

基于多信息深度学习的 3D 点云语义分割

刘友群, 敖建锋

(江西理工大学土木与测绘工程学院, 江西 赣州 341000)

摘 要: 点云语义分割是三维点云数据处理的基础步骤, 是三维场景理解分析、重建和目标识别的关键环节。针对目前对三维点云进行语义分割使用的点云信息少和精度不高的问题, 本文在利用点云三维坐标信息的基础上, 增加了点云 RGB 信息和所属房间的归一化坐标信息, 从而丰富了神经网络输入端的信息量, 进一步提高了模型的分割精度, 最后利用 PointNet++ 对改进后的三维点云语义分割效果进行检验, 实验表明: 在丰富了网络输入端的数据信息后, 模型的总体准确度提高了 6.65 %。

关键词: 三维点云; 深度学习; 语义分割; PointNet++

中图分类号: TP391.41 **文献标识码:** A **DOI:** 10.3969/j.issn.1001-5078.2021.05.022

3D point cloud semantic segmentation based on multi-information deep learning

LIU You-qun, AO Jian-feng

(School of Civil Engineering and Surveying Engineering, Jiangxi University of Science and Technology, Ganzhou 341000, China)

Abstract: Point cloud semantic segmentation is the basic step of 3D point cloud data processing and is the key link of 3D scene understanding analysis, reconstruction and target recognition. In view of the current problem of lack point cloud information and low accuracy used for semantic segmentation of 3D point clouds, this paper adds point cloud RGB information and normalized coordinate information of the room to which it belongs based on the use of point cloud 3D coordinate information to enrich the amount of information at the input end of the neural network and further improve the segmentation accuracy of the model. Finally, PointNet++ is used to test the improved 3D point cloud semantic segmentation effect. The experiment shows that after enriching the data information and inputting the network, the model accuracy is improved by 6.65 %.

Keywords: 3D point clouds; deep learning; semantic segmentation; PointNet++

1 引 言

深度学习是一种建立、模拟人脑进行分析学习的神经网络, 可以通过学习一种深层非线性网络结构, 实现复杂函数逼近, 从少量样本数据集中学习整

个数据集本质特征。点云语义分割是三维点云场景理解和分析的关键步骤, 目前, 基于深度学习的 2D 图像分割技术已经趋于成熟, 但是 2D 数据往往受遮挡等因素影响, 存在一定程度的数据缺失, 使得分

基金项目: 国家自然科学基金地区基金资助项目(No. 41561091); 江西省教育厅科学技术研究项目(No. GJJ150663); 江西省教育厅科学技术研究项目(No. GJJ150629); 江西省教育厅科学技术研究项目(No. GJJ180501) 资助。

作者简介: 刘友群(1998-) , 女, 本科, 主要从事点云数据处理研究。

通讯作者: 敖建锋(1984-) , 男, 讲师, 博士, 主要从事点云数据处理和开采沉降监测研究。

收稿日期: 2020-07-15

割的结果不尽如人意,针对这方面的局限性,人们逐渐将注意力转移到了 3D 点云数据中。虽然点云数据的嘈杂、稀疏和无序加大了语义分割的难度,但是可以充分利用点云的三维属性和几何特征分布等特性,避免转换成其他规则结构时造成的信息丢失。

因此本文直接利用基于原始点云的 PointNet++ 模型^[1]进行点云语义分割,避免了数据转换时产生的信息丢失。并在原始模型的基础上进行改进,丰富了网络输入端的数据信息,提高了模型的分割精度,得到了较好的分割效果。

2 相关工作

为了使点云适用于深度神经网络,有些研究者将三维点云数据转换成二维图像、体素等规则的结构,这样不但可以避免复杂的点云处理,还可以利用较成熟的基于规则结构的网络来实现点云的分割。深度学习对二维图像和体素的分割已经取得了很好的效果,并且已经有很多成熟的算法,加之点云数据也有其独特的优势,所以近几年来研究者逐渐把目光转向了基于原始点云的网络模型研究。根据以上几种思路,基于深度学习的点云分割方法的研究现状根据基本思想的不同分为以下三类。

2.1 基于投影和视图的点云分割方法

该方法的中心思想是先将点云光栅化为 2D 图像,或者通过球面、柱面投影的方式将点云的三维坐标投影到二维平面上。

受二维图像处理的启发,2015 年 Su 等人提出了基于图像的多视图卷积神经网络(MVCNN)^[2],此后点云投影转换为多视图的二维图像的方法开始广泛应用。2017 年,Lawin 等人^[3]提出了 DeePr3ss 算法,该方法的思想是先将点云投影到二维图像上,再将投影到的图像输入到 CNN 中进行分割,最后把图像分割结果重新投影回点云上,以此来实现点云的分割。2018 年, Qin 等人提出了一种将 ALS 点云转换为多视图进行地形分类的方法^[4],其中,提出了一个 TLFnet 网络,该网络首次利用多视图 CNN 进行大规模的 ALS 点云分割。2019 年, Chen 等人^[5]提出了一种通过过滤原始语义标签的投票策略来处理投影中标签错误的问题,这在降低标签错分概率上取得了很好的效果。2020 年, Alonso 等人^[6]提出了 3D-MiniNet 网络,提出了快速近邻搜索的方法和滑动框搜索球形投影后的像素点。

2.2 基于体素(Voxel)的点云分割方法

该方法的中心思想是先将点云体素化,变成体素这类规则的结构,然后使用深度神经网络对体素化的点云进行特征学习以实现点云的分割。

与深度图像网络一样,最早的深度体素网络也是 2015 年诞生的, VoxNet 网络^[7]由 Maturana 等人提出。同年, Wu 等人用卷积 DBN 网络构建了一个 3D ShapeNets,并首次建立了基于体素的数据集 ShapeNet 和 ModelNet^[8]。2016 年,国内学者 Huang 等人^[9]提出了一种基于 3D-CNN 的点云标记算法。随后,在 VoxNet 网络和 ShapeNet 的影响下,2017 年, Timo Hackel 等人提出了 FCNVoxNet 算法^[10],第一次将基于体素的方法用于室外场景分割,并建立了数据集 Semantic3D。同年, Tchapmi 等人提出了一种网络架构 SEGCloud^[11]。2018 年, Wang 等人提出了多尺度卷积网络(MSNet)^[12]。

2.3 基于原始点云数据的点云分割方法

近几年,基于原始点云数据的点云分割方法成为了国内外学者的研究热点,也成为了点云分割的主流发展方向。

2017 年, Qi 等人提出的直接对原始点云进行特征学习的网络架构 PointNet^[13],随后 Qi 等人在同一年又提出了 PointNet++^[1], PointNet++ 对 PointNet 做出了改进。2018 年, Jiang 等人参考了尺度不变特征变换(SIFT)特征子,设计了一个 PointSIFT 模块^[14]。Li 等人提出了 PointCNN 网络^[15]。Hua 等人提出了 Pointwise 卷积^[16]并用来构建神经网络。Wang 等人提出了一种将 GCNN 和 PointNet 结合用以点云分割的方法 DGCNN(Dynamic Graph CNN)^[17]。

为了更好地挖掘点云局部几何信息,2019 年, Komarichev 等人提出了环形卷积神经网络(A-CNN)^[18]。Zhang 等人对 DGCNN 做出了改进提出了 LDGCNN^[19],在 DGCNN 的基础上增加了节点的跨层连接,并把转换网络用多层感知机(MLP)代替,显著减小了网络的复杂度。

3 构建 PointNet++ 模型

PointNet++ 的主要结构如图 1 所示。其基本思想是:采用一种分层抽取特征的思想,可以递归地将 PointNet 循环使用以提取点云局部特征,并采用了多尺度分组(MSG)和多分辨率分组(MRG)来解决

点云密度不同的问题。

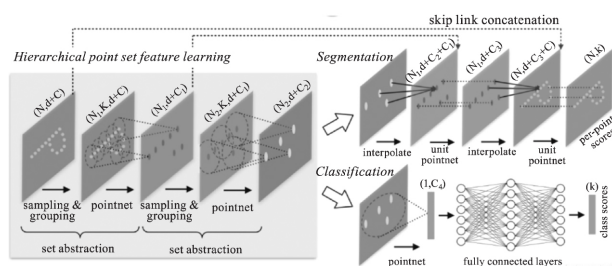


图1 PointNet++网络结构

Fig. 1 PointNet++ Network structure

PointNet++借鉴了卷积神经网络(CNN)这种多层感受野的思想,设计了一个 set abstraction 来对点云进行分层提取特征,每一个特征提取层都分为采样层、分组层和特征提取层三个部分组成,流程如图2所示。

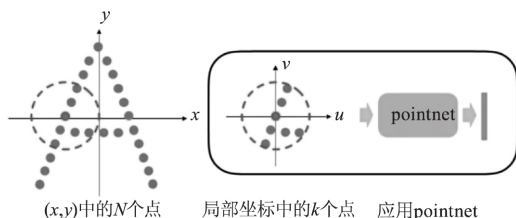


图2 Set abstraction 流程图

Fig. 2 Set Abstraction Flow

采样层: 采样算法使用迭代最远点采样方法(FPS)来选择一些点作为局部区域的中心点,其中中心思想是假设一共有 n 个点,先从所有点云中随机选择一个点作为起始点,然后选择离这个点最远的点也加入起始点集,而后再从剩余的 $(n-2)$ 个点中选择出离起始点集最远的点并加入到起始点集中,继续迭代,直到选出需要的个数为止。与随机采样法相比较,最远点采样法的优势是可以最大限度地覆盖所有点。

分组层: 又称聚类层,目的就是要将点云分组聚类划分成多个局部区域。由PointNet++网络结构图可看出,输入点云的大小为 $N \times (d+c)$, d 为点的坐标维度, c 为点的特征维度,在经过采样层采样后,选择了 $N_1 \times (d+c)$ 的中心点,再经过分组层后得到的每一个局部区域的大小为 $N_1 \times K \times (d+c)$, K 代表局部区域近邻点的数量,不同组的 K 的大小不一样。在PointNet++中采用了球查询(Ball query)分组算法,其中中心思想是把给定的中心点的固定半径内的近邻点都选取出来,同时给定近邻点的个数。与另一种聚类算法K近邻(KNN)算法相比较,球查询算法保

证了固定的查询区域尺度,从而使得局部区域特征在空间上更加一般化,这样更适合对区域进行划分。

特征提取层: 该层是对每一个局部区域通过小型PointNet网络进行卷积和池化得到的特征作为这个局部区域中心点的特征,输入数据为点 N_1 的局部区域,数据大小为 $N_1 \times K \times (d+c)$,输出数据为由其质心和编码质心邻域的局部特征抽象而成的局部区域,大小为 $N_1 \times (d+c_1)$, c_1 是这个局部区域的总体特征。为了方便PointNet网络进行特征提取,在这一层需先整个局部区域的点的坐标进行归一化处理,也就是将所有点的坐标转换成相对于质心的统一坐标系下。公式如下:

$$X_i^{(j)} = X_i^{(j)} - \hat{X}_i^{(j)} \quad (1)$$

式中 $i=1, 2, \dots, k; j=1, 2, \dots, d; \hat{X}_i$ 是质心坐标。使用这种相对坐标有助于捕获局部区域中点之间的关系。

由于采集设备和采集技术的局限,采集到的点集往往在不同区域会密度不均匀。为了使网络自适应密度不均匀下的特征提取,PointNet++提出了两种自适应密度的分组聚类方法,分别是多尺度分组方法(MSG)和多分辨率分组方法(MRG),示意图如图3所示。

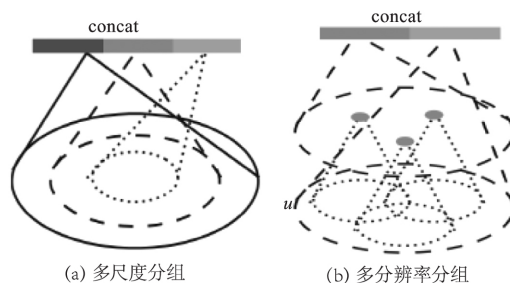


图3 密度自适应层

Fig. 3 Density adaptive layers

多尺度分组(MSG): 如图3(a)所示,其基本思想是产生不同尺度即不同半径的分组,用特征提取层来提取每个尺度的特征,最后将每个尺度提取出来的特征连结起来形成多尺度特征。

多分辨率分组(MRG): 如图3(b)所示,其中心思想是使每一层的特征都由两个向量组成,左边的向量使用几何抽象层来汇总子区域的特征,右边的向量则直接使用PointNet对原始点云进行特征提取,根据点云密度对两边赋予不同的权重。

最后,对于分类网络,只需要通过逐层提取出来

的局部特征,通过全连接层来总结出全局特征就可以得到分类的结果了。而分割网络则复杂很多,需要将特征从子采样点传播到原始点来获取原始点的特征,PointNet++使用了 interpolate 结构,采用了基于距离插值和跨层跳跃链接分层的方法。插值方式采用了 K 近邻(KNN)来反向加权求平均,定义如下:

$$f^{(j)}(x) = \frac{\sum_{i=1}^k w_i(x) f_i^{(j)}}{\sum_{i=1}^k w_i(x)} \quad (2)$$

其中 $w_i(x) = \frac{1}{d(x, x_i)^p}$; $j=1, \dots, c$; 默认情况下使用 $p=2, k=3$ 。

而后再将插值特征值与点集抽象层的特征值通过跨层跳跃链接的方法来连接在一起,然后将连接好的特征值传递到 PointNet 单元中,最后用一些共享的全连接层和 ReLU 层更新每个点的特征向量。重复此过程,直到将特征传播回原始点集为止。

4 实验结果分析与比较

4.1 实验环境

实验平台为 Nvidia Geforce RTX 2070Ti、8GB 显存,在 Ubuntu16.04 和 Python3.6.7 下搭建 CUDA 10.0、CUDNN 7.6.5、TensorFlow 1.13 深度学习环境。实验采用 S3DIS 数据集,S3DIS 数据集是斯坦福大型室内场景数据集,包含 3 个不同建筑的 6 个大型室内区域的 270 个房间,共标记了 13 类语义元素。将 KITTI 数据集划分为 3712 份训练集和 3769 份验证集。本文采用 S3DIS 数据集的 Area1-5 作为训练集,Area6 作为测试集。

4.2 训练

本次 PointNet++ 模型的训练一共用了 23 h。网络训练时超参数设置:初始学习率为 0.001、初始衰减率为 0.5、衰减速度为 300000、Batch_size 为 24、优化方法为 Adam、Momentum 为 0.9、迭代次数为 169800、num_point 为 4096、max_epoch 为 201。一次性采样 4096 个点,每次迭代训练的点数为 24×4096 个,每次迭代 300000 次学习率减半,并对所有训练样本迭代 201 次。由图 4 和图 5 可以看出:网络在训练时随着迭代次数的不断增加,loss 函数值呈现下降趋势,最后稳定在了 0.04~0.05 范围内,精度

也几乎趋近 1。

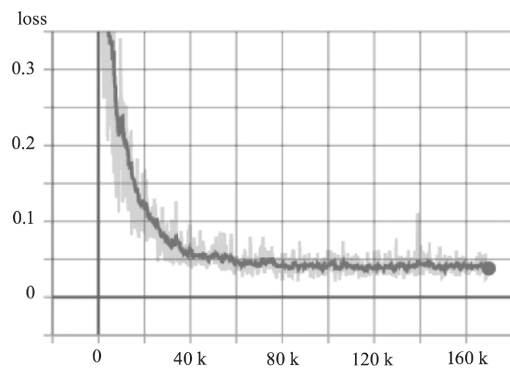


图 4 损失函数曲线

Fig. 4 Loss function curve

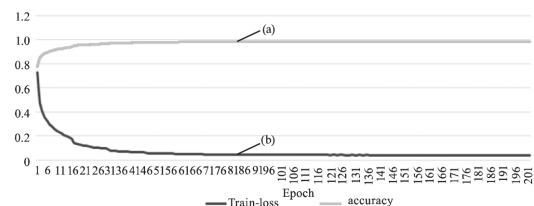


图 5 训练期间准确性(a)和损失(b)的演变曲线

Fig. 5 Training Accuracy(a) and Loss(b) curve

4.3 语义分割结果分析与比较

训练后模型的测试样本如图 6 所示,由此可见本文训练的模型可以对点云较好地分割,输出较平滑的预测边缘。训练后模型的总体准确度达到了 88.88%,每个类别的 IoU、准确率、召回率和 F1-Score 的值如表 1 所示。由表 1 的定量分析结果可看出,本文训练的模型对天花板、地板这类边界明显的结构元素可以达到很好的分割效果。

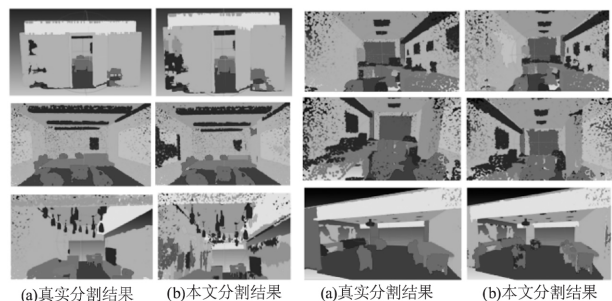


图 6 使用 PointNet++ 分割的部分结果展示

Fig. 6 Partial results of pointnet++ segmentation

为了验证本文训练的模型的高效性,本文将训练的模型与其他几种方法做了比较。比较结果见表 2 和图 7。MLP 是依赖人工提取点云特征的方法,SegCloud 是基于体素的方法,SPG 是基于原始点云的方法。从表 2 可以看出,改进后的 PointNet++ 模型

只有桌子、椅子和书架的 IoU 比 SPG 低,其余 5 种类别 IoU 均高于其他模型。由图 7 可看出,改进后的模型的 mIoU 和 oAcc 都优于其他几种方法,与原始的 PointNet++ 模型相比,总体准确度提高了 6.65%,可见输入更丰富的点云信息有助于训练出更高精度的语义分割模型。

表 1 各类别分割结果

Tab. 1 Segmentation results by category

评价指标/类别	天花板	地板	墙壁	横梁	柱子	窗户	门
IoU	94.19	97.61	78.23	68.53	47.95	76.30	80.14
准确率	0.959	0.987	0.873	0.865	0.803	0.810	0.873
召回率	0.981	0.989	0.883	0.767	0.544	0.929	0.907
F1-score	0.970	0.988	0.878	0.813	0.648	0.866	0.890

评价指标/类别	桌子	椅子	沙发	书架	木板	其他	平均值
IoU	74.37	73.79	39.81	58.37	72.58	63.94	71.22
准确率	0.855	0.881	0.465	0.678	0.834	0.791	0.816
召回率	0.851	0.820	0.734	0.808	0.849	0.770	0.831
F1-score	0.853	0.849	0.569	0.737	0.841	0.780	0.823

表 2 不同模型的各项 IoU 对比

Tab. 2 Comparison of categories' IoU of different models

IoU	天花板	地板	墙壁	门	桌子	椅子	书架	其他
PointNet ^[13]	88.8	97.33	69.8	10.76	58.93	52.61	40.28	33.22
SegCloud ^[11]	90.06	96.05	69.86	23.12	70.4	75.89	58.42	41.6
SPG ^[20]	91.49	96.87	75.89	52.29	77.4	86.35	65.49	50.67
PointNet++(本文)	94.19	97.61	78.23	80.14	74.37	73.79	58.37	63.94

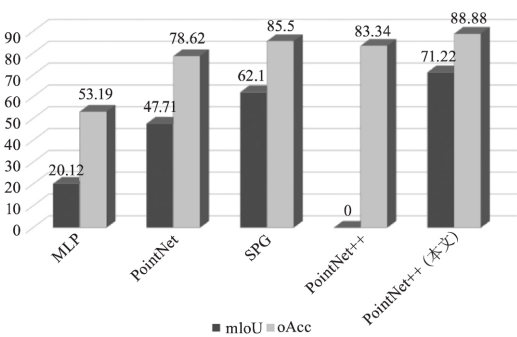
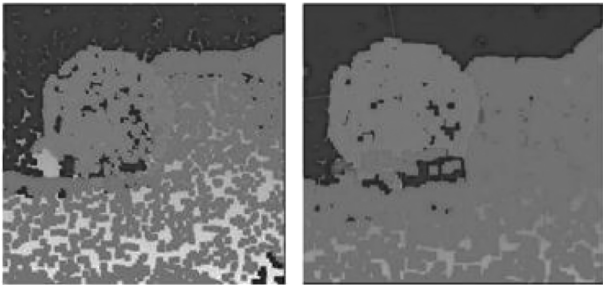


图 7 不同模型分割结果的 mIoU 和 oAcc 对比

Fig. 7 Comparison of mIoU and oAcc of Segmentation results of different models

对于桌子、椅子和书架这类家具元素,本文的分割效果不如 SPG 网络。SPG 网络不是逐点进行分割,而是将多个点组成的点集看作一个整体,对每个点集进行语义分割,这种方法可以较好地描述相邻物体之间的关系,这对基于上下文分类至关重要。图 8 展示了桌子的错分结果,其中,桌面部分点云被识别为沙发,正是由于 PointNet++ 不能很好地处理这种上下文关系导致的。



(a) 真实分割结果 (b) 本文分割结果

图 8 桌子的错分结果示例

Fig. 8 A result of table in wrong segmentation

5 结 语

针对目前对三维点云进行语义分割使用的点云信息少和分割精度不高的问题,本文利用 PointNet++ 实现三维点云的语义分割,并对 PointNet++ 网络输入端的点云信息做了改进,除了使用点云的三维坐标信息之外,还增加了点云 RGB 信息和所属房间的归一化坐标信息。与原始的 PointNet++ 模型相比,总体准确度提高了 6.65%,进一步提高了模型的分割精度。然而由于

PointNet++不能很好地描述相邻物体间的上下文关系,导致对房间内部的家具元素分割的效果不太理想。所以还需要进一步研究何更好的获取相邻物体间的上下文关系,以提高模型语义分割的精度。

参考文献:

- [1] 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 ,2017: 5099–5108.
- [2] Su H , Maji S , Kalogerakis E , et al. Multi-view convolutional neural networks for 3d shape recognition [C]//Proceedings of the IEEE International Conference on Computer vision ,2015: 945–953.
- [3] Lawin F J , Danelljan M , Tosteberg P , et al. Deep projective 3D semantic segmentation [C]//International Conference on Computer Analysis of Images and Patterns. Springer ,Cham ,2017: 95–107.
- [4] Qin N , Hu X , Dai H. Deep fusion of multi-view and multi-modal representation of ALS point cloud for 3D terrain scene recognition [J]. ISPRS Journal of Photogrammetry and Remote Sensing ,2018 ,143: 205–212.
- [5] Chen Y , Yang M , Wang C , et al. 3D semantic modelling with label correction for extensive outdoor scene [C]//2019 IEEE Intelligent Vehicles Symposium(IV) , IEEE , 2019: 1262–1267.
- [6] Alonso I , Riazuelo L , Montesano L , et al. 3D-MiniNet: learning a 2D representation from point clouds for fast and efficient 3D LIDAR semantic segmentation [J]. IEEE Robotic and Automation Lerrers ,2002 ,5(4) : 5432–5439.
- [7] 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) , IEEE ,2015: 922–928.
- [8] Wu Z , Song S , Khosla A , et al. 3d shapenets: a deep representation for volumetric shapes [C]//Proceedings of the IEEE conference on Computer Vision and Pattern Recognition ,2015: 1912–1920.
- [9] Huang J , You S. Point cloud labeling using 3d convolutional neural network [C]//2016 23rd International Conference on Pattern Recognition(ICPR) , IEEE ,2016: 2670–2675.
- [10] Hackel T , Savinov N , Ladicky L , et al. Semantic3d.net: a new large-scale point cloud classification benchmark [J]. arXiv preprint arXiv: 2017 ,1704: 03847.
- [11] Tchapmi L , Choy C , Armeni I , et al. Segcloud: semantic segmentation of 3d point clouds [C]//2017 International Conference on 3D Vision(3DV) , IEEE ,2017: 537–547.
- [12] Wang L , Huang Y , Shan J , et al. MSNet: multi-scale convolutional network for point cloud classification [J]. Remote Sensing ,2018 ,10(4) : 612.
- [13] Qi C R , Su H , Mo K , et al. Pointnet: Deep learning on point sets for 3d classification and segmentation [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition ,2017: 652–660.
- [14] Jiang M , Wu Y , Zhao T , et al. Pointsift: A sift-like network module for 3d point cloud semantic segmentation [J]. arXiv preprint arXiv: 2018 ,807: 00652.
- [15] Li Y , Bu R , Sun M , et al. Pointcnn: convolution on x-transformed points [C]//Advances in Neural Information Processing Systems ,2018: 820–830.
- [16] Hua B S , Tran M K , Yeung S K. Pointwise convolutional neural networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition , 2018: 984–993.
- [17] Wang Y , Sun Y , Liu Z , et al. Dynamic graph cnn for learning on point clouds [J]. ACM Transactions on Graphics (TOG) ,2019 ,38(5) : 1–12.
- [18] Komarichev A , Zhong Z , Hua J. A-CNN: annularly convolutional neural networks on point clouds [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition ,2019: 7421–7430.
- [19] Zhang K , Hao M , Wang J , et al. Linked dynamic graph CNN: learning on point cloud via linking hierarchical features [J]. arXiv preprint arXiv: 2019 ,1904: 10014.
- [20] Landrieu L , Simonovsky M. Large-scale point cloud semantic segmentation with superpoint graphs [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition ,2018: 4558–4567.