求和运算,更新并保留了关键的逐点特征  $\mathbf{F}_{\text{p-w}} \in \mathbb{R}^{N_h \times K_{t+1} \times \Omega}$ 。为了进一步提高模块对点云刚性变换的鲁棒性,第2次修正局部区域,得到该分支的局部形状  $\mathbf{F}_{\text{loc}} \in \mathbb{R}^{N_h \times \Phi}$ ,

整个流程可被简化为

$$\mathbf{F}_{\text{loc}} = \text{MLP}_4 \left\{ \left\{ \text{SUM} \sum_{j=1}^{t+1} (\mathbf{F}_{\text{p-w}}) \right\} \otimes \left\{ \text{Max} \sum_{j=1}^{t+1} [\text{MLP}_2(\mathbf{p}_i^{\text{coord}})] \right\} \right\}. \quad (8)$$

从局部点集的角度来看,邻点对空间结构变化并不敏感,可以将其忽略,而选择质心点坐标作为修正值,能够以点集为单位优化扰动的影响。其次选择 Summation 作为聚合函数,不仅能够完整地保留经由门控选择后的最优特征,同时也可强化局部形状的代表能力。最后,利用  $\text{MLP}_4$  强大的映射能力,将其编码为抽象的局部形状。这样的操作也能够降低通道数(维度变化为  $N_h \times \Omega \rightarrow N_h \times \Phi$ ),使模型维持在可接受的复杂度。

## 2.2 外-点集模块

如今的点云分析研究者大多聚焦于局部点集内潜藏的形状关系,但却忽略了单一点集之间存在的关联性。在第 2.1 节中,内-点集模块致力于感知局部形状,但由于不同局部区域是互相孤立的,该模块无法有效地在大范围视角下推理全局联系。二维图像处理与视频解析等领域的相关研究同样证明了关注区域之间全局关系的重要性。受图卷积方法<sup>[10,20,21]</sup>及相关研究<sup>[22,23]</sup>的启发,所提外-点集模块通过构建图网络来描述多元局部点集,建模局部区域之间的远程依赖关系,弥补点云分析中缺乏全局形状的不足,如图 4 所示。

首先,本实验组将相互独立的点集视为单一节点  $\mathbf{V}$ ,以点集内质心点的特征性质概括该节点。另一方面,某单一点集与其他点集之间的联系取决于两者质心点之间的空间布局,如图 2(b)所示。由于所构建的点集是以质心点为中心,在固定半径内

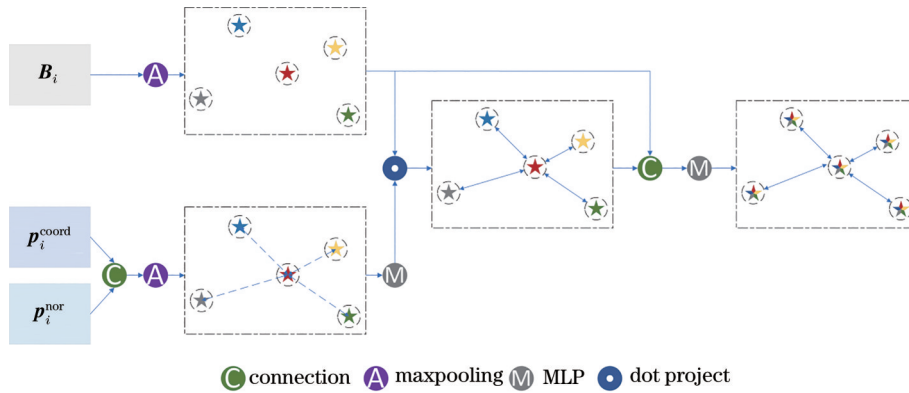


图 4 外-点集模块结构图

Fig. 4 Structure of external point set module



寻找邻点,所以选择点集质心点的法线和位置信息,作为点集之间存在的虚连接线  $E$ 。以上述设定为基础构建图  $G=(V,E)$ ,相关操作可描述为

$$\begin{cases} V = \text{Max} \sum_{j=1}^{l+1} (B_j) \\ E = \text{Max} \sum_{j=1}^{l+1} [p_i^{\text{coord}}, p_i^{\text{nor}}] \end{cases} \quad (9)$$

式(9)使用 MaxPooling 作为本模块的对称函数,获得了邻域内的显著特征,以此实现点云的置换不变性,有效概括了单一点集的整体形状,减少

了冗余信息的影响。上述基础的虚连接仅表示质心点之间的关系,无法与节点(点集)构建起联系。因而,第1步是将虚连接传播至原始节点上,使图  $G$  描述任意节点间存在的实连接。但是模块仅感知到节点之间的关联性是不够的,因为在处理点云数据时,例如飞机,模块仅学习到机翼点集节点与机身点集节点之间存在相关性是不可靠的,两部分之间的交界处依旧会因为差异性分辨的不明确产生分割偏差,所以模块还需继续探究不同节点之间的相异性关系。上述操作可描述为

$$G = \text{MLP}_2 \left[ \text{Max} \sum_{j=1}^{l+1} (B_j), \left\{ \text{MLP}_1 \left( \text{Max} \sum_{j=1}^{l+1} [p_i^{\text{coord}}, p_i^{\text{nor}}] \right) \right\} \cdot \left\{ \text{Max} \sum_{j=1}^{l+1} (B_j) \right\} \right], \quad (10)$$

式中:  $\text{MLP}_1$  用于实现节点间连接关系的实体化,建立节点互通的传播渠道。 $\text{MLP}_2$  的作用在于处理不同节点间的远程依赖联系,并更新至各个节点的状态。外-点集模块通过图  $G$  掌握了原本孤立的点集之间的内在关联性,可推理全局视角下完整的点集形状。

### 3 形状推理卷积网络(SR-Net)架构

对于点云分析任务,形状推理卷积网络(SR-Net)由前端与后端两部分组成,如图5所示。网络的输入是  $N \times 6$  格式的点云数据,其中  $N$  是数据中

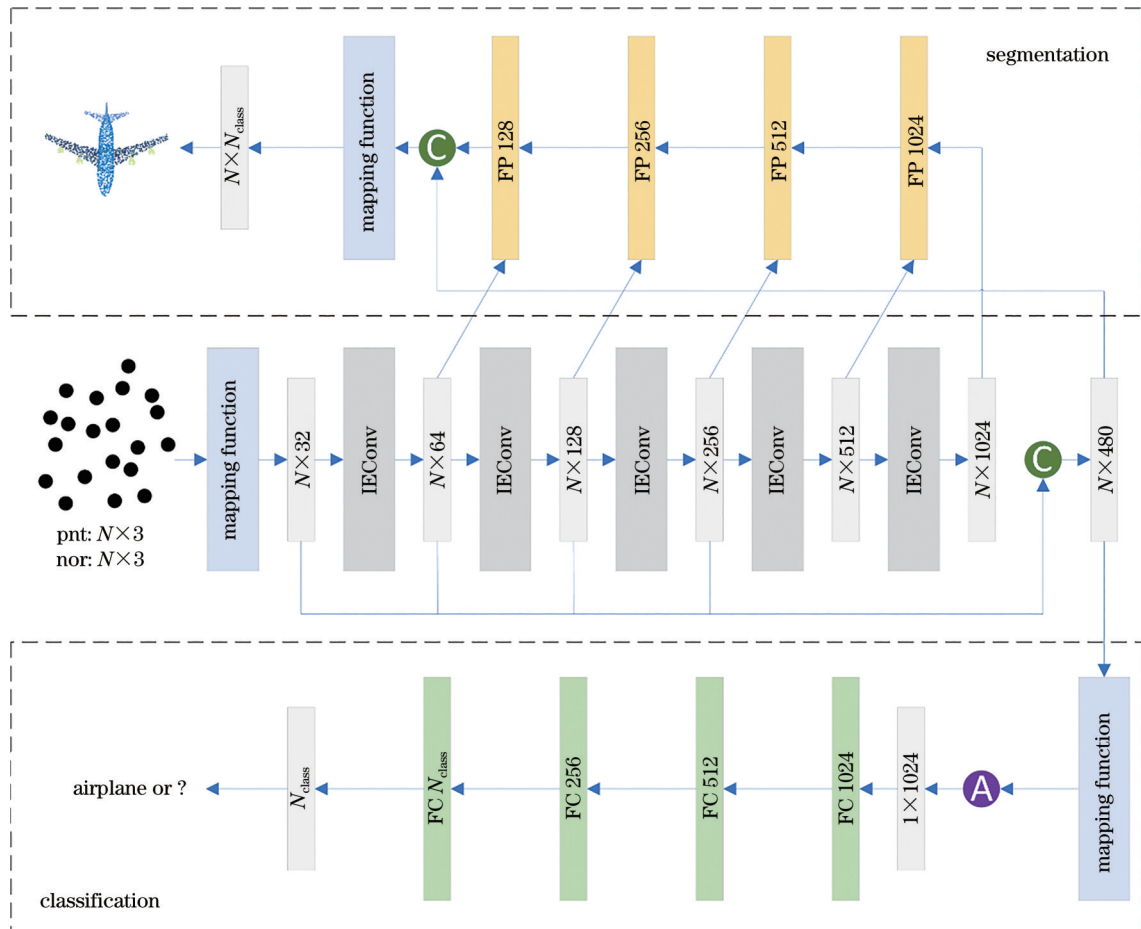


图5 内-外点集形状特征卷积网络架构图

Fig. 5 Architecture of interior-exterior point set shape feature convolutional network

点云的数量,每个点都拥有初始的原始三维坐标特征和三维法线特征。随后利用映射函数提升原始点云的通道数量。再以分层方式嵌入 IE-Conv,通过层级深度的增加,网络不断从点集内部及外部学习更为抽象化的局部与全局形状,实现对复杂点云数据集的分类与分割。同时,为了避免层级过深使网络出现梯度消失,进而影响形状学习准确性的问题。将映射函数的输出及前三层 IE-Conv 的数据聚合起来,避免了网络退化的隐患。

在网络后端,为应对分类与分割的任务需求,设计了不同的分支处理点云数据。在分类分支中,通过映射函数以及对称函数,将前端输出的形状特征处理为 1024 维度的特征向量。随后将其输入至全连接层(fully connected layer),得到分类任务所需的概率图。在分割分支中,使用分层嵌入的五层 IE-Conv 进行密集的形状学习;同时,利用特征传播层(feature propagation)<sup>[8]</sup>将不同层级的输出相互级联进行连续的上采样操作,达到恢复原始点云细节及推理形状向量的目的。同时,利用残差连接将特征传播层与前端聚合输出相融合并一同馈入映射函数,以产生对在对象中某一部分每一点的预测。

## 4 实验

将所提形状推理卷积网络(SR-Net)在 ModelNet40<sup>[24]</sup>和 ShapeNet<sup>[25]</sup>两个公开数据集上进行了相关实验,以 ADAM 作为优化算法。在分类、分割实验中,分别把 Batch-Size 设为 32、24;设置初始学习率为 0.001,同时经过每 30 个 epoch,便乘以 0.5 来衰减学习率;Batch-Norm 设定为 0.6,同时每 20 个 epoch 衰减 0.5。与文献[9,14]相同,选择交叉熵作为模型的损失函数。实验均在一台装有单个 NVIDIA GTX 2080 GPU 的台式计算机上进行,同时所有训练、测试代码都由 Pytorch 实现。

### 4.1 ModelNet40 形状分类任务

ModelNet40<sup>[24]</sup>数据集中包含 12311 个计算机辅助设计(CAD)模型。在本小节中,将数据集分为两部分,9843 个 CAD 用于训练,2468 个 CAD 用于测试。同时,从每个 CAD 模型的表面上均匀地采样 1024 个点,将点的三维坐标及三维法线作为原始属性输入网络。在增强策略中,随机各向异性缩放设置为 $[-0.66, 1.5]$ ,平移范围设置为 $[-0.2, 0.2]$ ,Dropout 设为 50%。在测试阶段,与 PointNet 等研

究策略相同<sup>[8,9]</sup>,通过随机标度进行了 10 次投票测试,并取平均值,得出最终结果。

从表 1 中的准确率(Acc)可以看出,在 ModelNet40 数据集中,形状推理卷积网络(SR-Net)仅选择 1024 个附加三维坐标和法线特征的点作为训练点数,实现了 93.9% 的准确率。从输入点属性的角度来看,在以 pnt 与 nor 为点属性的方法中,相比于 PAT<sup>[27]</sup>(91.7%)在准确率上提高了 2.2 个百分点;相比于 A-CNN(92.6%)在准确率上提高了 1.3 个百分点。即使是以 pnt 为输入点属性的方法中,准确率的提高也较明显:相比于 PointNet(89.2%)在准确率上提高了 4.7 个百分点;相比于 DGCNN(92.2%)在准确率上提高了 1.7 个百分点。从输入点个数的角度来看,所提方法以更少的点,取得了更高的准确率:相比于 PointNet++(5000,91.9%)在准确率上提高了 2.0 个百分点;相比于 SO-Net<sup>[29]</sup>(5000,92.4%)在准确率上提高了 1.5 个百分点。

为了进一步验证 SR-Net 的性能,对复杂度、运行时间和浮点运算数/样本(FLOPs/sample)进行了比较,结果如表 2 所示。从表 2 可以看出,SR-Net 在

表 1 ModelNet40 数据集的分类准确率比较

Table 1 Comparison of classification accuracy for ModelNet40 dataset

Method	Input	Points / 10 <sup>3</sup>	Acc / %
Pointwise-CNN <sup>[31]</sup>	pnt	1	86.1
ECC <sup>[19]</sup>	pnt	1	87.4
PointNet <sup>[8]</sup>	pnt	1	89.2
Point-CNN <sup>[33]</sup>	pnt	1	91.7
DGCNN <sup>[10]</sup>	pnt	1	92.2
SO-CNN <sup>[15]</sup>	pnt	1	93.1
Dense-Point <sup>[26]</sup>	pnt	1	93.2
RS-CNN <sup>[14]</sup>	nor	1	92.8
PAT <sup>[27]</sup>	pnt, nor	1	91.7
Spec-GCN <sup>[30]</sup>	pnt, nor	1	91.8
PointConv <sup>[11]</sup>	pnt, nor	1	92.5
A-CNN <sup>[11]</sup>	pnt, nor	1	92.6
PointASNL <sup>[13]</sup>	pnt, nor	1	93.2
ELM <sup>[28]</sup>	pnt, nor	1	93.2
RS-CNN <sup>[14]</sup>	pnt, nor	1	93.6
SO-Net <sup>[29]</sup>	pnt, nor	2	90.9
PointNet++ <sup>[9]</sup>	pnt, nor	5	91.9
Spider-CNN <sup>[32]</sup>	pnt, nor	5	92.4
SO-Net <sup>[29]</sup>	pnt, nor	5	93.4
Proposed method	pnt, nor	1	93.9

表 2 ModelNet40 数据集的分类复杂度与运行时间比较

Table 2 Comparison of classification complexity and time of the ModelNet40 dataset

Method	Number of parameters /MB	Acc /%	FLOPs /sample
PointNet	3.50	89.2	440
Spec-GCN	2.05	91.8	1112
PointNet++	1.48	91.9	1684
DGCNN	1.84	92.2	2767
Proposed method	<b>1.37</b>	<b>93.9</b>	<b>266</b>

模型参数量较少的前提下,实现了更高的准确率。例如,与 PointNet(3.50 MB, 89.2%)相比,SR-Net 的参数量仅约为其二分之一,而准确率却提高了 4.7 个百分点;与 PointNet++(1.48 MB, 91.9%)相比,虽然在参数量上只降低了 0.11 MB,但是准确率却提高了 2.0 个百分点。与 Spec-GCN<sup>[30]</sup>(1112 MB, 91.8%)相比,SR-Net 的 FLOPs/sample 仅约为其五分之一,准确率却提高了 2.1 个百分点。SR-Net 与 DGCNN(2767, 92.2%)相比, FLOPs/sample 仅约为其十分之一,准确率却提高了 1.7 个百分点。上述实验结果表明,SR-Net 在较少的参数、较低的计算成本条件下,达到了更高的分类精度。

#### 4.2 ShapeNet 形状分割任务

在 ShapeNet 标准数据集对 SR-Net 进行分割实验。在这项富有挑战性的实验中,数据集的任一点都需要被分类为预先定义好的形状标签。ShapeNet 数据集拥有 16881 个三维对象,囊括了 16 个类别,同时分别注释了 50 个形状。此外,任意形状由 2~6 个部分组成。依据文献[8],将

ShapeNet 数据集分为 12137 个训练对象,1870 个验证对象,2874 个测试对象,并设定每一形状含有 2048 个点。最后,选择平均交并比为形状的平均评估指标。






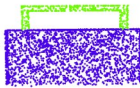


























表 3 为不同方法的分类结果。和一些典型的以法线和坐标作为输入点属性的方法相比,例如 SO-Net<sup>[29]</sup>(84.9%)、PointNet++(85.1%)、ELM<sup>[28]</sup>(85.3%),所提方法有了显著的提高(+1.5 个百分点、+1.3 个百分点、+1.1 个百分点)。特别的,SR-Net 在包、椅子、灯、杯子、桌子的形状分割中,取得了较好的表现。另一方面,表 4 展示了 SR-Net 在分割数据集中的可视化结果。从表 4 可以看出,点云数据集中隐藏有丰富的形状特征,同时一些对象中的类别构成部分较为复杂,例如轿车和摩托,将形状特征完整提取出来较为困难。但是 SR-Net 依旧可以鲁棒地、区别性地将其分割出来。究其原因,SR-Net 不仅通过内-点集模块挖掘局部区域内的细微关键特征,帮助模型对不同点云分属部分进行区分,同时通过外-点集模块在全局视角下,探究不同部分之间差异性边缘。

表 3 ShapeNet 数据集的分割准确率比较

Table 3 Comparison of segmentation accuracy of ShapeNet dataset

Method	Air	Bag	Cap	Car	Chair	Ear.	Gui.	Knife	Lamp	Lap	Moto	Mug	Pistol	Rock	Skate	Table	Mean
PointNet <sup>[8]</sup>	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	83.7
SONet <sup>[29]</sup>	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	84.9
PointNet++ <sup>[9]</sup>	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	85.1
DGCNN <sup>[10]</sup>	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	80.6	85.1
PCNN <sup>[11]</sup>	82.4	80.1	85.5	79.5	90.8	73.2	91.3	86.0	85.0	95.7	73.2	94.8	83.3	51.0	75.0	81.8	85.1
ELM <sup>[28]</sup>	84.0	80.4	88.0	<b>80.2</b>	90.7	77.5	91.2	86.4	82.6	95.5	70.0	93.9	<b>84.1</b>	55.6	75.6	82.1	85.3
SpiderCNN <sup>[32]</sup>	83.5	81.0	87.2	77.5	90.7	76.8	91.1	87.3	83.3	95.8	70.2	93.5	82.7	59.7	75.8	82.8	85.3
SO-CNN	83.9	84.1	85.0	77.4	91.3	78.3	<b>91.7</b>	87.4	83.8	<b>96.4</b>	69.7	93.5	83.1	58.9	76.2	82.9	85.7
A-CNN <sup>[11]</sup>	<b>84.2</b>	84.0	88.0	79.6	91.3	75.2	91.6	87.1	85.5	95.4	<b>75.3</b>	94.9	82.5	<b>67.8</b>	77.5	83.3	86.1
RS-CNN <sup>[14]</sup>	83.5	84.8	<b>88.8</b>	79.6	91.2	<b>81.1</b>	91.6	<b>88.4</b>	86.0	96.0	73.7	94.1	83.4	60.5	<b>77.7</b>	83.6	86.2
Proposed method	84.0	<b>86.2</b>	88.1	79.5	<b>91.6</b>	77.5	91.3	88.0	<b>86.3</b>	96.1	72.8	<b>95.0</b>	83.6	62.2	75.9	<b>83.9</b>	<b>86.4</b>

表 4 SR-Net 在 ShapeNet 数据集的可视化结果  
Table 4 Visualization results of SR-Net on ShapeNet dataset

Object	Visualization		Object	Visualization	
Air			Lamp		
Bag			Laptop		
Cap			Motor.		
Car			Mug		
Chair			Pistol		
Ear.			Rocket		
Gui.			Skate		
Knife			Table		

### 4.3 消融实验

为了进一步验证在点云分析任务中,SR-Net 各模块的必要性,在 ModelNet40 数据集上进行广泛的消融研究,结果如表 5 所示。A 是基线模型,准确率只有 90.1%;B 是另一参考模型,它是 RS-CNN 的改进型,由双层感知机组成,准确率为 92.0%;在添加了内-点集模块、外-点集模块后,模型 C、D 的准确率分别提高了 3.1 个百分点、0.9 个百分点,达到了 93.2%、92.9%。双边设计的 IE-Conv 准确率也随之提升至 93.9%。实验结果表明,利用门控机制强化点集内对关键形状的筛选能大幅度地提高分类精度,同时利用图卷积整合点集自身的特征及其之间的依赖联系能减少对错误形状的区别。

此外,对内-点集模块的核心如何选择最为合适

表 5 ModelNet40 数据集上对模块的消融实验

Table 5 Ablation experiments with modules on ModelNet40 dataset

Model	RS-CNN(*)	Internal	External	Internal-external	Acc / %
A					90.1
B	✓				92.0
C		✓			93.2
D	✓		✓		92.9
E				✓	93.9

的初始表达式作为门进行自适应学习这一关键问题进行了探讨,实验结果如表 6 所示。模型 A 仅使用了欧几里得距离和余弦距离作为几何门和法线门来提取相应两点间关联特征,得到了 92.4% 的准



表 6 ModelNet40 数据集上对不同先验表达式作为门的消融实验

Table 6 Ablation experiments on ModelNet40 dataset for different prior expressions as gates

Model	Priori expressions	Channel	Acc / %
A	$\ p_i - p_{i,j}\ _{ed}$	1	92.4
	$\ p_i - p_{i,j}\ _{cosd}$	1	
B	$p_i^{coord} \setminus p_{i,j}^{coord}$	6	93.4
	$p_i^{nor} \setminus p_{i,j}^{nor}$	6	
C	$p_i^{coord} \setminus p_{i,j}^{coord} \setminus \ p_i - p_{i,j}\ _{ed}$	7	93.9
	$p_i^{nor} \setminus p_{i,j}^{nor} \setminus \ p_i - p_{i,j}\ _{cosd}$	7	
	$p_i^{coord} \setminus p_{i,j}^{coord} \setminus \ p_i - p_{i,j}\ _{ed}$	7	
D	$p_i^{coord} \setminus p_{i,j}^{coord} \setminus \ p_i - p_{i,j}\ _{ed}$	7	93.0
	$p_i^{nor} \setminus p_{i,j}^{nor} \setminus \ p_i - p_{i,j}\ _{cosd}$	7	
E	$\ p_i - p_{i,j}\ _{cosd}$	7	92.5

准确率。而在模型 B 中,选择了采样点与邻点的坐标、法线,准确率为 93.4%。模型 C,作为 A 与 B 的升级版,准确率也升高为 93.9%。A、B、C 之间准确率不同的原因在于丰富的点间特征能够有效帮助门来选择出最能够合理描述单一点集的局部形状。另一方面,模型 D(93.0%)和模型 E(92.5%)均只从几何方面或法线方面,单独探寻质心点与邻点的联系,而模型 C 与模型 D、E 之间的正确率差异(+0.9%,+1.4%)也从侧面印证了内-点集模块中门控设计的必要性及其可从不同角度对点间特征进行自适应学习的有效性。

针对内-点集模块中自校准机制的可靠性进行了定量分析,结果如表 8 所示。装备了自校准功能的模型 C 和模型 E,它们的准确率达到 93.5% 和 93.9%。而与上述两者的基线:模型 B 与模型 D,在性能上都有了提高(+0.3 个百分点,+0.4 个百分点)。原因在于自我校准能够通过质心点坐标及其和邻点的坐标差来预测偏差,使网络能够降低点云对刚性变换的敏感度,进而利用点云自身属性生成门对修正后的逐点特征进行筛选,最终鲁棒地学习到任务所必要的形状特征。

为进一步验证 SR-Net 对点云置换和刚性变换的鲁棒性,在未进行数据增强的条件下进行了训练,并在测试阶段进行了平移(-0.2,+0.2)、逆时针旋转(90°、180°)处理,结果如表 7 所示。PointNet

表 7 ModelNet40 数据集上对点云添加平移或旋转后的鲁棒性实验

Table 7 Robustness experiments on ModelNet40 dataset after adding translations or rotations to point clouds

Method	Self-calibrate	Translation		Rotate	
		-0.2	+0.2	90°	180°
PointNet		70.8	70.6	42.5	38.6
PointNet++		88.2	88.2	47.9	39.7
Proposed method		90.9	90.9	90.9	90.9
Proposed method	✓	<b>92.1</b>	<b>92.1</b>	<b>92.1</b>	<b>92.1</b>

对平移或旋转都较为敏感(70.8/70.6/42.5/38.6),作为其升级版 PointNet++ 虽然减少了对平移扰动的影响,但是依旧无法鲁棒地处理旋转问题(88.2/88.2/47.9/39.7)。SR-Net(90.9/90.9/90.9/90.9,未添加自校准机制)即使添加了不同的平移或旋转干扰后,相较于 PointNet(+20.1/+20.3/+48.4/+52.3)和 PointNet++(+2.7/+2.7/+43.0/+51.2)均有显著提升,依旧能够鲁棒地从点云数据中感知形状。究其原因,SR-Net(未添加自校准机制)选择各点的几何关系和法线关系作为各点相关性特征的先验知识,很好地缓解了可能会出现扰动,此外门控机制也避免了模型在错误条件下学习形状。在添加了自校准机制后,SR-Net 的鲁棒性(92.1/92.1/92.1/92.1)进一步提升(+1.2/+1.2/+1.2/+1.2)。该实验结果表明,自校准机制通过在单位点集中利用质心点及其邻点的结构信息修正偏差及选择各点集质心的空间布局联系优化局部特征,能够显著提高网络的鲁棒性。

同时,表 8 也总结了在内-点集模块中,不同的聚合函数对分类任务准确率影响。可以发现使用平均池化(Avg)和最大池化(Max),准确率分别达到了 92.7% 和 93.2%。而求和(Sum)正确率相比前两者都有较大的提高,达到了 93.5%。另外,聚合函数 Summation 能够让自校准机制发挥更好的效

表 8 ModelNet40 数据集上不同聚合函数及自我校准功能的消融实验

Table 8 Ablation experiments of different aggregation functions and self calibration functions on ModelNet40 dataset

Model	Aggregation function	Self-calibrate	Acc / %
A	Avg		92.7
B	Max		93.2
C	Max	✓	93.5
D	Sum		93.5
E	Sum	✓	93.9

果。原因是,使用求和作为聚合函数,相比于 Max 来说,能够很好地避免点与点之间关键特征的损失。

## 5 结 论

提出了一种双边架构的 IE-Conv。在内点集模块中,IE-Conv 利用门控机制有区别性地筛选合适的点间关系自适应地挖掘单一点集所代表的局部形状,同时在学习的过程中辅以自校准功能,提高对模块自身的鲁棒性。在外点集中,通过不同的点集与其质心点之间的特征差构建图,推理出不同点集之间的远程依赖关系,传播并更新各点集状态。以 IE-Conv 为基础,分层构建的形状推理卷积网络(SR-Net),在 ModelNet 分类任务和 ShapeNet 分割任务中取得优异的成绩。在分类实验中,SR-Net 的准确率比 DGCNN、A-CNN 多 1.7 个百分点、1.3 个百分点,同时以更少的参数量比 PointNet++ 多 2 个百分点的准确率。在分割任务中,以平均交并比作为评价指标,相较于 DGCNN、ELM,分别高出 1.3 个百分点、1.1 个百分点。同时,在包、椅子、灯、杯子、桌子的形状分类中取得了不错的成绩。上述实验结果都有效验证了 SR-Net 在点云分析任务中的优秀性能。

## 参 考 文 献

- [1] Atzmon M, Maron H, Lipman Y. Point convolutional neural networks by extension operators[J]. *ACM Transactions on Graphics*, 2018, 37(4): 1-12.
- [2] Chen Y L, Liu S, Shen X Y, et al. Fast point R-CNN[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 9774-9783.
- [3] Choy C, Gwak J, Savarese S. 4D spatio-temporal ConvNets: minkowski convolutional neural networks [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 3070-3079.
- [4] 侯向丹,于习欣,刘洪普.基于图卷积网络的三维点云分类分割模型[J].*激光与光电子学进展*, 2020, 57(18): 181019.  
Hou X D, Yu X X, Liu H P. 3D point cloud classification and segmentation model based on graph convolutional network[J]. *Laser & Optoelectronics Progress*, 2020, 57(18): 181019.
- [5] 卢俊鑫,方志军,陈婕妤,等.点线特征结合的 RGB-D 视觉里程计[J].*光学学报*, 2021, 41(4):0415001.  
Lu J X, Fang Z J, Chen J Y, et al. RGB-D visualodometry combined with points and lines[J]. *ActaOptica Sinica*, 2021, 41(4): 0415001.
- [6] 戴仁月,方志军,高永彬.融合扩张卷积网络与 SLAM 的无监督单目深度估计[J].*激光与光电子学进展*, 2020, 57(6): 061007.  
Dai R Y, Fang Z J, Gao Y B. Unsupervised monocular depth estimation by fusing dilated convolutional network and SLAM[J]. *Laser & Optoelectronics Progress*, 2020, 57(6): 061007.
- [7] 刘航,汪西莉.基于注意力机制的遥感图像分割模型[J].*激光与光电子学进展*, 2020, 57(4): 041015.  
Liu H, Wang X L. Remote sensing image segmentation model based on attention mechanism[J]. *Laser & Optoelectronics Progress*, 2020, 57(4): 041015.
- [8] Charles R Q, Hao S, Mo K C, et al. PointNet: deep learning on point sets for 3D classification and segmentation[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 77-85.
- [9] Qi C R, Yi L, Su H, et al. Pointnet++: deep hierarchical feature learning on point sets in a metric space[EB/OL]. (2017-06-07) [2021-05-06]. <https://arxiv.org/abs/1706.02413>.
- [10] Wang Y, Sun Y B, Liu Z W, et al. Dynamic graph CNN for learning on point clouds[J]. *ACM Transactions on Graphics*, 2019, 38(5): 1-12.
- [11] Komarichev A, Zhong Z C, Hua J. A-CNN: annularly convolutional neural networks on point clouds[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 7413-7422.
- [12] Liu X H, Han Z Z, Liu Y S, et al. Point2Sequence: learning the shape representation of 3D point clouds with an attention-based sequence to sequence network [J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, 33: 8778-8785.
- [13] Yan X, Zheng C D, Li Z, et al. PointASNL: robust point clouds processing using nonlocal neural networks with adaptive sampling[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 5588-

- 5597.
- [14] Liu Y C, Fan B, Xiang S M, et al. Relation-shape convolutional neural network for point cloud analysis [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 8887-8896.
- [15] Zhang C Y, Song Y, Yao L N, et al. Shape-oriented convolution neural network for point cloud analysis [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12773-12780.
- [16] Wang X L, Girshick R, Gupta A, et al. Non-local neural networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 7794-7803.
- [17] Fu J, Liu J, Tian H J, et al. Dual attention network for scene segmentation[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 3141-3149.
- [18] Zhang H, Goodfellow I J, Metaxas D N, et al. Self-attention generative adversarial networks[C]// Proceedings of the 36th International Conference on Machine Learning, June 9-15, 2019, Long Beach, California, USA. London: PMLR, 2019: 7354-7363.
- [19] Simonovsky M, Komodakis N. Dynamic edge-conditioned filters in convolutional neural networks on graphs[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 29-38.
- [20] Wang L, Huang Y C, Hou Y L, et al. Graph attention convolution for point cloud semantic segmentation[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 10288-10297.
- [21] Li G H, Müller M, Thabet A K, et al. DeepGCNs: can GCNs go as deep as CNNs? [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 9266-9275.
- [22] 王旭娇, 马杰, 王楠楠, 等. 基于图卷积网络的深度学习点云分类模型[J]. 激光与光电子学进展, 2019, 56(21): 211004.
- Wang X J, Ma J, Wang N N, et al. Deep learning model for point cloud classification based on graph convolutional network[J]. Laser & Optoelectronics Progress, 2019, 56(21): 211004.
- [23] 王江安, 何娇, 庞大为. 基于动态图卷积网络的点云分类和分割网络[J]. 激光与光电子学进展, 2021, 58(12): 1215008.
- Wang J G, He J, Pang D W. Point cloud classification and segmentation network based on dynamic graph convolutional network[J]. Laser & Optoelectronics Progress, 2021, 58(12): 1215008.
- [24] Wu Z R, Song S R, Khosla A, et al. 3D ShapeNets: a deep representation for volumetric shapes[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition, June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 1912-1920.
- [25] Chang A X, Funkhouser T, Guibas L, et al. ShapeNet: an information-rich 3D model repository [EB/OL]. (2015-12-09) [2021-05-08]. <https://arxiv.org/abs/1512.03012>.
- [26] Liu Y C, Fan B, Meng G F, et al. DensePoint: learning densely contextual representation for efficient point cloud processing[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 5238-5247.
- [27] Yang J C, Zhang Q, Ni B B, et al. Modeling point clouds with self-attention and gumbel subset sampling [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 3318-3327.
- [28] Fujiwara K, Hashimoto T. Neural implicit embedding for point cloud analysis[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 11731-11740.
- [29] Li J X, Chen B M, Lee G H. SO-net: self-organizing network for point cloud analysis[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 9397-9406.
- [30] Wang C, Samari B, Siddiqi K. Local spectral graph convolution for point set feature learning[EB/OL]. (2018-03-15) [2021-05-07]. <https://arxiv.org/abs/1803.05827>.

- [31] Hua B S, Tran M K, Yeung S K. Pointwise convolutional neural networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 984-993.
- [32] Xu Y F, Fan T Q, Xu M Y, et al. Spidercnn: deep learning on point sets with parameterized convolutional filters[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11212: 90-105.
- [33] Li Y Y, Bu R, Sun M C, et al. Pointcnn: convolution on X-transformed points[EB/OL]. (2018-01-23)[2021-05-22]. <https://arxiv.org/abs/1801.07791>.