

结合空间注意力与形状特征的三维点云语义分割

郝雯^{1,2*}, 王红霄^{1,2}, 汪洋^{1,2}

¹西安理工大学计算机科学与工程学院, 陕西 西安 710048;

²陕西省网络计算与安全技术重点实验室, 陕西 西安 710048

摘要 目前已有的基于深度学习的点云分割方法比较关注点云全局或局部特征的提取, 忽略了点云的形状信息及点间语义特征的提取。针对已有方法的不足, 提出一种基于空间注意力机制的多特征融合动态图卷积神经网络。首先, 在提取点云局部边缘几何特征的基础上, 将点云的低维几何特征映射到高维特征空间, 以获取点云丰富的形状信息, 并利用多层感知机提取点的全局高维特征。然后, 引入一种空间注意力机制, 提取点间的语义特征。最后, 将几何形状特征与高层语义特征进行有效的融合, 丰富点云的全局、局部特征表征。在多个公开数据集上对所提网络模型进行了实验, 实验结果表明, 所提网络模型在目标分类、部件分割、语义分割任务中, 均取得了较好的性能。

关键词 遥感; 动态图卷积; 形状特征; 空间注意力; 多特征融合

中图分类号 P236

文献标志码 A

doi: 10.3788/LOP202259.0828004

Semantic Segmentation of Three-Dimensional Point Cloud Based on Spatial Attention and Shape Feature

Hao Wen^{1,2*}, Wang Hongxiao^{1,2}, Wang Yang^{1,2}

¹School of Computer Science and Engineering, Xi'an University of Technology, Xi'an, Shaanxi 710048, China;

²Shaanxi Key Laboratory for Network Computing and Security Technology, Xi'an, Shaanxi 710048, China

Abstract Existing segmentation methods based on deep learning focus only on the global or local feature extractions of the point cloud; they ignore the shape information and semantic features between points. We propose a multifeature fusion dynamic graph convolutional neural network based on spatial attention to solve the challenges mentioned above. First, on the basis of edge geometric feature extraction, the low-dimensional geometric features of point cloud are mapped to the high-dimensional feature space to obtain the rich shape information. The multilayer perceptron is used to extract the global high-dimensional features of points. Then, a spatial attention mechanism is introduced to extract contextual semantic features between points. Finally, geometric features and high-level semantic information are effectively fused to enrich the representation of global and local features. The proposed network is tested on several public datasets. Experimental results show that the proposed network has achieved superior performance in object classification, part segmentation, and semantic segmentation.

Key words remote sensing; dynamic graph convolution; shape feature; spatial attention; multifeature fusion

1 引言

近年来,随着三维(3D)传感器的广泛应用及激光扫描技术的快速发展,人类能够快速地获取场景

以及物体的点云数据。点云作为3D数据表示形式之一,能够较好地表达复杂场景及物体的几何形状,并且在物体空间关系及拓扑关系的表达方面具有独特的优势,已经被广泛地应用到自动驾驶、室内导

收稿日期: 2021-08-02; 修回日期: 2021-09-02; 录用日期: 2021-09-13

基金项目: 国家自然科学基金(61602373)、陕西省自然科学基金(2021JM-342)、陕西省教育厅重点实验室基金(18JS078)

通信作者: haowensxs@163.com

航、数字文化遗产保护等多个领域。

点云的语义分割作为 3D 场景理解和重建的基础,是点云数据处理和分析中的重要步骤。基于区域增长、基于聚类及基于模型拟合等传统的点云分割方法,往往需要手动设定特征,并需要手工调参,极大限制了点云的广泛应用。随着深度学习在二维(2D)图像的物体识别^[1]及语义分割^[2]等多个领域的成功应用,基于深度学习的点云分割方法^[3-4]也成为研究的热点,并取得了显著的成果。目前已有的基于深度学习的 3D 点云分割或分类方法可以分为 3 类:基于多层感知机(mlp)的方法、基于卷积的方法及基于图的方法^[5]。

基于多层感知机的方法首先通过堆叠共享的多层感知机来提取单个点的特征,然后使用对称聚合函数得到全局特征。PointNet^[6]开创性地以原始点云作为输入,引入转换网络进行空间对齐,使用多层感知机学习点特征,然后使用最大池化函数提取全局特征,但是该网络对点云数据局部特征的提取能力不足。为了进一步提升点云的语义分割性能,Qi 等^[7]提出 PointNet++,首先利用最远点采样法对输入点云进行采样和区域划分,然后在局部区域应用 PointNet 进行特征学习与提取。受 2D 图像形状描述符 scale-invariant feature transform(SIFT)的启发,Jiang 等^[8]设计了 PointSIFT 模块,该模块利用方向编码卷积提取点在八个方向的特征,进而将几个方向的特征组合形成多尺度特征,其采样模块与 PointNet++ 结构一致。基于多层感知机的方法往往忽略了点之间的相关性,缺乏捕捉点云局部几何特征的能力。尚鹏飞等^[9]在 PointNet 的基础上增加一个估计法矢量的模块,提出了一种基于点云法矢量的语义分割网络。

基于卷积的方法的核心思想是根据邻域点之间的空间位置关系学习点之间的权重参数,并根据学习到的权重参数自适应地聚合局部特征。由于卷积操作对于点的输入顺序敏感,point convolutional neural network(PointCNN)^[10]通过 χ -Conv 算子将输入点重排列为潜在的规范顺序,然后对转换后的特征应用卷积运算符,但该网络模型计算复杂度较高。Pointwise^[11]利用逐点卷积(pointwise convolution)获取点的局部特征信息实现语义分割,但是该网络需要对点云进行体素化,对每个栅格中的点设置相同的权重。在 PointConv^[12]中,卷积被定义为连续 3D 卷积的蒙特卡罗估计,卷积核由加权函数和密度

函数组成。为了提高计算效率,3D 卷积被简化为矩阵乘法和 2D 卷积运算。基于点卷积的方法能够有效提取点云的局部几何特征,但缺乏对点云形状信息进行提取的能力。

基于图的方法将点云中的每个点视为图的顶点,并根据每个点的邻域点构造有向边,通过在邻域点上应用多层感知机来实现卷积操作,并采用池化操作聚合来自邻域点的信息,通过在顶点之间学习边的权重以更新顶点的特征。Dynamic graph CNN(DGCNN)^[13]在特征空间构造图,并且逐层动态更新图结构。Edgeconv^[13]的核心是对边特征应用多层感知机进行特征学习,然后对与中心点相连的边特征使用通道级的对称聚合函数来获取点云的局部特征。然而,转换网络的使用增加了网络结构的复杂度和训练的参数量。为了显示建模局部区域点间的几何结构,GeoCNN^[14]将 GeoConv 应用于点的邻域,基于边特征的分解和聚合获取点间的局部几何关系。Linked dynamic graph CNN(LDGCNN)^[15]在 DGCNN 的基础上,将各层动态图的输出进行连接,有效解决了梯度消失问题,并且在训练后固定特征提取器,重复训练分类器以提升网络的性能,但该网络在语义分割方面效果并不明显。基于图的方法关注点间的关系,但对具有抽象语义信息的高级全局结构特征提取不足。

上述网络取得了一定的阶段性成果,但是这些方法主要关注点云数据全局或局部区域特征的提取,没有充分挖掘点云的语义特征信息;且缺乏对形状特征的感知能力,对每个点潜在的几何形状特征提取能力不足。而点的形状几何信息及点间的语义信息对于物体分类及分割也至关重要。因此,本文提出一种基于空间注意力机制的多特征融合动态图卷积神经网络(AMFF-DGCNN)。该网络具有以下特点:1)空间转换网络的去除降低了网络的复杂性、减小了计算量。2)在提取局部边缘特征的基础上,点的形状信息及点的全局特征的融入提高了网络模型的几何形状感知能力。3)空间注意力机制的引入可以捕捉点之间的相关性,挖掘全局单点特征隐含的深层次语义信息。

2 基于空间注意力的多特征融合动态图卷积神经网络

2.1 AMFF-DGCNN 框架

AMFF-DGCNN 的网络模型如图 1 所示,该网络以 $N \times D$ 的原始点云作为输入,其中 N 代表的是

点的个数, D 代表的是点的数据维度。图 1 上层为分类网络, 下层为分割网络。

分类网络包括 4 层特征提取模块, 每 1 层均由提取局部几何特征的 edgeconv 模块和利用空间注意力机制提取点间语义特征的 spatial attention 模块组成。上 1 层提取的局部几何特征与点间语义特征 point feature 拼接后作为下一层 edgeconv 模块的输入, $\{\}$ 中的数字代表特征的输出维度。每 1 层中

edgeconv 模块提取的局部几何特征和 spatial attention 所提取的语义特征的输出维度相同, 分别为 64, 64, 64, 128。对各层提取的特征进行拼接, 并将得到的维度为 $n \times 640$ 的融合特征输入到一个共享的全连接层(1024)后, 通过最大池化操作可得到 1×1024 维全局特征描述符, 最后使用三个全连接层(512, 256, k)即可得到目标所属类别的 k 维向量。

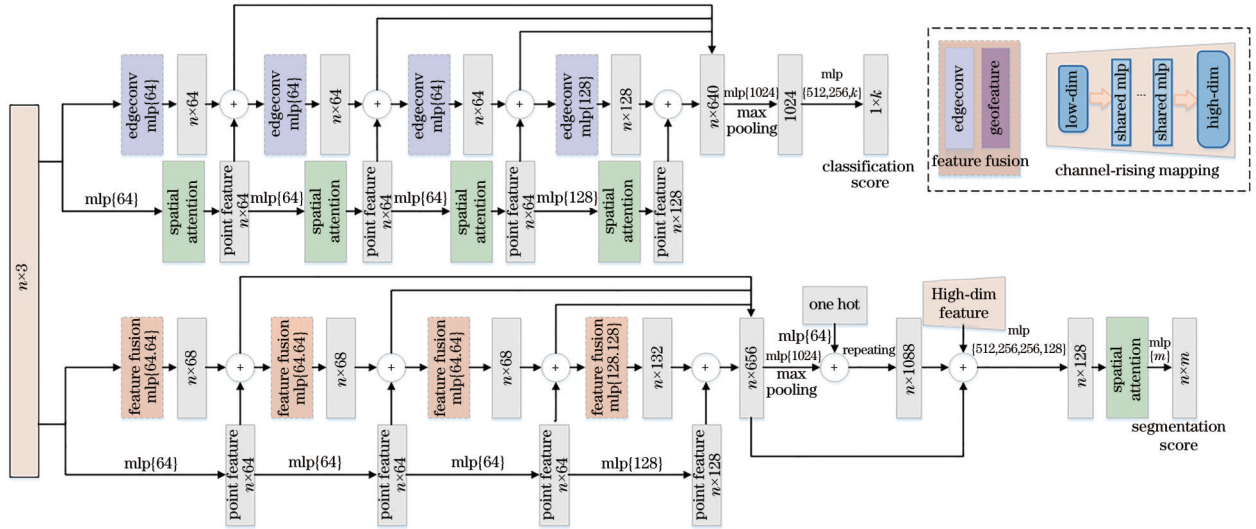


图 1 AMFF-DGCNN 结构
Fig. 1 AMFF-DGCNN structure

分割网络也包括 4 层特征提取模块。将 edgeconv 模块提取局部几何特征及低维几何特征的模块作为 feature fusion 模块, 该模块旨在提高网络的局部特征提取能力; 同时, 使用 mlp 直接对点云进行操作, 以提取点的全局特征, 丰富特征表达。每 1 层局部几何特征的输出维度与全局特征的输出维度相同, 分别是 64, 64, 64, 128。将每 1 层提取的全局、局部特征进行拼接可得到维度为 $n \times 656$ 的融合特征, 进一步对融合特征使用卷积操作(1024)和最大池化操作即可得到全局特征描述符。同时, 将低维几何特征映射到高维特征空间, 挖掘点云隐含在高维特征中潜在的几何形状信息, 并与融合特征、全局特征进行拼接, 然后使用全连接层(512, 256, 256, 128)进行转换, 并使用空间注意力机制挖掘空间语义信息, 最后使用一个全连接层(m)输出每个点属于 m 个类别的概率。

2.2 几何特征提取模块

所提 AMFF-DGCNN 在提取点云边缘特征的基础上, 提取了点云的形状特征及点的全局特征,

下面分别对这几个特征进行介绍。

2.2.1 平均对称边缘特征提取

对于任意一点 P_i , 寻找其 k 近邻点 $\{P_1, P_2, \dots, P_k\} \subset \mathbf{R}^3$, 连接近邻点构成有向图。本实验组基于 DGCNN 提出了一种平均对称边缘特征 e_{ij} 。

$$e_{ij} = g_\phi(P_i, P_j - P_i) / 2, \quad (1)$$

式中: ϕ 是一系列可学习的参数; g_ϕ 是非线性函数, 采用多层感知机实现。将中心点的坐标、中心点与邻域点之间的相对位置信息映射到高维特征空间, 从高维特征空间中学习点之间的几何相关性, 从而获取点云的边缘特征。边缘特征 e_{ij} 在获取中心点局部邻域信息的同时, 编码了中心点的位置信息。在提取到点间的局部邻域信息后, 本实验组使用最大池化操作作为聚合函数, 在选取最重要的特征的同时移除了冗余信息的干扰、降低了网络中的参数量。相比于 DGCNN 中的边缘函数, 基于平均的方式能够有效地减少冗余的特征信息, 提升分割精度。

2.2.2 形状特征提取

通过观察可以发现:许多人造物体多是由基本形状组成的,构成物体的点可以大致分为线状点、面状点和散状点。因此,本实验组通过分析每个点与邻域点的局部结构得到每个点的空间分布特征。对于任意一点 p_i ,寻找 p_i 的 k 个邻域点,建立点的局部邻域图 G ,构造协方差矩阵 M 。

$$M = \frac{1}{k} \sum_{i \in G} (p_i - \bar{p})(p_i - \bar{p})^T, \quad (2)$$

式中: \bar{p} 为点 p_i 的 k 个邻近点的平均位置,即 $\bar{p} = \frac{1}{k} \sum_{i=1}^k p_i$ 。通过奇异值分解对协方差矩阵 M 进行特征值分解,得到 M 的特征值 $\lambda_1^i, \lambda_2^i, \lambda_3^i, \lambda_1^i \geq \lambda_2^i \geq \lambda_3^i \geq 0$ 。计算点 p_i 的线状特征 L_λ 、面状特征 P_λ 、球形特征 S_λ 、各向异性 A_λ ,并将其作为低维几何特征 f_{low} ^[16]:

$$\left\{ \begin{array}{l} L_\lambda = \frac{\lambda_1^i - \lambda_2^i}{\lambda_1^i} \\ P_\lambda = \frac{\lambda_2^i - \lambda_3^i}{\lambda_1^i} \\ S_\lambda = \frac{\lambda_3^i}{\lambda_1^i} \\ A_\lambda = \frac{\lambda_1^i - \lambda_3^i}{\lambda_1^i} \end{array} \right. \quad (3)$$

低维度的几何特征对形状信息表达能力较弱,为了更大程度挖掘高维的几何特征信息,所提网络将低维的几何特征映射到高维特征空间^[17],以便获取更丰富的形状信息,从而得到具有形状感知鉴别能力的高维几何特征。所提网络使用 mlp 作为特征映射函数,最大池化作为聚合函数,以保证点的排列不变性。

$$f_{high} = \max [mlp(f_{low})]。 \quad (4)$$

2.2.3 点特征提取

边缘特征和形状特征都是通过构造某点的局部邻域图提取局部特征的,在建立图结构时,将整个点云划分为一个个独立的分组,该操作在一定程度上忽略了点云之间的相关性,对全局特征提取有所欠缺。为了弥补所提网络全局特征提取能力不足的缺陷,直接对原始的输入点云数据应用多层感知机,以提取高维全局特征,为提高点云分割的准确性奠定基础。

2.3 空间注意力模块

点间的语义特征也是一种非常重要的信息。因此,所提网络在提取边缘特征、形状信息的基础

上,引入空间注意力模块 DANet^[18]来捕捉点间的语义特征,以提升点云场景分割的准确性。图 2 为空间注意力模块示意图。对于给定的特征 $A \in \mathbf{R}^{N \times d}$,分别使用两个独立的卷积操作生成两个新的特征 $B \in \mathbf{R}^{N \times d}$ 和 $C \in \mathbf{R}^{N \times d}$ 。在 DANet 中,对特征 A 应用卷积操作得到特征 B 和 C 后,对 B 和 C 进行 reshape 操作,然后在 B 和 C 之间进行矩阵相乘,再使用 Softmax 得到空间注意力图。但是,reshape 操作会改变特征的形状,导致特征信息的损失,因此本实验组删除 reshape 操作,直接对 B 和 C 进行矩阵相乘操作,得到自注意力权重矩阵 $U \in \mathbf{R}^{N \times N}$,并应用 Softmax 得到归一化的空间注意力权重矩阵。

$$U_{ij} = \frac{\exp(B_i \cdot C_j^T)}{\sum_{i=1}^N \exp(B_i \cdot C_j^T)}, \quad i, j = 1, 2, 3, \dots, N, \quad (5)$$

式中: B_i 和 C_j 分别是点 i 和点 j 在 B 和 C 中的特征表示; U_{ij} 衡量的是第 i 个点与第 j 个点之间的相似性,相似性越高,值越大,在一定程度上反映了点之间的相关性。

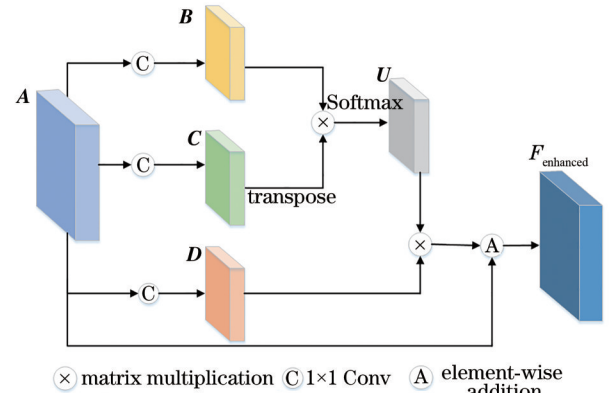


图 2 空间注意力模块

Fig. 2 Spatial attention module

此外,本实验组对特征 A 应用卷积操作得到新的特征向量 $D \in \mathbf{R}^{N \times d}$,并将其与注意力权重 U_{ij} 进行矩阵相乘操作,最后与标度参数 α 相乘,并且与输入特征 A 进行元素求和操作,即可得到增强的输出特征 $F_{enhanced}$ 。

$$F_{enhanced} = \alpha \sum_{j=1}^N (U_{ij} D_j) + A_i, \quad (6)$$

式中: α 是初始值为 0 的可学习的标度参数。在点云的模型分割任务中,距离很远的两个点也有可能属于同一个语义类别,因此这样的点应该被结合起来考虑。空间注意力模块的引入可以建立点之间的全局依赖关系,提取点之间的深层次语义相关性。

3 实验结果与分析

3.1 实验设置

将所提网络 AMFF-DGCNN 分别在目标分类 ModelNet40^[19]、部件分割 ShapeNet^[20]、场景语义分割 S3DIS^[21] 3 个公共数据集上进行了测试,以评估网络模型的性能。实验环境基于 Windows 10 操作系统、Intel i7 8700 CPU、32 GB 内存、RTX 2080 Ti 显卡,深度学习框架为 Tensorflow-GPU 1.13.0。训练过程中采用 Adam 作为优化器,设置动量为 0.9,权重衰减为 0.0001,初始学习率为 0.001,学习率衰减指数为 0.5,衰减速度为 300000,实验中的 k 近邻设置为 20。为了防止过拟合,全连接层中 dropout 的参数保留率在目标分类、部件分割、场景语义分割实验中分别设置为 0.5、0.5、0.8。

3.2 目标分类

为评估模型的分类效果,在 ModelNet40 标准数据集^[19]上进行训练和测试,该数据集包含 40 个类别的 12311 个 CAD 模型,其中 9843 个形状用于训练,2468 个形状用于测试。同 PointNet 一致,从每个模型随机采样 1024 个点,并将其标准化到单位球体中,将采样点的坐标值 x 、 y 、 z 作为输入, batch size 设置为 16。

在图 1 中, AMFF-DGCNN 利用 edgeconv 模块获取局部几何特征,并使用空间注意力机制提取点与点之间的空间语义信息,将局部几何特征与全局语义特征融合,增强点的特征表示,弥补了 DGCNN 对全局特征提取不足的缺陷。

表 1 为所提算法与其他算法在 ModelNet40 数据集的实验结果,本实验组采用了总体分类精度(OA)和平均分类精度(mACC)作为衡量指标来评估分类结果,其中加粗数据表示最优结果。从表 1 中可以看出,无论是总体分类精度还是平均分类精度,所提算法均优于其他算法。在 OA 方面,所提算法比 PointNet 高 2.6 个百分点,比 DGCNN 高 0.6 个百分点;在 mACC 方面,所提算法比 PointNet 高 3.2 个百分点,比 DGCNN 高 0.3 个百分点。

3.3 部件分割

为了验证网络模型在处理三维点云细粒度分割任务上的有效性,在 ShapeNet 数据集^[20]上进行实验,并与已有算法进行了比较,该数据集共包括 16 个对象类别的 16881 个 CAD 模型,总共被标记为 50 个部件,每个对象都由 2~5 个部件组成。按照

表 1 ModelNet40 数据集上的分类实验结果对比

Table 1 Comparison of classification experimental results on ModelNet40 dataset

| Algorithm | mACC / % | OA / % |
|---------------------------|-------------|-------------|
| VoxNet ^[22] | 83.0 | 85.9 |
| PointNet ^[6] | 86.0 | 89.2 |
| PointNet++ ^[7] | | 90.7 |
| ECC ^[23] | 83.2 | 87.4 |
| SO-Net ^[24] | 87.2 | 90.9 |
| RGCNN ^[25] | 87.3 | 90.5 |
| DGCNN ^[13] | 88.9 | 91.2 |
| Proposed algorithm | 89.2 | 91.8 |

DGCNN 的设置进行数据集划分,从每个模型随机采样 2048 个点作为输入。

在图 1 中,部件分割网络在提取的边缘特征里加入几何特征信息来增强特征表示的丰富性;此外使用多层感知机学习每个独立点的高维表征,以提取点云的全局特征,提高网络在分割任务中的性能。在部件分割任务中, batch size 设置为 8。

表 2 为所提算法与其他算法在三维模型 ShapeNet 数据集上的实验对比结果,采用平均交并比(mIOU)作为衡量分割任务性能的指标。表 2 中第 1 列为使用的算法,第 2 列为在 ShapeNet 数据集上得到的 mIOU,第 3 列到最后 1 列,每 1 列对应一个对象类别(例如:飞机、包、茶杯等)的 mIOU。从表中可以看到,所提算法的 mIOU 比 PointNet 高出 1.2 个百分点,比 GAPNet 高出 0.2 个百分点,比 DGCNN 高出 0.2 个百分点,具有更好的部件分割性能。另外,所提 AMFF-DGCNN 在椅子、刀具及手枪 3 个类别取得较好的分割结果。

图 3 为所提算法与 DGCNN 在 ShapeNet 数据集上的部件分割可视化效果图,第 1 列为分割真值图,第 2 列为 DGCNN 的分割效果图,第 3 列为所提算法的分割效果图。从图中可以看出:对于汽车模型(图 3 第 1 行),所提算法能够较好地将汽车的车轮分割出来;对于摩托车模型(图 3 第 2 行),所提算法提取的车轮点与真值更接近;对于桌子模型(图 3 第 3 行),所提算法可以将桌面与桌沿的连接处清楚地分割出来;对于椅子模型(图 3 第 4 行),所提算法对椅背与座位的粘连处等细粒度细节的分割准确率有明显提高。实验结果表明,与 DGCNN 相比,所提算法的分割错误率明显减小,语义分割结果与分割真值高度吻合,进一步验证了所提多特征融合动态图卷积神经网络对于点云细粒度几何特征的

表 2 ShapeNet 数据集上的部件分割实验结果对比

Table 2 Comparison of experimental results of component segmentation on ShapeNet dataset

| Algorithm | mIOU | Aero | Bag | Cap | Car | Chair | Earphone | Guitar | Knife | Lamp | Laptop | Motor | Mug | Pistol | Rocket | Skateboard | Table |
|-------------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| PointNet ^[6] | 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 ^[26] | 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 |
| GAPNet ^[27] | 84.7 | 84.2 | 84.1 | 88.8 | 78.1 | 90.7 | 70.1 | 91.0 | 87.3 | 83.1 | 96.2 | 65.9 | 95.0 | 81.7 | 60.7 | 74.9 | 80.8 |
| SCN ^[28] | 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 |
| RSNet ^[29] | 84.9 | 82.7 | 86.4 | 84.1 | 78.2 | 90.4 | 69.3 | 91.4 | 87.0 | 83.5 | 95.4 | 66.0 | 92.6 | 81.8 | 56.1 | 75.8 | 82.2 |
| DGCNN ^[13] | 84.7 | 83.6 | 85.7 | 84.8 | 78.2 | 90.5 | 75.4 | 91.1 | 87.2 | 82.4 | 95.6 | 62.9 | 94.5 | 80.7 | 63.6 | 75.4 | 81.8 |
| Proposed algorithm | 84.9 | 84.1 | 80.2 | 85.5 | 78.7 | 90.8 | 74.5 | 90.7 | 88.3 | 82.1 | 95.7 | 67.0 | 93.8 | 83.0 | 56.3 | 73.7 | 81.9 |

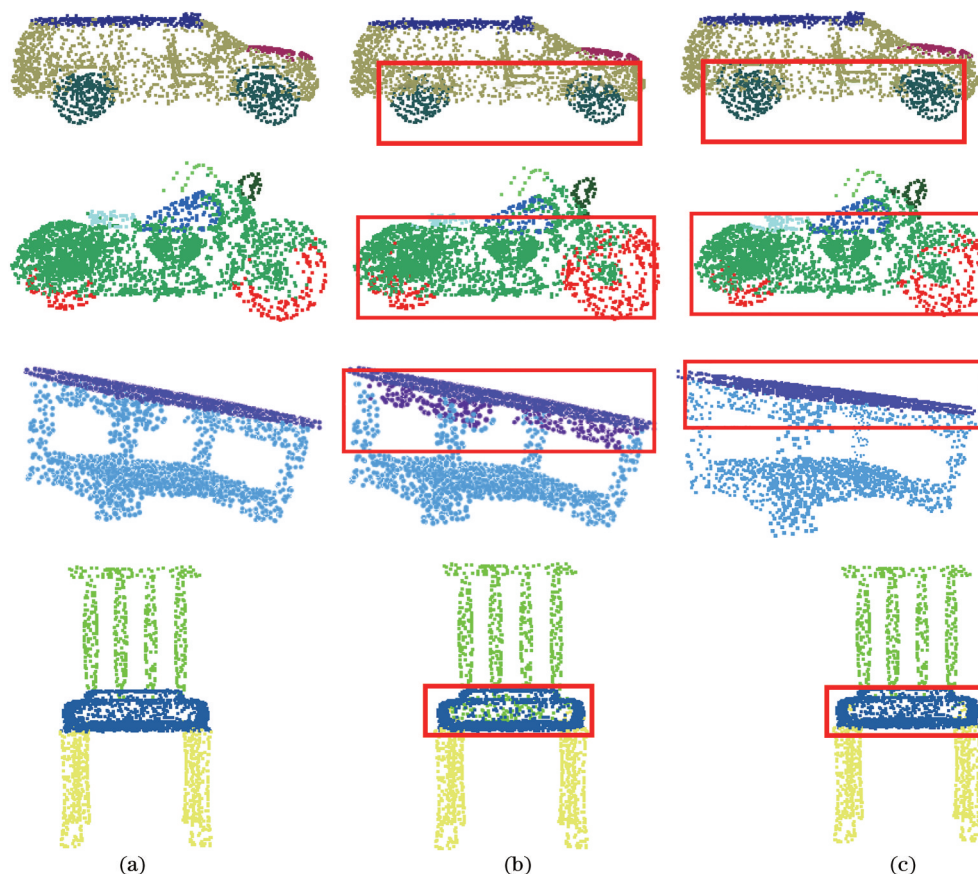


图 3 ShapeNet 数据集上 DGCNN、AMFF-DGCNN 的部件分割效果。(a) 分割真值图; (b) DGCNN; (c) AMFF-DGCNN
 Fig. 3 Component segmentation effects of DGCNN, AMFF-DGCNN on ShapeNet dataset. (a) Ground truth; (b) DGCNN; (c) AMFF-DGCNN

捕获能力。

3.4 语义分割

为了验证所提算法在大规模场景点云的分割性能,在三维室内场景点云语义分割数据集 S3DIS^[21]上进行训练和测试。S3DIS 数据集包含 6 个区域的 272 个小房间,场景中的每个点都标注有一个标签,一共有 13 类,包括天花板、墙面、横梁、椅子等,每个点由它的 3D 坐标、RGB 颜色信息、归一化空间坐标的 9 维向量表示。

场景语义分割任务与部件分割任务相比,每个场景中包含丰富的语义类别,并且含有噪声,因此更能体现网络对于模型分割任务的泛化能力。语义分割所使用的网络结构与图 1 所示的分割网络相似,只使用了 3 层特征提取模块(64, 64, 64),batch size 设为 6。

本实验组分别使用 6 折交叉验证和单独将 Area 5 作为测试区域对网络模型进行评估,采用 mIOU 和 OA 作为衡量分割任务性能的指标,并与已有算法进行比较。表 3 为将 Area 5 作为测试区

域时的实验结果。使用 Area 5 作为测试集时,由于 Area 5 和其他区域中的房间有所不同,包含的物体存在差异,可以评估模型的可推广性。表 3 第 1 列为使用的算法,第 2 列为 OA,所提算法比 DGCNN 高出 1.29 个百分点,第 3 列为 mIOU,所提算法比 PointNet 高出 10.47 个百分点,比 MS3_DVS 高出 5.24 个百分点,比 DGCNN 高出 4 个百分点。表 3 中第 4 列至最后 1 列,每 1 列代表不同类别(例如:天花板、地板、墙面、柱子、窗户等)的分割精度,所提算法在天花板、地板、墙面、柱子、窗户等 8 个类别都取得较好的分割精度。表 4 不同算法使用 6 折交叉验证在 S3DIS 数据集的实验结果,6 折交叉验证即将数据集划分为 6 份,并将每 1 份都分别作为测试集进行 6 次测试和训练,可以充分评估模型的泛化能力。

从表 4 可以看出:在 OA 方面,所提算法比 SPG 高出 2.71 个百分点,比 DGCNN 高出 1.29 个百分点,比 Octant-CNN 高出 1.01 个百分点,比 Contextual attention CNN 高出 0.41 个百分点;在 mIOU 方面,所提算法比 SPG 高出 5.84 个百分点,比 DGCNN 高出 3.04 个百分点,比 Octant-CNN 高出 1.6 个百分点,比 Contextual attention CNN 高出 2.3 个百分点;并在天花板、梁、柱子、窗户等 6 个类别都取得较好的分割精度。

由表 3 可知,所提算法在天花板,窗户等类别中显示出较好的分割性能,主要原因在于所引入的空间注意力机制在有效提取每个点之间语义信息的同时,计算了点的形状信息,增强了平坦区域中点的特征表示。

表 3 S3DIS 数据集中 Area 5 语义分割结果

Table 3 Semantic segmentation results of Area 5 on S3DIS dataset

| Algorithm | OA | mIOU | Ceiling | Floor | Wall | Beam | Column | Window | Door | Chair | Table | Bookcase | Sofa | Board | Clutter |
|--------------------------|--------------|--------------|--------------|--------------|--------------|------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| PointNet ^[6] | | 41.09 | 88.80 | 97.33 | 69.80 | 0.05 | 3.92 | 46.26 | 10.76 | 52.61 | 58.93 | 40.28 | 5.85 | 26.38 | 33.22 |
| MS3_DVS ^[30] | | 46.32 | 79.03 | 88.07 | 53.55 | 0.00 | 20.47 | 29.01 | 37.29 | 68.84 | 63.72 | 47.44 | 61.62 | 16.50 | 36.64 |
| SegCloud ^[31] | | 48.92 | 90.06 | 96.05 | 69.86 | 0.00 | 18.37 | 38.35 | 23.12 | 75.89 | 70.40 | 58.42 | 40.88 | 12.96 | 41.60 |
| DGCNN ^[13] | 83.31 | 47.56 | 92.77 | 97.54 | 74.86 | 0.00 | 11.77 | 50.72 | 23.71 | 66.38 | 69.58 | 8.58 | 48.72 | 31.58 | 42.06 |
| Proposed algorithm | 84.60 | 51.56 | 92.81 | 97.78 | 78.12 | 0.00 | 24.91 | 51.92 | 31.05 | 69.11 | 73.91 | 16.21 | 52.68 | 39.51 | 42.18 |

表 4 6 折交叉验证在 S3DIS 数据集上的语义分割结果

Table 4 Semantic segmentation results of 6-fold cross-validation on S3DIS dataset

| Algorithm | OA | mIOU | Ceiling | Floor | Wall | Beam | Column | Window | Door | Chair | Table | Bookcase | Sofa | Board | Clutter |
|--|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|-------------|--------------|
| PointNet ^[6] | 78.60 | 47.60 | 88.00 | 88.70 | 69.30 | 42.40 | 23.10 | 47.50 | 51.60 | 42.00 | 54.10 | 38.20 | 9.60 | 29.40 | 35.20 |
| RSNet ^[29] | | 56.47 | 92.48 | 92.83 | 78.56 | 32.75 | 34.37 | 51.62 | 68.11 | 59.72 | 60.13 | 16.42 | 50.22 | 44.85 | 52.03 |
| SPG ^[32] | 82.90 | 54.06 | 92.17 | 95.00 | 71.91 | 33.46 | 15.03 | 46.53 | 60.92 | 65.05 | 69.45 | 56.82 | 38.21 | 6.86 | 51.29 |
| DGCNN ^[13] | 84.32 | 56.86 | 92.78 | 93.74 | 76.42 | 53.10 | 35.59 | 56.46 | 61.24 | 64.16 | 51.43 | 15.93 | 48.26 | 43.06 | 46.99 |
| Octant-CNN ^[33] | 84.6 | 58.3 | 92.1 | 94.5 | 76.3 | 48.9 | 30.8 | 56.9 | 62.9 | 65.8 | 55.5 | 28 | 48.1 | 50.3 | 48.4 |
| Contextual attention CNN ^[34] | 85.2 | 57.6 | | | | | | | | | | | | | |
| Proposed algorithm | 85.61 | 59.90 | 92.86 | 94.73 | 78.27 | 54.04 | 43.25 | 59.06 | 62.60 | 67.39 | 59.08 | 18.28 | 51.29 | 47.40 | 50.51 |

图 4 为 AMFF-DGCNN 与 DGCNN 在 S3DIS 数据集中的分割结果,第 1 列为输入的原始点云数据,第 2 列为分割真值图,第 3 列为 DGCNN 的分割效果图,第 4 列为所提 AMFF-DGCNN 的分割效果图。从图中可以看出:在第 1 个场景中(图 4 第 1 行),所提算法对于左墙面上白板的预测比 DGCNN 更为完整,而且对于天花板与墙面连接处的杂乱点的识

别与真实值高度吻合;在第 2 个场景中(图 4 第 2 行),DGCNN 将左边一部分墙面误分割为白板,而对嵌入在右墙面上的白板的预测结果残缺不全,所提算法对于物体连接处的局部细节的预测结果较为准确,较为完整地保留了白板;在第 3 个场景中(图 4 第 3 行),DGCNN 将横梁误分为天花板,将书架误分为墙面和杂乱点,而所提算法完整地分割出整个

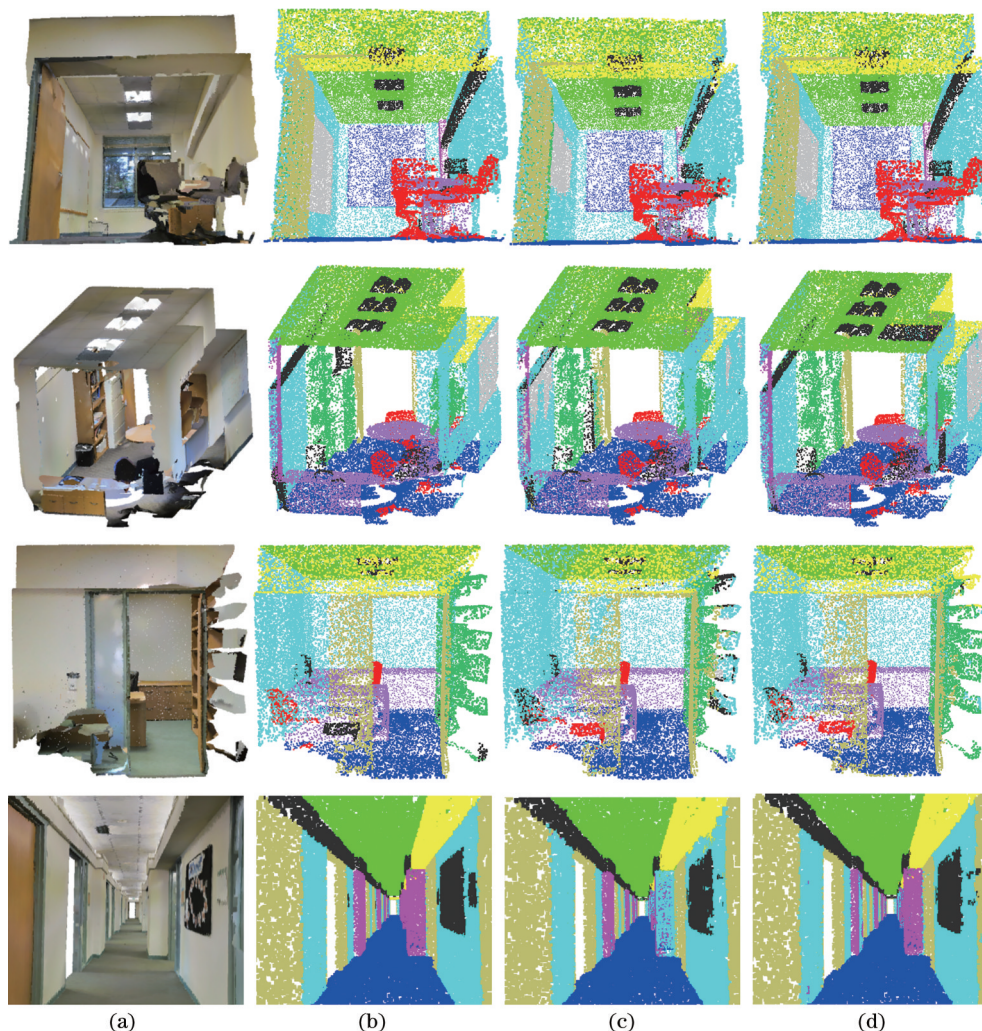


图4 DGCNN、AMFF-DGCNN在S3DIS数据集的语义分割结果。(a)输入点云数据;(b)分割真值图;(c) DGCNN分割结果;(d) AMFF-DGCNN分割结果

Fig. 4 Semantic segmentation results of DGCNN, AMFF-DGCNN on S3DIS dataset. (a) Input point cloud; (b) ground truth; (c) DGCNN segmentation results; (d) AMFF-DGCNN segmentation results

横梁;在第4个场景中(图4第4行),DGCNN将与墙面相连的柱子误分为墙面,而所提算法可以较为清晰地分割出柱子的边缘。对比DGCNN,所提算法具有更好的分割结果,主要原因在于所提AMFF-DGCNN可以较好地捕捉点云局部几何特征,并且结合了点间语义特征的优势,从而提高点云语义分割精度。

3.5 消融实验

为了进一步探究网络结构中不同模块的重要性,在S3DIS数据集的Area 5上进行了消融实验。对点特征、形状特征及空间注意力机制依次进行实验,来研究每个模块对测试结果的影响,结果如表5所示。

从表5可以看出:单纯的利用边缘特征对点云场景进行语义分割,模型A的分割准确率为

表5 模块重要性分析

Table 5 Module importance analysis

| Model | Point feature | Geometry feature | Spatial attention | OA / % |
|-------|---------------|------------------|-------------------|--------|
| A | | | | 83.3 |
| B | ✓ | | | 84.15 |
| C | ✓ | ✓ | | 84.3 |
| D | ✓ | ✓ | ✓ | 84.6 |

83.3%;为了弥补DGCNN在捕捉全局特征方面的缺陷,将边缘特征除以2,然后在每层加入点特征,使得网络模型的分割准确率达到84.15%;然后,每层又融合了形状几何特征,网络模型的分割准确率提升到84.3%;最后为了进一步增强点的上下文语义特征的表达,引入了空间注意力机制,最终

的分割准确率达到 84.6%。

4 结 论

为了弥补动态图卷积神经网络对于点云形状特征、语义特征提取不充分的缺陷,提出了一种基于空间注意力的多特征融合动态图卷积神经网络 AMFF-DGCNN。在每一层网络中利用边卷积提取点的局部几何特征,并且加入形状特征增强网络模型对于几何实例的有效表征。同时提取全局单点特征来增强网络由局部到全局的特征提取能力;此外,引入注意力机制提取点之间的语义信息,充分挖掘点之间的相关性。实验结果表明,所提网络在三维目标分类、部件分割和语义分割任务上都展现了良好的性能。但是,所提算法在提取边缘特征时,无法对不同边缘特征的重要性进行有效鉴别。因此,如何对边缘特征设置有效的权重值,充分挖掘局部更细粒度的几何特征,是下一步工作的重点。

参考文献

- [1] Fang W, Ding Y W, Zhang F H, et al. DOG: a new background removal for object recognition from images[J]. *Neurocomputing*, 2019, 361: 85-91.
- [2] Lateef F, Ruichek Y. Survey on semantic segmentation using deep learning techniques[J]. *Neurocomputing*, 2019, 338: 321-348.
- [3] Zhang J Y, Zhao X L, Chen Z. Review of semantic segmentation of point cloud based on deep learning [J]. *Laser & Optoelectronics Progress*, 2020, 57(4): 040002.
张佳颖, 赵晓丽, 陈正. 基于深度学习的点云语义分割综述[J]. *激光与光电子学进展*, 2020, 57(4): 040002.
- [4] Wen P, Cheng Y L, Yu W C. Point cloud classification methods based on deep learning: a review[J]. *Laser & Optoelectronics Progress*, 2021, 58(16): 1600003.
文沛, 程英蕾, 余旺盛. 基于深度学习的点云分类方法综述[J]. *激光与光电子学进展*, 2021, 58(16): 1600003.
- [5] Guo Y L, Wang H Y, Hu Q Y, et al. Deep learning for 3D point clouds: a survey[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(12): 4338-4364.
- [6] 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 (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 77-85.
- [7] Qi C R, Yi L, Su H, et al. Pointnet++: deep hierarchical feature learning on point sets in a metric space[C]//Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA. [S.l.: s.n.], 2017, 30: 5099-5108.
- [8] Jiang M Y, Wu Y R, Zhao T Q. PointSIFT: a SIFT-like network module for 3D point cloud semantic segmentation[EB/OL]. (2018-07-02) [2021-05-02]. <https://arxiv.org/abs/1807.00652>.
- [9] Shang P F, Chen Y, Lü W J, et al. A point cloud semantic segmentation network considering normals [J/OL]. *Laser & Optoelectronics Progress*: 1-16 [2021-01-25]. <http://kns.cnki.net/kcms/detail/31.1690.tn.20210802.1730.058.html>.
尚鹏飞, 陈义, 吕伟嘉, 等. 一种考虑法矢量的点云语义分割网络[J/OL]. *激光与光电子学进展*: 1-16 [2021-01-25]. <http://kns.cnki.net/kcms/detail/31.1690.tn.20210802.1730.058.html>.
- [10] Li Y Y, Bu R, Sun M C, et al. PointCNN: convolution on X-transformed points[C]//Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montreal, Canada. [S.l.: s.n.], 2018: 828-838.
- [11] 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.
- [12] Wu W X, Qi Z, Fuxin L. PointConv: deep convolutional networks on 3D 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: 9613-9622.
- [13] 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.
- [14] Lan S Y, Yu R C, Yu G, et al. Modeling local geometric structure of 3D point clouds using geo-CNN[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: 998-1008.
- [15] Zhang K G, Hao M, Wang J, et al. Linked dynamic graph CNN: learning on point cloud via linking hierarchical features[EB/OL]. (2019-04-22)[2021-05-02]. <https://arxiv.org/abs/1904.10014>.
- [16] Weinmann M, Jutzi B, Hinz S, et al. Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2015, 105: 286-304.
- [17] 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.
- [18] 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.
- [19] 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 (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 1912-1920.
- [20] Yi L, Kim V G, Ceylan D, et al. A scalable active framework for region annotation in 3D shape collections[J]. *ACM Transactions on Graphics*, 2016, 35(6): 1-12.
- [21] Armeni I, Sener O, Zamir A R, et al. 3D semantic parsing of large-scale indoor spaces[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 1534-1543.
- [22] Maturana D, Scherer S. VoxNet: a 3D convolutional neural network for real-time object recognition[C]//2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), September 28-October 2, 2015, Hamburg, Germany. New York: IEEE Press, 2015: 922-928.
- [23] Simonovsky M, Komodakis N. Dynamic edge-conditioned filters in convolutional neural networks on graphs[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 29-38.
- [24] 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.
- [25] Te G S, Hu W, Zheng A M, et al. RGCNN: regularized graph CNN for point cloud segmentation [C]//Proceedings of the 26th ACM international conference on Multimedia, October 22-26, 2018, Seoul, Republic of Korea. New York: ACM, 2018: 746-754.
- [26] Klokov R, Lempitsky V. Escape from cells: deep Kd-networks for the recognition of 3D point cloud models [C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 863-872.
- [27] Chen C, Fragonara L Z, Tsourdos A. GAPNet: graph attention based point neural network for exploiting local feature of point cloud[EB/OL]. (2019-05-21) [2021-05-02]. <https://arxiv.org/abs/1905.08705>.
- [28] Xie S N, Liu S N, Chen Z Y, et al. Attentional ShapeContextNet for point cloud recognition[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: 4606-4615.
- [29] Huang Q G, Wang W Y, Neumann U. Recurrent slice networks for 3D segmentation of point clouds [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: 2626-2635.
- [30] Roynard X, Deschaud J E, Goulette F. Classification of point cloud scenes with multiscale voxel deep network [EB/OL]. (2018-04-10) [2021-05-02]. <https://arxiv.org/abs/1804.03583>.
- [31] Tchampi L, Choy C, Armeni I, et al. SEGCloud: semantic segmentation of 3D point clouds[C]//2017 International Conference on 3D Vision (3DV), October 10-12, 2017, Qingdao, China. New York: IEEE Press, 2017: 537-547.
- [32] Landrieu L, Simonovsky M. Large-scale point cloud semantic segmentation with superpoint graphs[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: 4558-4567.
- [33] Xu X, Shuai H, Liu Q S. Octant convolutional neural

network for 3D point cloud analysis[J/OL]. Acta Automatica Sinica:1-10[2021-09-02]. <http://kns.cnki.net/kcms/detail/11.2109.TP.20200811.1020.002.html>.

许翔, 帅惠, 刘青山. 基于卦限卷积神经网络的3D点云分析[J/OL]. 自动化学报:1-10[2021-09-02]. <http://kns.cnki.net/kcms/detail/11.2109.TP.20200811.1020.002.html>.

html.

[34] Yang J, Dang J S. Semantic segmentation of 3D point cloud based on contextual attention CNN[J]. Journal on Communications, 2020, 41(7): 195-203.

杨军, 党吉圣. 基于上下文注意力CNN的三维点云语义分割[J]. 通信学报, 2020, 41(7): 195-203.