

融合多维空间信息的三维点云语义分割方法

魏 东,张潇瀚,刘欢,孙天翼,李昌恺

(沈阳工业大学,辽宁 沈阳 110870)

摘要: 融合多维空间信息的三维点云语义分割方法以较低的时间复杂度处理场景级三维点云。针对局部特征不足导致感知能力下降的问题,通过引入法向量估计和混合池化用以增强局部特征。针对网络结构难以处理大规模三维点云的细微语义对象,通过使用长跳跃多层融合和短跳跃多维特征融合,可以保留多层特征多样性。经实验结果表明,在 S3DIS (Stanford Large-Scale 3D Indoor Spaces Dataset)数据集 6 折交叉验证平均交并比为 70.2%。

关键词: 语义分割;多层融合;法向量估计;多维特征融合;混合池化

中图分类号:TP391.41

文献标识码:A

文章编号:2096-9759(2023)06-0090-04

3D Point Cloud Semantic Segmentation Method Fused with Multidimensional Spatial Information

WEI Dong,ZHANG Xiaohan,LIU Huan,SUN Tianyi,LI Changkai

(Shenyang University of Technology, Shenyang 110870,China)

Abstract: The 3D point cloud semantic segmentation method fused with multi-dimensional spatial information can process scene level 3D point cloud with low time complexity. Aiming at the problem that the lack of local features leads to the decline of perception ability, the normal vector estimation and hybrid pooling are introduced to enhance the local features. For the network structure that is difficult to handle the subtle semantic objects of large-scale 3D point clouds, multi-layer feature diversity can be preserved by using long-hop multi-layer fusion and short-hop multi-dimensional feature fusion. The experimental results show that the average intersection and union ratio of 6-fold cross-validation in the S3DIS (Stanford Large-Scale 3D Indoor Spaces Dataset) dataset is 70.2%.

Key words: semantic segmentation; multilayer fusion; normal vector estimation; multi dimension feature fusion

0 引言

随着近年来三维计算机视觉获得更多的关注,特别是无人驾驶、虚拟现实、机器人、智慧交通等应用得到了有效的发展。三维计算机视觉的表达方式包含深度图像、三维点云、体素化网格等。其中三维点云分割任务根据作用的对象不同分为:部分分割(细微级)、实例分割(对象级)和语义分割(场景级)^[1]。本文重点研究基于深度学习的三维点云语义分割。

本文方法是基于原始点云的语义分割方法,是一种针对不同规模的三维点云语义分割网络,能够在大规模室外场景三维点云或者连续多帧三维点云进行处理,并提高三维点云语义分割精度。针对基于原始点云的语义分割方法出现的局限性,所提的方法解决方案如下:(1)针对三维点云空间特征不足的问题:在编码层引入法向量估计增强空间特征。(2)针对随机采样导致局部区域感知力降低的问题:引入混合池

化增强局部区域感知力。(3)针对网络模型预测精度的问题:采用 6 层编码层,10 层解码层结构,并使用长跳跃多层特征融合和短跳跃多维特征融合,提高网络在更高尺度下的精度。

1 算法描述

本文方法的网络结构如图 1 所示,整体设计结构为端到端网络结构,由预处理层,编码层,解码层组成。因不同数据集提供原始特征维度不同,因此预处理层使用共享 MLP 提取原始三维点云的每个点的特征,并输出特定维度。编码层由 6 层组成,每层由短跳跃多维特征融合模块和随机采样两个部分组成,其中短跳跃多维特征融合模块由两组局部空间编码和混合池化组成。经过 6 层编码层后输入到 10 层解码层,解码层先通过最近邻插值(Nearest Neighbor)进行上采样,再与对应编码层进行长跳跃多层特征融合,并且每层之间使用共享 MLP 进行降维。

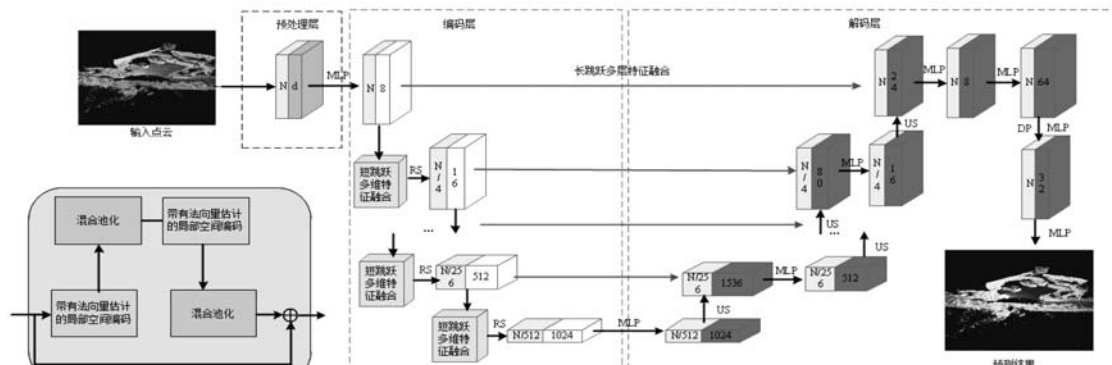


图 1 网络整体结构

Fig.1 Overall network architecture

收稿日期:2023-03-13

基金项目:辽宁省教育厅科学基金资助项目(No. LJGD2020006)。

作者简介:魏东(1968-),男,辽宁沈阳人,副教授,硕士,主要研究方向:计算机图形学;张潇瀚(1994-),男,辽宁大连人,硕士研究生,CCF 会员(C8072G),主要研究方向为点云语义分割。

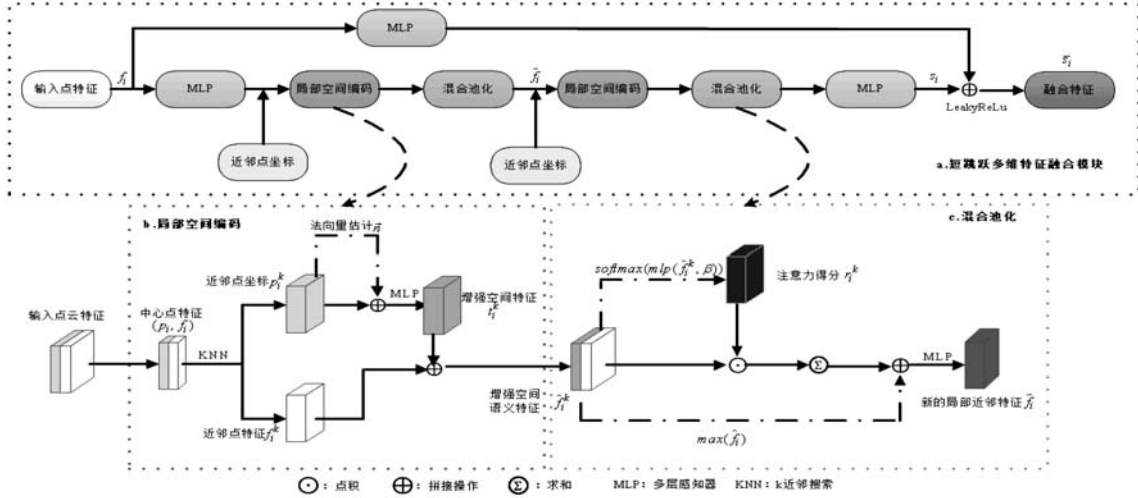


图2 短跳跃多维特征融合模块

Fig.2 Short jump multi-dimensional feature fusion module

1.1 共享 MLP 语义特征提取

整个网络结构的语义特征提取采用共享 MLP 的方式提取,由卷积层、批归一化、激活函数组成的多层感知器,将输入特征通过共享 MLP 输出为需要的特定特征维度,如公式(1)所示:

$$MLP = LReLU\left(BN\left(Conv_{1 \times 1}^{C_{out}}(\theta)\right)\right), \theta \in \mathbb{R}^{N \times C_m} \quad (1)$$

其中 MLP 表示多层感知器操作, $Conv$ 表示卷积操作并且下标 1×1 表示卷积核尺寸, 上标 C_{out} 表示输出维度, R 表示点集, C_m 表示输入特征维度, N 表示输入三维点云数量, θ 表示输入信息, BN 表示批归一化, $LReLU$ 表示使用 Leaky ReLU 的激活函数。

1.2 局部空间编码

局部空间编码结构如图2中b所示,局部空间编码方式主要为以下几个方面:(1)K近邻寻找局部区域点:包含收集局部区域点并构建索引;(2)法向量估计和欧式距离计算:法向量估计是基于局部表面拟合进行估计,法向量估计和欧式距离作为几何特征进行编码,增强网络结构的几何理解能力;(3)增强空间语义特征:将空间特征与语义特征相结合,使得构造的新特征具有更强的空间感知能力。

1.2.1 K近邻寻找局部区域点

为了使网络模型更好的感知细节,因此需要在编码层中构建局部区域。首先需要随机选取局部区域中心点集合为 $\chi = \{p_1, p_2, \dots, p_i\}, (1 \leq i \leq N)$, 并使用基于欧式距离的 KNN(K-Nearest Neighbor)算法收集 K 个近邻点并构建 KD-tree(K-Dimensional)索引,从而构建出局部区域。每个局部区域内点的表示即为 $P_i = \{p_1^i, p_2^i, \dots, p_K^i\}, (1 \leq i \leq N)$, 收集到的近邻点特征表示为 $F_i = \{f_1^i, f_2^i, \dots, f_K^i\}, (1 \leq i \leq N)$ 。 N 表示三维点云数量,其中下标 i 表示以每个局部区域中心点为 p_i 进行排序,上标 k 表示以中心点 i 构建的局部区域 K 个近邻点。

1.2.2 法向量估计

本文方法通过 K 近邻获取局部区域点后使用最小二乘法拟合局部平面,拟合平面中需要获取法向量并满足 $\|\vec{n}\|_2 = 1$, 因此采用主元分析法对协方差矩阵 W 进行特征值分解,其中协方差矩阵 W 的最小特征值所对应的特征向量,即为该局部区域拟合平面的法向量 \vec{n} 。构建局部区域拟合平面计算方法公

式(2)所示,协方差矩阵如公式(3)所示:

$$L(\vec{n}, d) = \underset{k=1}{\operatorname{argmin}} \sum_{k=1}^K (\vec{n} p_i^k - d), (1 \leq i \leq N) \quad (2)$$

$$W = \frac{1}{K} \sum_{k=1}^K (p_i^k - p_i)(p_i^k - p_i)^T, (1 \leq i \leq N) \quad (3)$$

其中 d 表示局部区域拟合平面到坐标原点距离, p_i^k 表示以 p_i 为中心的局部区域近邻点。

1.2.3 增强空间语义特征

通过寻找到的局部区域,将中心点坐标、局部区域内邻近点坐标、局部区域内近邻点坐标与局部区域的中心点坐标差、局部区域内点相对于中心点的欧式距离、法向量估计进行拼接,构造空间信息,其公式如(4)所示:

$$\phi_i^k = \operatorname{Concat}\left(\vec{n} p_i^k, p_i, p_i^k, (p_i - p_i^k), \|p_i - p_i^k\|\right) \quad (4)$$

其中 Concat 表示拼接操作, ϕ_i^k 表示空间信息, $\|p_i - p_i^k\|$ 表示局部区域内点相对于中心点的欧式距离。

通过 MLP 后得到局部空间特征 $t_i = \{t_i^1, t_i^2, \dots, t_i^k\}$, 与提取到的三维点云特征 \hat{f}_i^k 进行拼接操作,构造出增强空间语义特征 $\hat{f}_i = \{\hat{f}_i^1, \hat{f}_i^2, \dots, \hat{f}_i^k\}$, 使得整个网络结构更好的理解局部特征,如公式(5)、(5)所示:

$$t_i^k = MLP(\phi_i^k) \quad (5)$$

$$\hat{f}_i^k = \operatorname{Concat}(t_i^k, f_i^k) \quad (6)$$

1.3 混合池化

混合池化如图2中c所示,为了使得网络结构学习到更精细的局部区域特征,因此引入注意力池化,并通过与最大池化进行拼接用以弥补特征信息损失。将增强空间语义特征与权重超参数 β 作为输入,通过 MLP 和 softmax 组成的函数,求取每个点在局部区域的注意力得分 r_i^k , 将得到的注意力得分与构造后的新的特征 \hat{f}_i^k 进行点积操作,得到局部区域每个点的加权注意力特征,再进行求和就获得聚合后的局部区域特征,再将注意力池化与最大池化进行拼接输入进多层感知器 MLP, 如公式(7)、(8)所示:

$$\text{attentive}(\hat{f}_i) = \sum_{k=1}^k (r_i^k \cdot \hat{f}_i^k) \quad (7)$$

$$\tilde{f}_i = \text{MLP}\left(\text{Concat}\left(\text{attentive}(\hat{f}_i), \max(\hat{f}_i)\right)\right) \quad (8)$$

其中 $\text{attentive}(\cdot)$ 表示注意力池化, $\max(\cdot)$ 表示最大池化, \tilde{f}_i 表示新的局部近邻特征。

1.4 短跳跃多维特征融合模块

虽然网络模型随着层数增加可以扩大局部区域的感受野获取全局信息,但是三维点云的数量也逐步降低,从而会丢失一部分局部空间信息导致特征多样性的缺失。本文方法在编码层的每一层来扩大局部区域的感受野,使用多组连续的局部空间编码和混合池化逐步增加局部区域的感受野,这样在保留局部空间信息的同时逐步获取全局信息。由于随着层数加深会降低整体网络模型的计算效率,因此为了防止在训练过程中出现计算效率的损失同时防止梯度消失或梯度爆炸,采用如图2中a所示的短跳跃多维特征融合模块,使用连续的两组局部空间编码和混合池化。

首先将输入的三维点云特征通过 MLP 将维度下降至原来的一半,与局部区域近邻点坐标一起输入进第一组局部空间编码和混合池化,再将输出的语义特征 \tilde{f}_i 和局部区域近邻点坐标作为第二组局部空间编码和混合池化的输入,输出的局部扩大特征为 $s_i = \{s_i^1, s_i^2, \dots, s_i^k\}$,再将输入的三维点云特征通过 MLP 将维度提升一倍与局部扩大特征 s_i 拼接,最后通过 Leaky ReLu 函数进行激活,得到融合特征 $\bar{s}_i = \{\bar{s}_i^1, \bar{s}_i^2, \dots, \bar{s}_i^k\}$ 。

1.5 长跳跃多维特征融合

编码层与解码层间采用长跳跃多层特征融合,在获取全局信息同时保留下局部信息。如图1所示,编码层前3层随机采样保留25%的点,后3层随机采样保留50%点,即三维点云数量为 $(N, \frac{N}{4}, \frac{N}{16}, \frac{N}{64}, \frac{N}{128}, \frac{N}{256}, \frac{N}{512})$,从预处理后编码层三维点云特征维度变化为(8, 16, 64, 128, 256, 512, 1024)。解码层第1层为最后1层编码层通过后作为下一层解码层输入,从第2层到第7层解码层,每层先进行最近邻插值方法进行上采样,其中每层三维点云数量变化为 $(\frac{N}{512}, \frac{N}{256}, \frac{N}{128}, \frac{N}{64}, \frac{N}{16}, \frac{N}{4}, N)$,每层再与编码层进行长跳跃多层融合,解码层前7层三维点云特征维度变化为(1024, 512, 256, 128, 64, 16, 8)。如公式(10)所示:

$$s_j^D = \text{MLP}\left(\text{Concat}\left(s_j^D, s_{7-j}^E\right)\right), (2 \leq j \leq 7) \quad (10)$$

上式中 s_j^D 表示当前解码层特征,上标 D 表示解码层,下标表示层数, s_{7-j}^E 表示编码层特征,上标 E 表示编码层,其中 s_0^E 表示预处理层输出的特征。

1.6 损失函数

在训练过程中使用的总损失函数 L_{all} 是由多分类交叉熵损失 L_{CE} 和增强空间特征损失 $L(p_i)$ 组成。其中 L_{CE} 是作用在反向传播中预测多个语义标签置信度得分,如公式(11)所示:

$$L_{CE} = -\sum_{j=0}^{C-1} y_j \log(x_j) \quad (11)$$

C 表示样本标签数量, $y = \{y_0, \dots, y_{C-1}\}$ 表示使用 onehot 表示的样本语义标签, $x = \{x_0, \dots, x_{C-1}\}$ 每个元素表示样本属于第 j 类的概率。增强空间特征损失 $L(p_i)$ 是为了使局部区域更加紧密,增强空间语义特征可以约束局部区域,更具几何完整性。

通过使用 l_2 损失,使得局部区域内的近邻点更加靠近局部区域中心点 p_i ,如公式(12)所示:

$$L(p_i) = \left\| \frac{1}{K} \sum_{k=1}^K \text{Concat}(t_i^k, p_i^k) - p_i \right\|_2 \quad (12)$$

因为网络结构由多个局部空间编码堆叠每层网络局部空间编码范围不一,所以通过权重超参数来限制每层网络对局部区域收紧程度,权重超参数随着层数加深而增大,同时将多分类交叉熵损失和各层的增强空间损失求和构造出总体损失。总体损失函数如公式(13)所示:

$$L_{all} = L_{CE} + \sum_{m=1}^M \left(\lambda \cdot \sum_{i=1}^B L(p_i) \right) \quad (13)$$

式中 λ 表示权重超参数, M 表示编码层数, m 表示编码层数的序号, i 表示局部区域中心点序号, B 表示局部区域数量。

2 实验结果与分析

2.1 实验数据集及实验设置

本文方法采用 S3DIS^[10] 数据集来验证网络模型。网络模型在训练和测试中的实验环境采用 Ubuntu20.04 操作系统、Intel(R) i7 6800K CPU、NVIDIA RTX3090Ti GPU、以及 Tensorflow-1.15.0、CUDA11.0、Python3.6 框架实现。采用 Adam 优化网络结构,初始学习率为 0.01, dropout 失活率为 50%, KNN 搜索近邻点数量为 16,初始输入三维点云样本容量最大值为 45056。

2.2 实验结果与分析

表1是在S3DIS数据集上进行三维点云语义分割6折交叉验证获得的结果。本文方法采取的评估指标为总体精度(OA)、类别平均准确度(mAcc)、平均交并比(mIoU)。相比于其他网络模型,本文方法在mAcc、OA、mIoU指标下均表现较为优秀。PointNet^[3]/PointNet++^[4]均采用最大池化和最远点采样的方式,并且仅在第一层解码层只使用一次隔层特征拼接,在解码层很难获得更多特征信息。SPG^[5]采用图卷积的方法学习上下文,同时采用最大池化,存在语义信息的缺失。InsSem-SP^[8]使用一种训练自预测语义标签的方法使主干网络模型探索更多特征,从而提高总体语义分割精度,虽然OA比本文方法高0.6%,但是mAcc和mIoU均低于本文方法。KPCConv^[9]提出的Kernel卷积方法,使用适合复杂场景的可变形卷积提取特征,虽然mIoU比本文方法高0.4%但是mAcc相较本文方法低2.6%。

表1 S3DIS数据集的三维点云语义分割结果

Table 1 Semantic Segmentation Results of 3D Point Cloud of S3DIS Dataset

方法	mAcc(%)	OA(%)	mIoU(%)
PointNet ^[3]	66.2	78.6	47.6
PointNet++ ^[4]	67.1	81.0	54.5
SPG ^[5]	73.0	85.5	62.1
PointCNN ^[6]	75.6	88.1	65.4
PointWeb ^[7]	76.2	87.3	66.7
InsSem-SP ^[8]	74.3	88.5	64.1
KPCConv ^[9]	79.1	-	70.6
Ours	81.7	87.9	70.2

图3为本文方法在S3DIS进行可视化预测,因为该数据集提供三维坐标和RGB颜色作为原始输入特征,这导致墙壁与画板之间的边缘部分由于颜色变化不明显所以存在预测错误的情况,而在地面和墙壁等输入颜色明显的区域,三维点云语义分割表现优秀。由于在三维空间中不同物体的点云分布

基于深度学习的轻量化图像风格迁移方法研究

刘 瑛,徐亦丹

(南昌交通学院,江西 南昌 330100)

摘要: 由于传统图像风格迁移方法在风格化过程中会造成图像显著区域扭曲,导致艺术视觉效果不佳,为此文章研究基于深度学习的轻量化图像风格迁移方法。在图像风格迁移中,利用平滑处理、特征提取操作预处理风格化图像。通过显著区域检测,保留原始风格化图像中的显著区域。引入注意力机制结合轻量化卷积网络 SqueezeNet,设计一个轻量化风格迁移模型。在 INIT 数据集上的定性与定量测试实验结果表明,文章设计迁移方法生成的风格化图像具有良好的视觉效果,验证了该方法的性能优越性。

关键词: 深度学习;轻量化;图像风格迁移;注意力机制

中图分类号: TP391

文献标识码: A

文章编号: 2096-9759(2023)06-0093-04

0 引言

图像风格迁移,简单地说就是将一幅风格图像的风格经过处理后运用到另一幅图像上的转变过程,属于一种高超的图像编辑技术。该技术可以实现艺术大师作品的短期快速生成,为艺术创作工作带来了一定的启发。图像风格迁移方法的研究已经成为我国艺术领域的重点课题之一,文献[1]作者李文书等人引入高斯采样技术,设计一种图像局部相似区域的风格迁移;文献[2]作者孙天鹏等人通过 GAN 进行图像风格

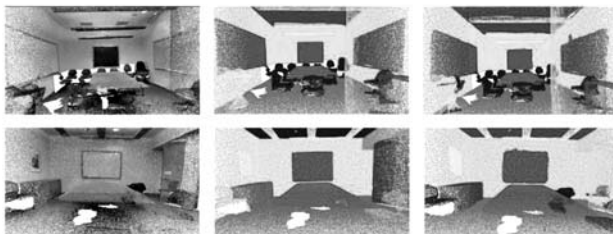
迁移,有效避免了图像风格迁移中细节丢失、色彩失真等问题;文献[3]作者兰天等人针对无监督的图像风格迁移受限等问题,利用对抗网络提升了迁移速度。但各式各样的新模型仍存在一定不足之处,大多数的图像风格迁移模型中算法较为复杂,难以在移动设备上运用。因此,本文针对图像风格迁移方法进行了深入研究,希望可以使迁移模型更加快速、轻量地在移动端环境运行,在学术界和实际的应用中具有很强的技术创新价值和实践意义。

收稿日期: 2023-02-03

基金项目: 2021 年度江西省教育厅科技项目:基于深度学习的轻量化图像风格迁移算法研究(项目编号:GJJ218408)。

作者简介: 刘瑛(1982-),女,江西南昌人,本科,讲师,研究方向:软件工程。

极不均匀,因此在细微部分容易出现预测错误的现象,但是总体预测结果接近真值结果,本文方法取得较好的效果。



输入点云 真值标签 预测结果

图3 S3DIS 数据集语义分割结果可视化

3 结语

本文提出一种深层端到端网络结构,构建了 6 层编码层和 10 层解码层结构,使用随机采样的方法以较低的时间复杂度对三维点云下采样。引入法向量估计增强空间特征编码从而提高特征多样性。构建混合池化融合空间特征和语义特征增强局部空间的感知能力。构建短跳跃多维特征融合模块在扩大局部区域范围的同时防止梯度消失和梯度爆炸并逐步从局部信息扩展到全局信息。构造长跳跃多层融合可以保留不同尺度下的特征信息。通过实验结果表明本文方法在多项评价指标中取得更优秀的性能。但是针对物体边缘处语义分割精度较低,以及场景中较少点云的物体类别语义分割精度较低。

参考文献:

- [1] 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):49-75.
- [2] Armeni I, Sener O, Zamir A R, et al. 3d semantic parsing of large-

scale indoor spaces[C]//Proceedings of the IEEE conference on computer vision and pattern recognition.IEEE,2016:1534-1543.

- [3] 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. IEEE,2017:652-660.
- [4] Qi C R, Yi L, Su H, et al. Pointnet++:Deep hierarchical feature learning on point sets in a metric space[J].Advances in neural information processing systems, 2017:5100-5109.
- [5] 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. IEEE,2018:4558-4567.
- [6] Li Y,Bu R,Sun M, et al.Pointcnn:Convolution on x-transformed points[J].Advances in neural information processing systems, 2018,31.
- [7] Zhao H, Jiang L,Fu C W,et al.Pointweb:Enhancing local neighborhood features for point cloud processing[C]// Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. IEEE, 2019:5565-5573.
- [8] Liu J,Yu M,Ni B,et al.Self-prediction for joint instance and semantic segmentation of point clouds[C]//European Conference on Computer Vision. Springer,Cham, 2020:187-204.
- [9] Thomas H,Qi C R,Deschaud J E,et al.Kpconv:Flexible and deformable convolution for point clouds[C]// Proceedings of the IEEE/CVF international conference on computer vision. IEEE/CVF,2019:6411-6420.