

基于密度感知和自注意力机制的点云分割算法

鲁斌^{1,2}, 刘亚伟^{1,2*}, 张宇航^{1,2}, 杨振宇^{1,2}

¹华北电力大学计算机系, 河北 保定 071003;

²河北省能源电力知识计算重点实验室, 河北 保定 071003

摘要 针对现有三维点云语义分割算法对点间密度信息以及空间位置特征利用不充分的问题,提出一种基于密度感知和自注意力机制的三维点云语义分割算法。首先,基于自适应K近邻(KNN)算法和局部密度位置编码构建密度感知卷积模块,从而有效地提取点间关键密度信息,加强初始输入特征的信息表达深度,提升算法捕获局部特征的能力。然后,构建空间特征自注意力模块,基于自注意力和空间注意力机制强化全局上下文信息和空间位置信息的关联性,对全局特征和局部特征进行有效聚合,从而提取更深层次的上下文特征,有效提升算法的分割性能。最后,在公开的S3DIS数据集和ScanNet数据集上进行了大量实验。实验结果表明,算法的平均交并比分别达到了69.11%和72.52%,与其他算法相比有明显提升,验证了所提算法有着良好的分割性能和泛化性能。

关键词 三维点云; 语义分割; 密度信息; 注意力机制; 密度位置编码

中图分类号 TP391.4

文献标志码 A

DOI: 10.3788/LOP231450

Point Cloud Segmentation Algorithm Based on Density Awareness and Self-Attention Mechanism

Lu Bin^{1,2}, Liu Yawei^{1,2*}, Zhang Yuhang^{1,2}, Yang Zhenyu^{1,2}

¹Department of Computer, North China Electric Power University, Baoding 071003, Hebei, China;

²Hebei Key Laboratory of Knowledge Computing for Energy & Power, Baoding 071003, Hebei, China

Abstract We propose a 3D point cloud semantic segmentation algorithm based on density awareness and self-attention mechanism to address the issue of insufficient utilization of inter point density information and spatial location features in existing 3D point cloud semantic segmentation algorithms. First, based on the adaptive K-Nearest Neighbor (KNN) algorithm and local density position encoding, a density awareness convolutional module is constructed to effectively extract key density information between points, enhance the depth of information expression of initial input features, and enhance the algorithm's ability to capture local features. Then, a spatial feature self-attention module is constructed to enhance the correlation between global contextual information and spatial location information based on self-attention and spatial-attention mechanisms. The global and local features are effectively aggregated to extract deeper contextual features, enhancing the segmentation performance of the algorithm. Finally, extensive experiments are conducted on the public S3DIS dataset and ScanNet dataset. The experimental results show that the mean intersection over union of our algorithm reaches 69.11% and 72.52%, respectively, shows significant improvement compared with other algorithms, verifying the proposed algorithm has good segmentation and generalization performances.

Key words 3D point clouds; semantic segmentation; density information; attention mechanism; density position encoding

1 引言

近年来,随着激光雷达扫描仪、红绿蓝-深度(RGB-D)相机和深度摄像机等三维采集设备的快速发展与不断成熟,三维点云的采集效率和采集精度得到

显著提升。与三维体素和三维网格等其他三维数据表示形式相比,三维点云不仅更容易获得,而且能以较小的数据规模表示丰富的三维信息。因此,三维点云在自动驾驶、工业制造和机器人等领域的应用越来越广泛^[1]。但是由于三维点云具有不均匀、不规则和无序等

收稿日期: 2023-06-02; 修回日期: 2023-07-04; 录用日期: 2023-07-31; 网络首发日期: 2023-08-18

基金项目: 河北省重点研发计划(20310103D)、河北省在读研究生创新能力培养资助项目(CXZZBS2023153)

通信作者: *1425655356@qq.com

特性,在三维点云场景中完成语义分割是一项具有挑战性的任务。根据处理三维点云的方式^[2],基于三维点云的语义分割方法分为基于多视图的方法、基于体素的方法和基于点的方法^[3]。

基于多视图^[4-7]的语义分割方法将不同视角下的三维点云投影到二维平面上,直接应用卷积神经网络(CNN)来提取获取的二维平面图的特征。该方法在投影转换过程中容易受到视角选择和投影遮挡等因素的影响,并且投影过程不仅会破坏点间的几何关系还会丢失大量关键空间信息,进而限制点云分割精度。基于体素^[8-10]的语义分割方法将点云转换为规则的三维网格后进行分割。点云体素化形成的特征分辨率是影响这类方法的重要因素。低分辨率特征可以加快训练速度,但会丢失空间信息;高分辨率特征保留了空间信息,但对计算资源有较高的需求,不利于方法的部署应用。因此,在保留足够特征的同时提升性能是设计这类方法的关键和难点。

基于点的语义分割方法直接将原始点云作为三维 CNN 的输入以获取点云的语义标签。Charles 等^[11]提出的 PointNet 利用多层感知机(MLP)逐点处理数据,以获得高维特征,然后使用最大池化操作降低特征维度并聚合特征信息,最终得到全局特征。然而,PointNet 只考虑点云的全局特征,忽略了点间的局部特征。随后提出的 PointNet++^[12]引入了点云分区策略,将点云进行分区后使用 PointNet 对每个区域进行特征提取,然后聚合特征以获取全局特征,进一步提高了算法的分割性能。尽管 PointNet 和 PointNet++ 使用最大池化操作保持了点云的排列不变性,但也导致了一些局部信息的丢失。之后,Li 等^[13]提出 PointCNN,该模型通过自定义算子在保持点云排列不变性的同时保留了局部信息。Wang 等^[14]提出的 DGCNN(dynamic graph CNN)采用了边缘卷积运算,利用 K 近邻(KNN)算法构建图结构,通过卷积和对称聚合函数获得点云的局部特征信息。Thomas 等^[15]提出的 KPConv(kernel point convolution)使用不同数量的核点构建核卷积,以提取点云的局部特征。这些方法在提取点云的局部特征时较为被动且不够灵活,分割性能仍有很大的提升空间。

为了更好地聚合点云的局部特征与全局特征^[16-17],一些研究者在三维点云语义分割中引入了注意力机制^[18]。注意力机制不仅可以增大算法对重要信息的关注权重,还可以灵活聚合全局特征与局部特征。Point Transformer^[19]使用 Transformer 处理点云,在分类分割任务中都取得了不错的结果。PCT(point cloud transformer)^[20]使用 Transformer 处理点云的无序性,通过注意力机制学习点云特征。Fu 等^[21]使用自注意力机制来选择通道和位置的关键上下文信息,以实现语义分割。这些方法的语义分割性能虽然有所提升,但没有考虑注意力机制的输入的处理方式,大都使用普通 CNN 处理原始点云,输入注意力机制的特征对

点云信息的表达不够丰富和全面。

针对以上问题,为了更准确地完成三维点云语义分割任务,本文提出一个基于密度感知和自注意力机制的点云分割算法。所提算法主要由密度感知卷积(DAC)模块和空间特征自注意力(SFSA)模块组成,前者充分利用三维点云密度信息,为注意力模块提供特征表达更加丰富的输入,而后者则结合点云的空间位置关系,加强自注意力机制对空间信息的提取和利用,实现局部特征与全局特征的充分聚合与利用。在真实场景数据集上进行了大量实验,结果表明,与现有算法相比,所提算法在语义分割性能上有着较为明显的优势。

2 所提算法

2.1 算法框架

所提三维点云语义分割算法的结构如图 1 所示。考虑到原始点云包含丰富的空间几何信息,为了降低计算代价,所提算法采用基于点的方法,将大规模点云数据划分为不同的批次,分批次进行输入,最后输出每个点的类别标签(类别标签有 m 个)。输入点云定义为包含 N 个点的集合,表示为 $P = \{p_1, p_2, \dots, p_N\} \in \mathbf{R}^{N \times d}$,其中每个点具有 d 维特征,包括 (x, y, z) 三维坐标和其他特征(如反射率等)。使用 X_i 表示第 i 个点的坐标向量, F_i 表示第 i 个点的特征向量。

算法采用“编码器-解码器”结构。首先使用一个 MLP 将尺寸为 $N \times d$ 的原始点云映射到高维特征空间,再将其输入 DAC 模块;使用自适应 KNN 算法查询邻域点,并基于局部密度位置编码获取点云的空间密度特征。之后结合预定义的权重矩阵得到动态卷积核,通过卷积操作和最大池化操作得到点云的局部特征。接着,将局部特征输入 SFSA 模块,基于全局自注意力和空间注意力的双注意力机制生成特征的注意力权重,实现对全局信息和局部信息的聚合,得到点云的全局特征。然后,将全局特征输入上采样(US)层,同时应用跳跃连接,级联对应的 DAC 模块的局部特征,有效融合高维空间特征和低维局部特征,并映射回原始点云空间。最后,使用 MLP 进行线性变换,获得三维点云的语义分割结果。

2.2 DAC 模块

三维 CNN 已经被广泛地应用于点云数据的处理,其感受野的设置对算法的语义分割性能有着直接的影响。三维 CNN 的感受野越大,其在原始点云中对应的作用范围就越大,卷积操作得到的特征就会包含更丰富的语义信息,但同时也会包含无用的冗余信息。反之,其感受野越小,对应的作用范围就越小,得到的特征对局部细节信息有更多的表达,但缺乏对点间关系等重要信息的足够表达。

为了更好地解决上述问题,提出了 DAC 模块,其结构如图 2 所示。首先,利用自适应 KNN 算法学习中心

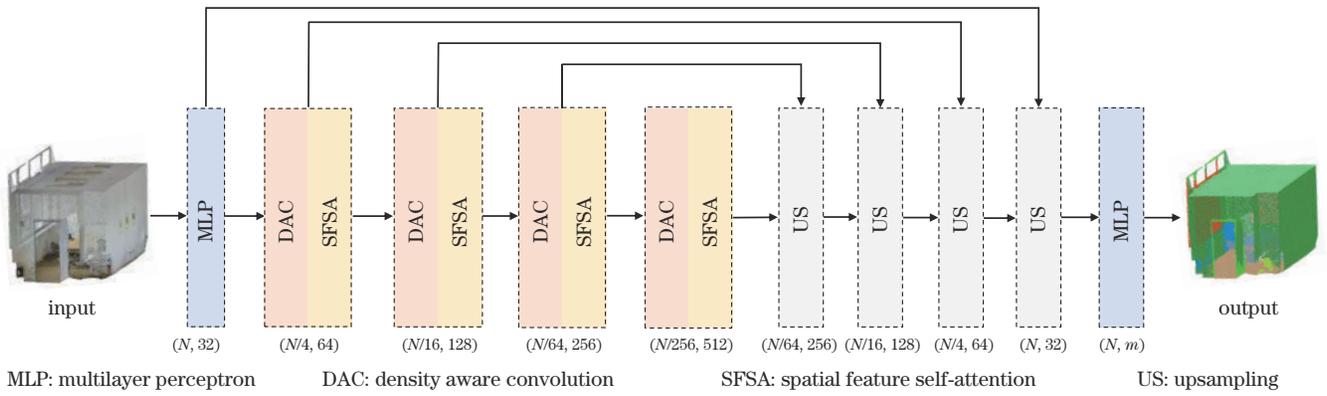


图 1 算法结构

Fig. 1 Algorithm structure

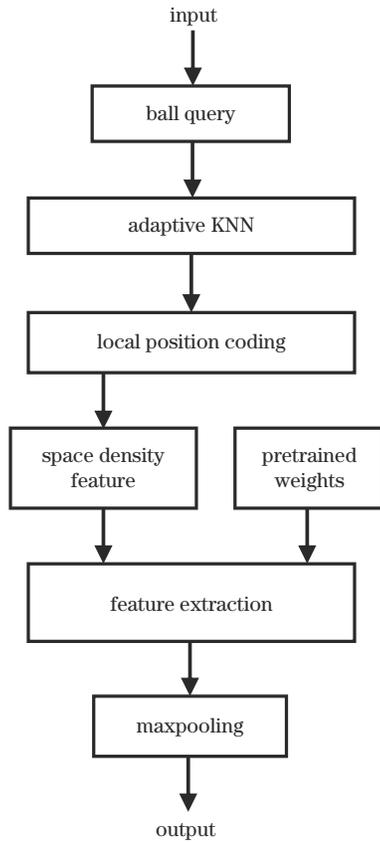


图 2 DAC 模块的结构

Fig. 2 Structure of DAC module

点与邻域点的点间关系,再结合由局部密度位置编码输出的点密度特征及预定义的权重矩阵,构建动态卷积核。DAC 模块可以自适应地选择感受野,结合点云密度信息构建卷积核,增强算法对三维点云空间特征的感知能力,动态地学习多层次的局部点间关系,有效提取点云的局部特征。

2.2.1 自适应 KNN 算法

在真实场景的三维点云中,常常存在密度分布不一致的区域,如图 3 所示。在这些区域使用固定 K 值的 KNN 算法时,如果点云十分密集,获取的点间关系会包含过多冗余信息;反之则无法提取足够的点间关

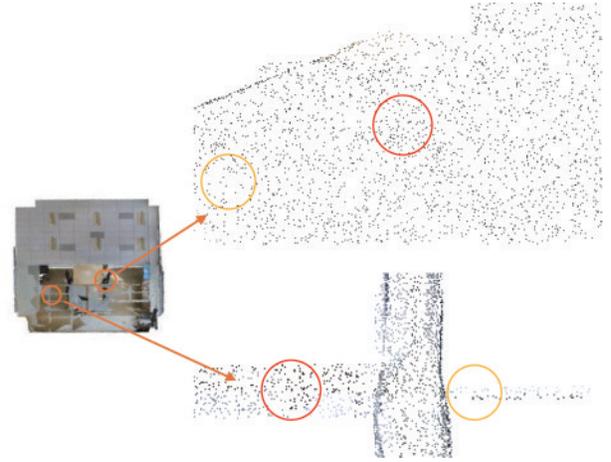


图 3 点云密度分布

Fig. 3 Density distribution of point clouds

系。点云关键空间信息的丢失或者冗余信息的增多都会对算法性能造成影响,选择合适的 K 值对于提高算法的分割精度来说十分重要。

为了应对三维点云局部密度不均匀的问题,提出自适应 KNN 算法。对于输入点云 P ,首先获取以 X_i 为中心点、半径 $r=0.2^{[12]}$ 范围内的所有邻域点 X_j ,其中 $j=1,2,\dots,N'$ 。然后计算中心点与其邻域点之间的欧氏距离以及所有邻域点与其中心点之间的平均距离,判断点云密度的稀疏程度,得到自适应系数。最后,将自适应系数与预定义 K 值相乘得到自适应 K 值,表示为

$$\varphi = \text{Sigmoid} \left[\text{Mean} \left(\sum_{i=1, j=1}^n d_{ij} \right) \right], \quad (1)$$

$$K' = K \times \varphi, \quad (2)$$

式中: φ 是自适应系数; K' 是自适应 K 值; K 是预定义 K 值; $\text{Sigmoid}(\cdot)$ 是激活函数; $\text{Mean}(\cdot)$ 是平均池化操作; $d_{ij} = \sqrt{(X_i - X_j)^2}$ 表示中心点 X_i 与其邻域点 X_j 之间的欧氏距离。

2.2.2 局部密度位置编码模块

真实场景中三维点云的密度信息可以准确地反映物体的空间分布情况。现有采集设备获取到的点云的

密度在分布上差异较大,这对点云密度信息的使用提出了较高要求。因此,提出局部密度位置编码模块,如图 4 所示。通过核密度估计(KDE)获取密度系数,然

后结合点的空间位置特征进行密度位置编码,得到局部结构的密度位置特征,从而获取更紧密的特征联系,学习潜在的局部特征。

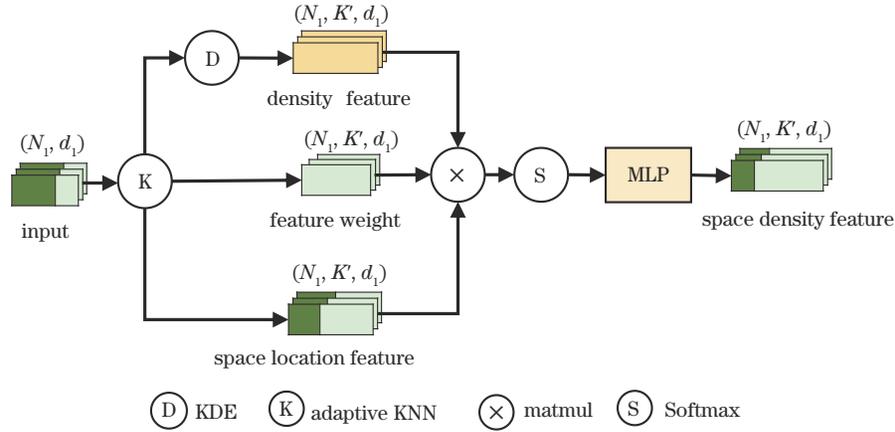


图 4 局部密度位置编码模块的结构

Fig. 4 Structure of local density position encoding module

具体地,利用自适应KNN算法对输入点云中的每个点 X_i 进行处理,得到以 X_i 为中心点的所有邻域点 X_j ,其中 $j=1,2,\dots,K'$ 。接着通过 KDE 估计每个点 X_i 的局部密度,对密度进行非线性变换得到密度特征,将密度特征、特征权重和空间位置特征进行矩阵乘法。最后利用 Softmax 和 MLP 进行处理,得到具有更大区分度的点的空间密度特征集合,即 $H=\{D_i|i=1,2,3,\dots,N'\}$ (D_i 是空间密度特征)。此过程表示为

$$\mathbf{W}(X_i, X_j) = \frac{1}{N' \delta} \sum_{i=1, j=1}^N \text{KEN}[d_{ij}/\delta], \quad (3)$$

$$\mathbf{D}_i = \text{MLP}\left\{\text{Softmax}\left[\mathbf{W}(X_i, X_j) \odot \mathbf{C}(X_i) \odot \mathbf{S}(X_i, X_j)\right]\right\}, \quad (4)$$

式中: N' 是当前点的数量; δ 是搜索带宽; $\text{KEN}(\cdot)$ 表示核函数; $\text{Softmax}(\cdot)$ 表示归一化函数; $\mathbf{W}(X_i, X_j)$ 表示点 X_i 的密度特征; 连续函数 $\mathbf{C}(X_i)$ 可以输出每个点对应的特征的权重,其输入是邻域点的相对坐标,用于描述中心点与邻域点之间的位置关系; $\mathbf{S}(X_i, X_j)$ 表示中心点 X_i 与其邻域点 X_j 之间的空间位置特征; \odot 表示哈达玛积,即对应位置的元素逐项相乘。

2.2.3 密度自适应特征提取模块

在 DAC 模块中,首先使用自适应 KNN 算法查找中心点的邻域点集合,接着利用局部密度位置编码得到点的空间密度特征,然后结合预定义的权重矩阵构建自适应卷积核^[22],实现密度自适应特征的提取,如图 5 所示。该过程表示为

$$\kappa = \sum_{i=1}^N (\mathbf{D}_i \mathbf{B}), \quad (5)$$

式中: \mathbf{B} 是预定义的权重矩阵; κ 是 DAC 核的集合。

基于对点的密度信息的有效利用,所提算法能够以动态自适应的方式构建卷积核,更加灵活地处理密度不均匀的三维点云,更加精确地分割复杂的局部结构,提

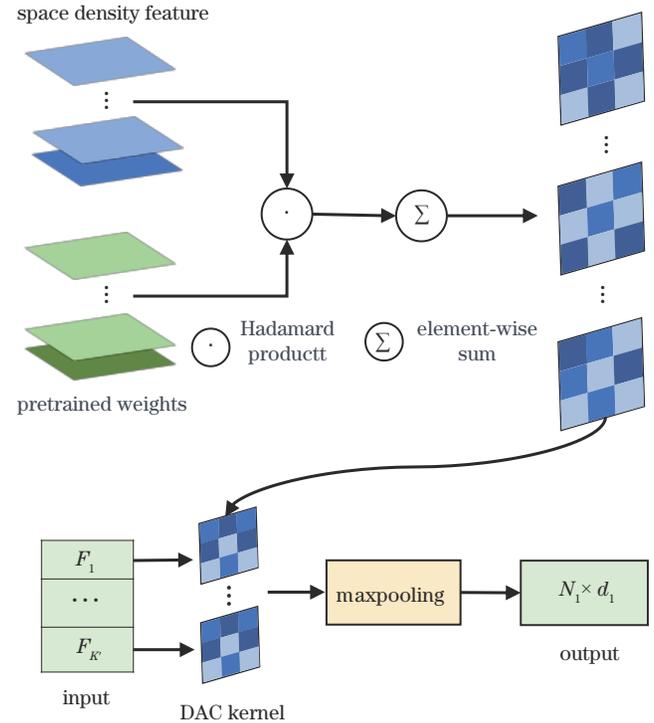


图 5 密度自适应特征提取模块的结构

Fig. 5 Structure of density adaptive feature extraction module

高分割性能。此外,利用预定义的权重矩阵而不是点的位置关系来构建卷积核,极大地降低了算法的复杂性,减少了计算资源的消耗,提高了算法的分割效率。

2.3 SFSA 模块

Transformer^[19]模型在自然语言处理方面展现了强大的性能,自注意力机制也随之受到了越来越多的关注。在计算机视觉领域,ViT^[23]等方法将自注意力机制应用在图像处理任务上,同样展现了不错的效果。自注意力机制可以很好地对数据的全局上下文特征进行有效聚合,这对提升点云分割的效率和精度有着重

要意义。尽管基于 CNN 的语义分割算法展现了一定的性能,但 CNN 受制于感受野和计算资源之间的平衡问题,不能很好地聚合点云场景中的全局特征。因此,

提出 SFSA 模块,引入自注意力机制和空间注意力机制来高效聚合全局特征,有效提升算法的分割性能。该模块的结构如图 6 所示。

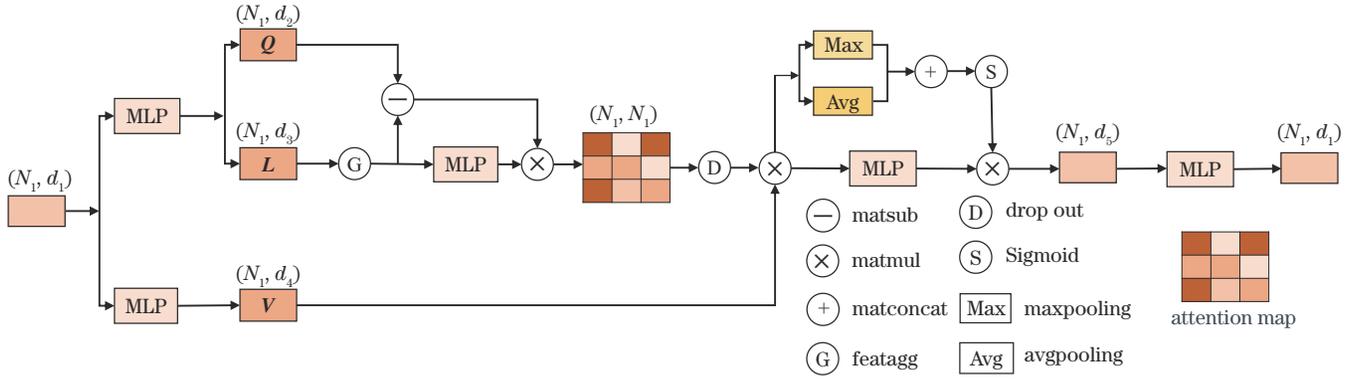


图 6 SFSA 模块的结构

Fig. 6 Structure of SFSA module

SFSA 模块由全局自注意力模块和空间注意力模块两部分组成。全局自注意力模块使用矩阵操作和不同的 MLP 自适应地生成和分配注意力权重,进而对点云场景的全局特征进行有效提取与聚合。空间注意力模块则基于两个池化操作对空间特征进行强化,为点云特征补充关键的空间位置信息。

2.3.1 全局自注意力模块

在全局自注意力模块中,首先使用两个 MLP 对输入的点云特征分别进行映射,得到查询矩阵、键矩阵和值矩阵。接着应用分组聚合将键矩阵分割出不同的区域,并将聚合后的键矩阵与查询矩阵相减,提取查询矩阵中的点与键矩阵中的点之间的空间关系。再与经过一个 MLP 层的键矩阵逐元素相乘得到全局注意力权重图,之后对注意力权重图进行随机失活操作,以缓解过拟合现象。然后,将值矩阵与经过随机失活的全局注意力权重图进行矩阵相乘,最终得到全局特征矩阵。通过全局自注意力机制,能够有效地捕捉点云中不同区域之间的依赖关系,并得到更加全局、准确的特征表示。同时,随机失活操作还能提高模型的泛化能力和鲁棒性。全局自注意力机制表示为

$$\omega = (L - Q) \odot \text{MLP}(L), \quad (6)$$

$$M = \text{Drop}(\omega) \odot V, \quad (7)$$

式中: Q 、 L 、 V 分别表示通过线性函数映射得到的输入特征的查询矩阵、键矩阵和值矩阵; ω 表示全局注意力权重图; $\text{Drop}(\cdot)$ 表示随机失活操作,用于增加算法的泛化能力; M 是输出的全局特征矩阵。点云特征经过全局自注意力模块后,特征间的相关性得到增强,特征对关键信息的表达更加丰富,感知范围也有一定程度的增大,全局的上下文信息得到了有效的聚合。

2.3.2 空间注意力模块

为了弥补全局自注意力模块对三维点云的重要空间特征以及特征间空间关联性的获取与利用不足,在

全局自注意力模块之后增加空间注意力模块,从而有效地提取和利用空间信息,增强算法对上下文空间特征的理解,更好地捕获特征间的空间语义信息。首先,对输入特征矩阵分别执行全局平均池化和全局最大池化操作,提取每个维度的特征信息并进行拼接,获取空间注意力权重。然后,应用激活函数以保证空间注意力权重非负,并得到空间注意力的权重图。最后将空间注意力权重图与应用 MLP 后的全局特征矩阵相乘得到注意力输出特征,接着通过一个 MLP 对输出特征进行维度变换得到 SFSA 模块的最终输出。具体表示为

$$G = \text{Sigmoid}[\text{Max}(M) \oplus \text{Avg}(M)], \quad (8)$$

$$T = \text{MLP}[\text{MLP}(M) \odot G], \quad (9)$$

式中: $\text{Max}(\cdot)$ 表示全局最大池化操作; $\text{Avg}(\cdot)$ 表示全局平均池化操作; $\text{Sigmoid}(\cdot)$ 表示激活函数; \oplus 表示矩阵拼接; G 表示空间注意力权重图; T 表示 SFSA 模块最终的输出特征。

点云中的每个点具有不同的相关性,类别相同的点的相关性大于类别不同的点。因为感受野的限制较大,所以每次卷积局限于局部的点云子集,导致模型缺乏捕获全局上下文信息和空间信息的能力。空间注意力机制可以获取整个输入点云的相关性分布图,强化点点的空间关联性,对特征相似的点进行语义关联,从而提高了算法对全局上下文信息和空间信息的理解能力,能更好地提取更深层次的全局特征。

3 实验与结果分析

实验环境如表 1 所示。训练参数与训练设置如下:批处理大小为 12,初始学习率为 0.006,使用多步长衰减策略调整学习率,调整倍率为 0.1,使用 AdamW 优化器进行优化。

3.1 数据集说明

所提算法的训练和验证均在真实世界三维点云室

表 1 实验环境

Table 1 Experimental environment

Environment	Version
System	Ubuntu 18.04
GPU	Nvidia RTX 3090
CUDA	11.3
PyTorch	1.8.0

内数据集 S3DIS 和 ScanNet 上进行。S3DIS 数据集是由斯坦福大学推出的真实室内场景三维点云语义分割数据集,包含 6 个室内区域(从 area 1 到 area 6), 总共 271 个房间的三维点云数据。每个室内区域包含大量具有语义标签的三维点,每个点都被表示为由三维坐标、RGB 信息和归一化空间坐标信息组成的 9 维向量。点云语义标签包含 13 个类别,包括地板、门和椅子等。ScanNet 数据集是由斯坦福大学、普林斯顿大学和慕尼黑工业大学共同推出的室内场景数据集,包含 20 个类别,包括墙、地板和桌子等,和 1513 个场景。训练使用其中的 1201 个场景,测试使用其中的 312 个场景。

3.2 评估指标说明

采用平均交并比(mIoU)、所有类别的平均精确率(mAcc)和总体精确率(OA)评估所提算法的性能。它们分别表示为

$$V_{mIoU} = \frac{1}{c} \sum_{i=1}^c \left[\frac{f_{i,i}}{\left(\sum_{j=1}^c f_{i,j} + \sum_{j=1}^c f_{j,i} - f_{i,i} \right)} \right], \quad (10)$$

$$V_{mAcc} = \frac{1}{c} \sum_{i=1}^c \left(\frac{f_{i,i}}{\sum_{j=1}^c f_{i,j}} \right), \quad (11)$$

$$V_{OA} = \frac{n_t}{N_T}, \quad (12)$$

式中: c 是点云的标签数; $f_{i,i}$ 是点云中预测与实际匹配的点数; $f_{i,j}$ 和 $f_{j,i}$ 是点云中预测与实际不匹配的点数; n_t 是点云中所有匹配准确的点数; N_T 是点云中的所有点数。

3.3 对比实验结果分析

表 2 和表 3 展示了所提算法与其他先进算法在 S3DIS 数据集 area 5 上的实验结果,表 4 和表 5 展示了所提算法与其他先进算法在 ScanNet 数据集上的实验结果。从表 2 可以看出,所提算法在 area 5 上的 mIoU 为 69.11%,与 KPConv^[15]相比提高了 2.01 个百分点,与 LLGF-Net^[24]相比提高了 1.11 个百分点。此外,所提算法的 mAcc 达到了 75.09%,与 KPConv^[15]相比提高了 2.29 个百分点,与 LLGF-Net^[24]相比提高了 0.69 个百分点。在类别标签方面,所提算法在 chair 和 bookcase 等类别上取得了最高的准确率。从表 4 可以看出,所提算法在 ScanNet 上的 mIoU 为 72.52%,与 KPConv 相比提高了 4.12 个百分点,与 FG-Net 相比提高了 3.52 个百分点。

表 2 S3DIS 数据集各类别分割结果的 IoU 比较

Table 2 Comparison of IoU of segmentation results for all categories on the S3DIS dataset

unit: %

Class	PointCNN ^[13]	DSPoInt ^[25]	BAAF-Net ^[26]	LLGF-Net ^[24]	Ours
Ceiling	92.31	94.20	92.90	94.10	93.81
Floor	98.24	98.10	97.90	98.20	98.52
Wall	79.41	82.40	82.30	85.10	85.99
Beam	0	0	0	0	0
Column	17.60	19.10	23.10	30.60	40.31
Window	22.77	49.90	65.50	60.60	60.03
Door	62.09	66.20	64.90	73.50	68.17
Table	74.39	85.60	78.50	89.50	84.95
Chair	80.59	78.20	87.50	79.60	91.09
Sofa	31.67	67.90	61.40	72.30	57.66
Bookcase	66.67	59.00	70.70	63.10	77.21
Board	62.05	62.30	68.70	79.80	80.69
Clutter	56.74	59.90	57.20	57.60	59.93

为了更好地展示所提算法的性能,选取部分 S3DIS 数据集中的场景数据进行可视化,如图 7 和图 8 所示。图 7 中从左到右依次为原始点云、真实标签、baseline 和所提算法分割标签。基于图 7 的结果可以分析得出,所提三维点云语义分割算法的分割结果与真实标签之间存在高度的一致性,较 baseline 的分割结果明显更好,表明该算法能够有效地实现点云数据的语义

分割任务。特别地,从图 8 中可以看出,对于桌椅等具有复杂几何形状的物体,该算法表现出了更加准确和真实的分割结果。这一结果表明,该算法具有较强的适应性和泛化性能,能够获取此类物体丰富的形状和细节信息,并且能够通过更有效地提取更准确的局部特征来捕获物体的几何信息,从而更好地地区分物体的边界,在不同场景下进行有效的语义分割任务。

表 3 S3DIS 数据集上分割结果的各评估指标的比较

Table 3 Comparison of evaluation indicators of segmentation results on the S3DIS dataset

Method	mIoU / %	mAcc / %
PointNet ^[11]	41.09	48.98
SegCloud ^[8]	48.92	57.35
PointWeb ^[27]	60.28	66.64
SegGCN ^[28]	63.60	70.44
PointCNN ^[13]	57.26	63.86
KPConv ^[15]	67.10	72.80
LLGF-Net ^[24]	68.00	74.40
Ours	69.11	75.09

表 4 ScanNet 数据集上分割结果的各评估指标比较

Table 4 Comparison of evaluation indicators of segmentation results on the ScanNet dataset

Method	mIoU / %	mAcc / %
PointNet ^[11]	14.69	19.90
PointNet++ ^[12]	34.26	43.77
PointConv ^[29]	55.60	—
RSNet ^[30]	39.35	48.37
FG-Net ^[31]	69.00	—
PointCNN ^[13]	43.70	57.90
KPConv ^[15]	68.40	—
Ours	72.52	80.63

表 5 ScanNet 数据集各类别分割结果的 IoU 比较

Table 5 Comparison of IoU of segmentation results for all categories on the ScanNet dataset

Class	PointNet++ ^[12]	PointCNN ^[13]	PointASNL ^[32]	KPConv ^[15]	Ours
Wall	77.48	74.50	80.60	81.90	85.62
Floor	92.50	90.70	95.10	93.50	95.10
Bed	51.32	56.10	78.10	75.80	80.66
Chair	64.55	68.80	83.00	81.40	89.75
Sofa	52.27	60.10	75.10	78.50	81.52
Table	46.60	55.30	55.30	61.40	72.29
Door	2.02	7.50	53.70	59.40	68.74
Desk	12.69	28.80	47.40	60.50	62.86
Sink	30.23	41.90	67.50	69.00	64.58
Toilet	31.37	73.40	81.60	88.20	92.20
Cabinet	23.81	26.40	65.50	64.70	65.50
Picture	0	1.30	27.90	18.10	36.56
Counter	20.04	22.50	47.10	47.30	61.26
Curtain	32.97	36.10	76.90	77.20	74.09
Window	3.56	11.00	70.30	63.20	68.24
Bathtub	42.72	73.50	70.30	84.70	87.70
Bookshelf	52.93	38.90	75.10	78.40	81.57
Refrigerator	18.51	43.40	63.50	58.70	64.06
Shower curtain	27.43	40.60	69.80	80.50	60.46
Other furniture	2.20	23.60	47.50	45.00	57.58

综合 S3DIS 和 ScanNet 数据集上的实验及可视化结果可知,所提算法对具有复杂几何结构的物体(如椅子、书柜)的检测效果最好。这是因为此类物体通常结构复杂,具有更多的几何结构和空间信息,点密度特征信息更多。而所提局部密度位置编码、自适应 KNN 算法和 DAC 模块能够提升算法处理点密度特征信息和点空间位置信息的能力,帮助算法有效提取点密度特征和点空间特征,增强多种特征间的语义关联性,提高算法对局部复杂特征的辨别能力以及对复杂几何结构物体的分割精度,使算法对密度分布不均匀的三维点云具有很好的鲁棒性。而对于 DAC 模块使用最大池化操作造成的信息丢失的问题,所提 SFSA 模块对丢失的信息进

行了补充,从而可以有效地获取全局语义信息和空间信息,准确地感知特征间的强弱联系并保持点云的整体连续性。同时,该模块还可以聚合全局语义特征和空间特征,进而有效提高算法的语义分割精度。

3.4 消融实验结果分析

3.4.1 自适应 K 值消融实验

为了探究自适应 KNN 算法中 K 值对算法分割精度的影响,在 S3DIS 数据集上进行消融实验,结果如表 6 所示。由表 6 可知,自适应 KNN 算法比普通 KNN 算法具有更好的分割效果,能够有效地增大感受野,学习点云的局部特征信息。

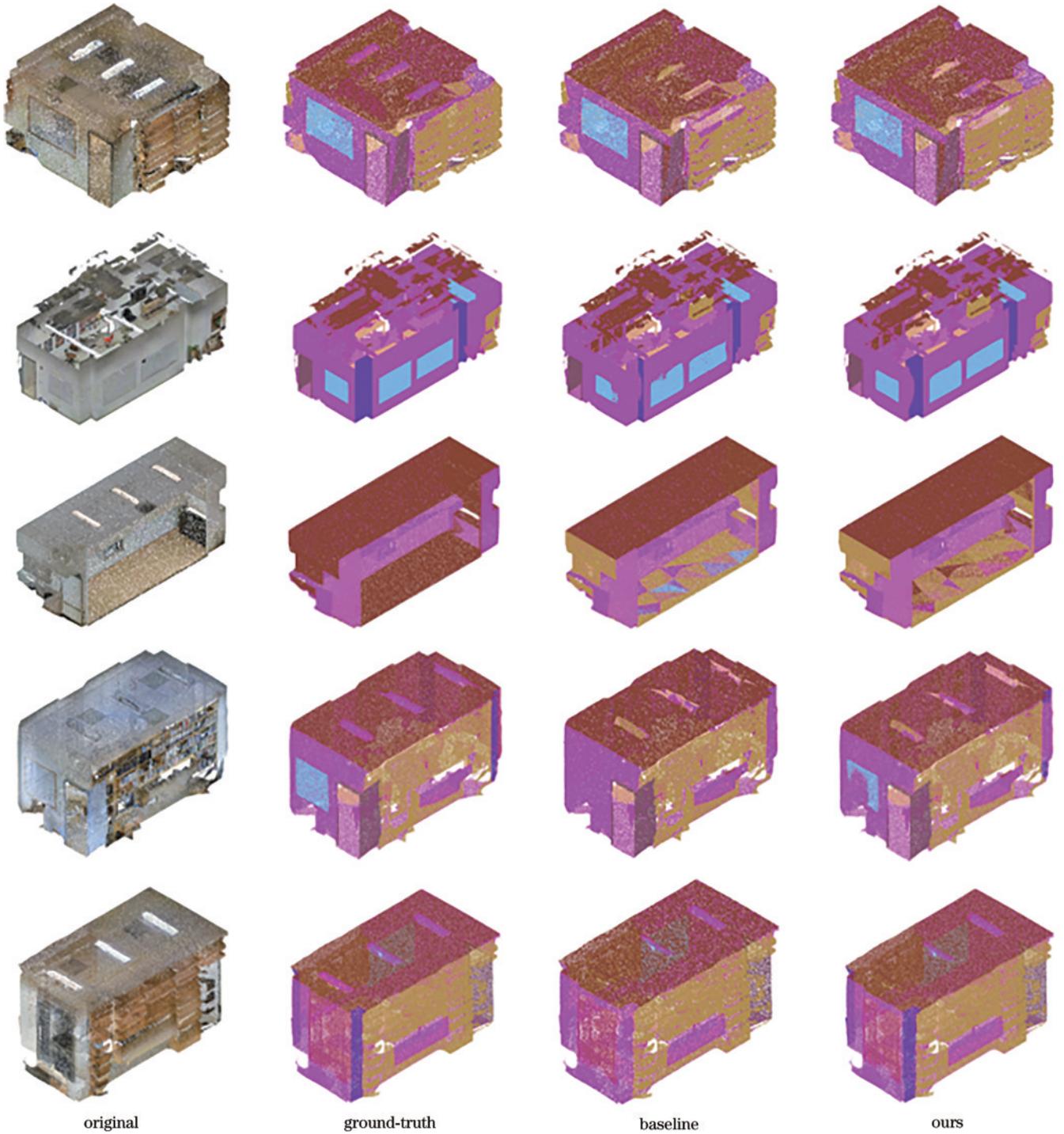


图 7 部分分割结果的可视化对比
Fig. 7 Visual comparison of partial segmentation results

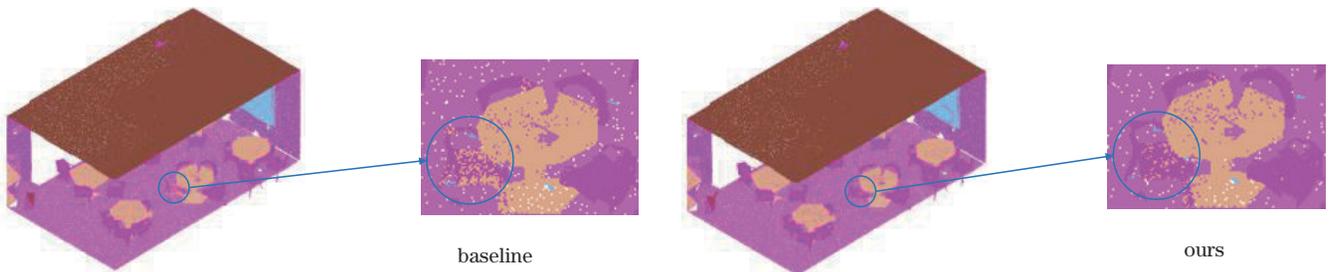


图 8 部分分割结果的可视化放大对比图
Fig. 8 Visual magnification comparison of partial segmentation results

表 6 不同 K 值的实验结果

Table 6 Experimental results of different K values

K value	KNN		Adaptive KNN	
	mIoU / %	mAcc / %	mIoU / %	mAcc / %
8	65.10	72.01	65.54	72.76
16	68.62	74.38	69.11	75.09
32	68.11	74.03	68.83	74.36
64	65.93	74.52	66.79	73.15

3.4.2 模块消融实验

为了进一步验证 DAC 模块和 SFSA 模块对提升算法分割性能的有效性,在 S3DIS 数据集上对两个模块分别进行消融实验。使用 area 5 进行验证,其余 5 个 area 为训练集,实验结果如表 7 所示,其中:Mod. 1 为 DAC 模块,Mod. 2 为 SFSA 模块。算法 0 将 PACConv^[22] 网络作为基线模型。算法 1 把算法 0 的特征提取模块替换为 DAC 模块,mIoU 比算法 0 提高了 0.68 个百分点。说明所提 DAC 模块通过关注点密度信息和空间位置信息能更有效地获取局部特征,提高算法识别和分割复杂结构物体的能力。算法 2 在算法 0 的基础上加入了 SFSA 模块,mIoU 比算法 0 提高了 0.21 个百分点。说明 SFSA 模块能利用自注意力和空间注意力对丢失的特征信息进行补充,并对多尺度的特征信息进行聚合,抑制无用特征,增强有用特征。算法 3 则同时采用了 DAC 和 SFSA 模块,算法的分割性能获得了进一步提升,mIoU 达到了 69.11%,比算法 0 提高了 2.53 个百分点,总体分割精度达到了 90.99%,而平均分割精度达到了 75.09%,比算法 0 提高了 2.09 个百分点。表 7 中的结果表明,所提算法通过着重关注点云的密度信息和空间信息,有效地提取和识别复杂结构的特征,同时聚合点云的局部特征和全局特征,显著提高了算法的分割精度,促进了点云语义分割任务的完成。

表 7 不同模块的消融实验结果

Table 7 Ablation experimental results of different modules

Algorithm	Mod. 1	Mod. 2	mIoU / %	OA / %	mAcc / %
0	×	×	66.58	—	73.00
1	√	×	67.26	89.70	73.13
2	×	√	66.79	89.67	73.57
3	√	√	69.11	90.99	75.09

3.4.3 子模块消融实验

为了探究 DAC 模块和 SFSA 模块中各个子模块对算法性能的影响,在 S3DIS 数据集上对各个子模块分别进行消融实验。使用 area 5 进行验证,其余 5 个 area 为训练集,实验结果如表 8 所示。Smod 表示不同的子模块,其中:Smod. 1 和 Smod. 2 分别是 DAC 模块的子模块,即自适应 KNN 和局部密度位置编码(当不采用自适应 KNN 和局部密度位置编码时,算法采用普

表 8 不同子模块的消融实验结果

Table 8 Ablation experimental results of different submodules

Algorithm	Smod.	Smod.	Smod.	mIoU	OA	mAcc
	1	2	3	/ %	/ %	/ %
0	×	×	√	66.74	89.65	73.23
1	√	×	√	67.37	89.42	73.61
2	×	√	√	67.52	89.61	73.88
3	√	√	√	69.11	90.99	75.09
4	√	√	×	68.32	89.86	74.69

通的 KNN 和位置编码);Smod. 3 是 SFSA 模块的子模块,即空间注意力子模块。从算法 0 与算法 1、2、3 的结果对比以及算法 3 与算法 4 的结果对比可以看出,自适应 KNN、局部密度位置编码和空间注意力子模块对提升算法的分割性能具有促进作用,同时应用这三个子模块的算法的分割性能最好。这表明自适应 KNN 和局部密度位置编码能够有效地缩放感受野,捕获点云中更多的局部特征和点间关系,同时能够利用密度特征信息来更好地处理不同尺度的物体对象,帮助算法更好地理解点云的空间分布;此外,空间注意力不仅能有效获取点云的重要空间特征,增强特征间的空间关联性,还能有效利用点云不同区域的上下文信息和空间特征,帮助算法更好地处理点云数据,提高分割结果的准确性和鲁棒性。

4 结 论

提出了一种基于密度感知和自注意力机制的语义分割算法。一方面,算法通过 DAC 模块有效地处理点云的密度信息和空间信息,实现感受野的动态变化,从而提取点间不同尺度的局部特征;另一方面,算法通过 SFSA 模块,利用自注意力和空间注意力双注意力机制对特征提取过程中的信息进行补充和增强,同时强化点间的空间关联性和语义关联性,聚合全局特征,完成语义分割任务。实验结果表明,所提算法在真实室内场景下具有较好的分割性能以及良好的泛化性能。

点云密度是点云的一个重要属性特征,它直接反映了点云的空间分布特点和疏密程度,包含丰富的几何信息和空间信息。因此未来的研究重点是深入探索点云密度特征的利用,继续提升算法的分割性能。

参 考 文 献

- [1] 赵亮, 胡杰, 刘汉, 等. 基于语义分割的深度学习激光点云三维目标检测[J]. 中国激光, 2021, 48(17): 1710004.
Zhao L, Hu J, Liu H, et al. Deep learning based on semantic segmentation for three-dimensional object detection from point clouds[J]. Chinese Journal of Lasers, 2021, 48(17): 1710004.
- [2] 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.
- [3] 姜枫, 顾庆, 郝慧珍, 等. 基于内容的图像分割方法综述[J]. 软件学报, 2017, 28(1): 160-183.
Jiang F, Gu Q, Hao H Z, et al. Survey on content-based image segmentation methods[J]. Journal of Software, 2017, 28(1): 160-183.
- [4] 张坤, 朱亚薇, 王晓红, 等. 基于空间图卷积的三维点云语义分割[J]. 激光与光电子学进展, 2023, 60(2): 0228007.
Zhang K, Zhu Y W, Wang X H, et al. Semantic segmentation of 3D point cloud based on spatial graph convolution[J]. Laser & Optoelectronics Progress, 2023, 60(2): 0228007.
- [5] 刘友群, 敖建锋, 潘仲泰. DGPoint: 用于三维点云语义分割的动态图卷积网络[J]. 激光与光电子学进展, 2022, 59(16): 1610014.
Liu Y Q, Ao J F, Pan Z T. DGPoint: dynamic graph convolution network for semantic segmentation of 3D point clouds[J]. Laser & Optoelectronics Progress, 2022, 59(16): 1610014.
- [6] Wu B C, Wan A, Yue X Y, et al. SqueezeSeg: convolutional neural nets with recurrent CRF for real-time road-object segmentation from 3D LiDAR point cloud[C]//2018 IEEE International Conference on Robotics and Automation (ICRA), May 21-25, 2018, Brisbane, QLD, Australia. New York: IEEE Press, 2018: 1887-1893.
- [7] Su H, Maji S, Kalogerakis E, et al. Multi-view convolutional neural networks for 3D shape recognition [C]//2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2016: 945-953.
- [8] Tchapmi 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, 2018: 537-547.
- [9] Riegler G, Ulusoy A O, Geiger A. OctNet: learning deep 3D representations at high resolutions[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6620-6629.
- [10] Wang P S, Liu Y, Guo Y X, et al. O-CNN: octree-based convolutional neural networks for 3D shape analysis [J]. ACM Transactions on Graphics, 36(4): 72.
- [11] 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.
- [12] Qi C R, Yi L, Su H, et al. PointNet++: deep hierarchical feature learning on point sets in a metric space [C]//Proceedings of the 31st International Conference on Neural Information Processing Systems, December 4-9, 2017, Long Beach, California, USA. New York: ACM Press, 2017: 5105-5114.
- [13] Li Y Y, Bu R, Sun M C, et al. PointCNN: convolution on X-transformed points[EB/OL]. (2018-01-23)[2023-02-03]. <https://arxiv.org/abs/1801.07791>.
- [14] Wang Y, Sun Y B, Liu Z W, et al. Dynamic graph CNN for learning on point clouds[J]. ACM Transactions on Graphics, 38(5): 146.
- [15] Thomas H, Qi C R, Deschaud J E, et al. KPConv: flexible and deformable convolution for point clouds[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Republic of Korea. New York: IEEE Press, 2020: 6410-6419.
- [16] 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]. 激光与光电子学进展, 2022, 59(16): 1610011.
Shang P F, Chen Y, Lü W J, et al. A semantic segmentation network for point cloud considering normal vector[J]. Laser & Optoelectronics Progress, 2022, 59(16): 1610011.
- [17] 郝雯, 王红霄, 汪洋. 结合空间注意力与形状特征的三维点云语义分割[J]. 激光与光电子学进展, 2022, 59(8): 0828004.
Hao W, Wang H X, Wang Y. Semantic segmentation of three-dimensional point cloud based on spatial attention and shape feature[J]. Laser & Optoelectronics Progress, 2022, 59(8): 0828004.
- [18] 李涛, 高志刚, 管晟媛, 等. 结合全局注意力机制的实时语义分割网络[J]. 智能系统学报, 2023, 18(2): 282-292.
Li T, Gao Z G, Guan S Y, et al. Global attention mechanism with real-time semantic segmentation network [J]. CAAI Transactions on Intelligent Systems, 2023, 18(2): 282-292.
- [19] Zhao H S, Jiang L, Jia J Y, et al. Point Transformer[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV), October 10-17, 2021, Montreal, QC, Canada. New York: IEEE Press, 2022: 16239-16248.
- [20] Guo M H, Cai J X, Liu Z N, et al. PCT: point cloud transformer[J]. Computational Visual Media, 2021, 7(2): 187-199.
- [21] Fu J, Liu J, Jiang J, et al. Scene segmentation with dual relation-aware attention network[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(6): 2547-2560.
- [22] Xu M T, Ding R Y, Zhao H S, et al. PACConv: position adaptive convolution with dynamic kernel assembling on point clouds[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE Press, 2021: 3172-3181.
- [23] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: transformers for image recognition at scale[EB/OL]. (2020-10-22) [2023-03-04]. <https://arxiv.org/abs/2010.11929>.
- [24] Zhang J Z, Li X W, Zhao X F, et al. LLGF-Net: learning local and global feature fusion for 3D point cloud semantic segmentation[J]. Electronics, 2022, 11(14): 2191.

- [25] Zhang R R, Zeng Z Y, Guo Z Y, et al. DSPoint: dual-scale point cloud recognition with high-frequency fusion [EB/OL]. (2021-11-19)[2023-03-05]. <https://arxiv.org/abs/2111.10332>.
- [26] Qiu S, Anwar S, Barnes N. Semantic segmentation for real point cloud scenes via bilateral augmentation and adaptive fusion[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE Press, 2021: 1757-1767.
- [27] Jiang L, Zhao H S, Liu S, et al. Hierarchical point-edge interaction network for point cloud semantic segmentation [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Republic of Korea. New York: IEEE Press, 2020: 10432-10440.
- [28] Lei H, Akhtar N, Mian A. SegGCN: efficient 3D point cloud segmentation with fuzzy spherical kernel[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 11608-11617.
- [29] Wu W X, Qi Z A, Li F X. 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, 2020: 9613-9622.
- [30] 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.
- [31] Liu K C, Gao Z, Lin F, et al. FG-Net: fast large-scale LiDAR point clouds understanding network leveraging correlated feature mining and geometric-aware modelling [EB/OL]. (2020-12-17)[2023-03-05]. <https://arxiv.org/abs/2012.09439>.
- [32] 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.