http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2022.0775

融合双注意力机制和动态图卷积的点云语义分割

杨军[1,2,*]，张琛[1]

(1. 兰州交通大学 电子与信息工程学院，兰州 730070；2. 兰州交通大学 测绘与地理信息学院，兰州 730070)

摘 要：针对现有基于深度学习的三维点云语义分割算法通常在提取局部特征时忽略邻域点间深层次语义信息，聚合局部邻域特征时忽略其他邻域特征中有用信息的问题，提出融合双注意力机制和动态图卷积神经网络(DGCNN)的三维点云语义分割算法。通过动态图卷积操作构造边缘特征，并将中心点与邻域点的相对距离输入到核点卷积操作得到增强后的边缘特征，进一步加强中心点与邻域点之间的联系；引入空间注意力模块以建立邻域点之间的依赖关系，将特征相似点相互关联，从而在局部邻域内提取到更深层次的上下文信息，丰富邻域点的几何特征；在聚合局部邻域特征时引入通道注意力模块，通过给不同通道赋予不同的权值以达到增强有用通道同时抑制无用通道的目的，从而提高语义分割的准确率。在S3DIS数据集和SemanticKITTI数据集上的实验结果表明：所提算法的语义分割精度分别达到了66.0%和59.4%，与其他经典的网络模型相比，取得了更好的点云分割效果。

关 键 词：三维点云；语义分割；深度学习；注意力机制；动态图卷积
中图分类号：TP391
文献标志码：A **文章编号**：1001-5965(2024)10-2984-11

近年来，随着深度传感器技术(如激光雷达设备、RGB-D相机)的不断发展，三维点云被广泛应用于室内导航[1]、机器人[2]、自动驾驶[3]、三维重建[4]等领域。作为计算机视觉长期研究的课题之一，语义分割旨在利用计算机对场景进行逐点分类，并将场景分割成若干具有特定语义类别的区域，是三维场景理解和分析的基础[5]。然而，由于点云数据的不规则、无序性和稀疏性等特点，使自动、精确的点云语义分割成为一项具有挑战性的任务，设计针对大规模、复杂场景三维点云的语义分割网络模型成为当前的研究重点。早期学者们借鉴二维图像处理的方法，将三维点云数据转化为二维数据，进而使用二维图像处理中较为成熟的卷积神经网络(convolutional neural network, CNN)、全连接神经网络(fully connected neural network, FCN)算法，则基于投影和基于体素化的方法应运而生。基于投影的方法先获取三维点云形状在不同视角下的二维图像，再对每个视图进行卷积处理，通过池化层和全连接层将每个视角得到的特征进行聚合，得到最终的语义分割结果，但投影过程不可避免地会产生信息损失，导致分割结果并不理想。基于体素化的方法将点云规则化为网格结构，很大程度上保留了物体的几何信息，但该结构仍然无法细分物体边界的几何信息，并且存在大量无效、冗余的信息，大大增加了计算量，空间复杂度高。PointNet[6]提出直接对点云数据进行处理，但其注重独立提取点云全局特征而忽略了点之间的几何关联，从而丢失了许多局部特征。为解决该问题，文献[7]在动态图卷积神经网络(dynamic graph convolutional neural network, DGCNN)中提出了边缘卷积，但在实验过程中发

收稿日期：2022-09-14；**录用日期**：2023-01-02；**网络出版时间**：2023-01-10 13:51
网络出版地址：link.cnki.net/urlid/11.2625.V.20230109.1713.007
基金项目：国家自然科学基金(42261067,61862039)；兰州市人才创新创业项目(2020-RC-22)；兰州交通大学天佑创新团队(TY202002)
*通信作者. E-mail：yangj@mail.lzjtu.cn

引用格式：杨军，张琛. 融合双注意力机制和动态图卷积的点云语义分割[J]. 北京航空航天大学学报，2024，50(10)：2984-2994.
YANG J，ZHANG C. Semantic segmentation of point clouds by fusing dual attention mechanism and dynamic graph convolution[J]. Journal of Beijing University of Aeronautics and Astronautics，2024，50(10)：2984-2994 (in Chinese).

现, 该方法仍会丢失局部深层次语义信息。

针对上述问题, 本文提出一种融合双注意力机制和 DGCNN 的架构。主要创新点如下:

1) 构建空间注意力模块, 用来建立局部邻域内的邻域点间的依赖关系, 提取更深层次的语义信息。

2) 构建通道注意力模块, 用来增强有用的通道, 并尽可能保留局部邻域内的重要信息, 从而提高网络的训练效果。

1　相关工作

目前, 基于深度学习的三维点云语义分割算法主要分为三大类: 基于投影的方法、基于体素化的方法和基于点云的方法。

1.1　基于投影的方法

基于投影的方法通常将三维点云投影到二维图像中, 利用二维卷积方法提取三维模型的表面信息。Tatarchenko 等[8] 提出了 TangentConv 网络, 将点云投影到曲面上, 利用切线卷积进行点云分割, 但该方法会丢失点云的几何和结构信息。Wu 等[9] 提出并设计了一种利用球面投影完成三维点云语义分割的 SqueezeSeg 网络, 该方法先将三维点云投影到球面中, 每个球面都是一张二维图像, 再利用 SqueezeNet 网络进行特征提取, 最终通过条件随机场(conditional random field, CRF)判别模型的优化, 实现对输入的三维点云进行每个点的语义信息预测。之后, 为优化点云语义信息预测的结果, 利用无监督的管道方法提出了 SqueezeSegV2 网络[10], 并提升了点云语义分割的精度。Milioto 等[11] 提出了 RangeNet++网络, 将二维范围图像的语义标签传输到三维点云, 使用基于 GPU 加速的后处理方法来减轻离散化错误和推理输出模糊的问题。球形投影虽然保留了更多信息, 并且适合于激光雷达 LiDAR 点云的标记, 但这种通过投影的中间表示仍会不可避免地损失一些几何和结构信息。

1.2　基于体素化的方法

体素化方法主要是将不规则的点云转化为规则的体素网格, 利用标准的三维卷积分析和处理。VoxNet 模型[12] 最早将点云采样成体积网格后输入到 Volumetric CNN 网络中实现点云分割, 相较于二维图像数据, 其计算开销更大, 效率也更低。针对体素网格低分辨率的限制, SEGCloud 网络[13] 引入了三线性插值(trilinear interpolation, TI)和全连接条件随机场(fully connected conditional random field, FCCRF)来实现细粒度和全局一致性。为减少不必要的计算和内存消耗, Riegler 等[14] 采用八叉树构建网络 OctNet, 用于将存储器分配和计算集中到相关的密集区域, 在不影响分辨率的情况下实现更深层网络; Zeng 和 Gevers[15] 构建了基于 KD-tree 的 KD-Net, 用于提高点云计算和存储效率, 但此类树结构方法依赖体素边界, 并未充分利用局部几何结构。

1.3　基于点云的方法

基于点云的方法直接对不规则点云进行处理。PointNet++[16] 是在原有 PointNet 的基础上对点的特征的构建提出了更加有效的方法, 其利用点的邻域空间, 通过下采样的方式同时构建每个采样点的局部空间, 对每个局部空间利用 PointNet 的基础结构进行空间的特征提取, 并进行对应的上采样特征插值, 最终得到了每个点的特征, 但忽略了单个点与其邻域点的几何关系。针对这一问题, 文献 [17] 提出了一种多特征融合方法, 通过构建注意力融合层学习全局单点特征和局部几何特征的隐含关系来充分挖掘更能表征模型类别的细粒度几何特征, 并将全局单点特征和细粒度几何特征进一步融合, 达到优势互补、增强特征丰富性的目的。然而, 点云数据中, 并不是所有点之间都有关联, 有些点与点之间的关系强度很低。为此, Zhao 等[18] 设计了一种基于点对联系的点云分割网络 PointWeb, 其主要模块为自适应特征调整(adaptive feature adjustment, AFA), 该算法完整描述了在点集构成的局部空间中, 每个点与其余点之间的强度联系, 可以根据这种联系使每个点都能从所有其他点收集特征, 从而增强特征描述局部邻域的能力, 但对于点云边缘部分的分割效果并不理想。为解决这一问题, Wang 等[19] 提出一种图注意力卷积网络 GACNet, 主要通过建立每个点与周围点的图结构, 并通过引入注意力机制计算中心点与每一个邻接点的边缘权重, 根据动态学习到的特征有选择地关注它们中最相关的部分, 捕获点云的结构化特征以进行细粒度分割, 但该算法在一定程度上忽略了邻域点间的深层次语义信息。

综上所述, 目前大多数基于深度学习的点云语义分割算法过分关注中心点与邻域点的特征提取, 而忽略了邻域点间相互结构关系的学习, 并且在聚合局部邻域特征时, 通常选择最大值特征来代表局部邻域特征, 导致丢失一些具有鉴别性的特征。因此, 如何在局部邻域内将中心点的局部特征与邻域点间的深层次语义特征相结合, 并在聚合局部邻域特征时保留其他通道的重要特征是三维点云语义分割研究的一个亟待解决的问题。

2　融合双注意力机制和动态图卷积神经网络

本文构建的融合双注意力机制和 DGCNN 的架

构主要由改进的边缘卷积(EdgeConv++)模块、空间注意力机制(spatial attention mechanism, SAM)模块和通道注意力机制(channel attention mechanism, CAM)模块组成,网络结构如图1所示。在提取局部邻域特征时,首先,通过边缘卷积操作EdgeConv++构造边缘特征,并将中心点与邻域点的相对距离 $\boldsymbol{y}_{ij}$ 输入到核点卷积(kernel point convolution, KPConv模块)[20],得到增强后的局部特征 $\boldsymbol{c}_{ij}$,通过多层感知机(multilayer perceptron, MLP)将局部特征和全局特征连接起来得到边缘特征 $\boldsymbol{h}_{ij}$;然后,通过SAM模块学习局部邻域内的邻域点深层语义特征,并对边缘特征进行增强,通过CAM模块对边缘特征各通道赋予不同的权重;最后,使用最大池化操作得到聚合后的局部邻域特征作为中心点的新特征。

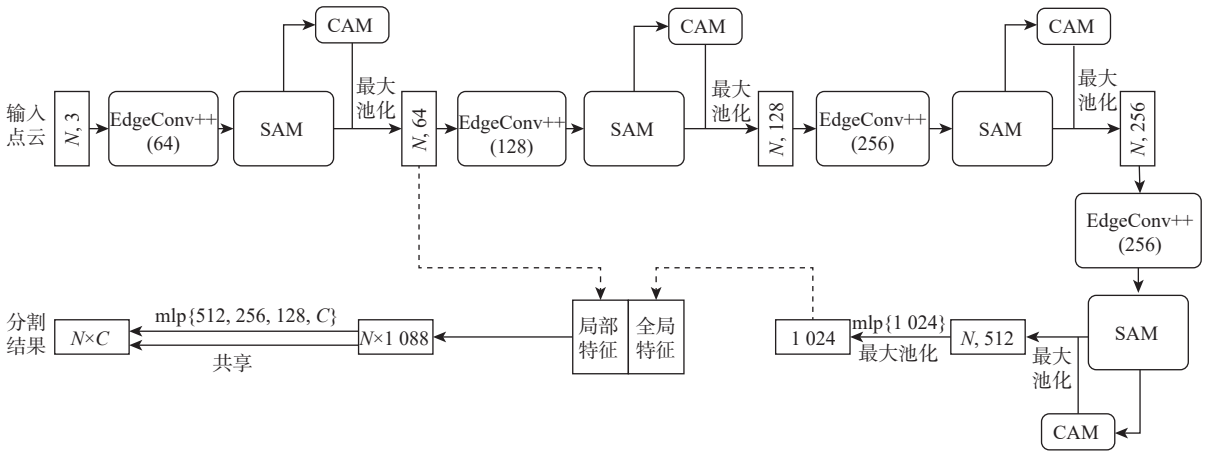


图1 融合双注意力机制和DGCNN的模型结构

Fig. 1 Structure of model fusing dual attention mechanism and DGCNN

2.1 EdgeConv++模块

对EdgeConv模块进行改进,与DGCNN中定义的边缘特征 $\boldsymbol{e}_{ij}$ 相似,既考虑中心点 $\boldsymbol{x}_i$ 的全局信息,也考虑中心点与邻域点的局部信息,如下:

$$\boldsymbol{e}_{ij}=l_\theta\left(\boldsymbol{x}_i,\boldsymbol{x}_{ij}-\boldsymbol{x}_i\right) \tag{1}$$

式中:$l_\theta:\mathbf{R}^F\times\mathbf{R}^F\to\mathbf{R}^{F'}$ 为具有一组可学习参数 θ 组成的非线性函数。

但本文的局部信息不只是通过计算中心点与邻域点的相对距离 $\boldsymbol{y}_{ij}=\boldsymbol{x}_{ij}-\boldsymbol{x}_i$,而是受KPConv的启发。将 $\boldsymbol{y}_{ij}$ 作为输入,经过一个KPConv后得到增强后的局部特征 $\boldsymbol{c}_{ij}$,KPConv的计算过程如图2所示。将全局特征与局部特征拼接形成新的边缘特征 $\boldsymbol{h}_{ij}$,如下:

$$\boldsymbol{h}_{ij}=l_\theta\left(\boldsymbol{x}_i,\boldsymbol{c}_{ij}\right)=l_\theta\left(\boldsymbol{x}_i,\mathrm{KPConv}\left(\boldsymbol{y}_{ij}\right)\right) \tag{2}$$

EdgeConv++模块结构如图3所示。

具体做法为:首先,对于点云中的任意一个点 $\boldsymbol{x}_i$,通过K最近邻(K-nearest neighbor, KNN)算法构建一个以点 $\boldsymbol{x}_i$ 为中心点的局部邻域图,得到 $\boldsymbol{x}_i$ 的 k 个邻居点 $N(\boldsymbol{x}_i)=\{\boldsymbol{x}_{i1},\boldsymbol{x}_{i2},\cdots,\boldsymbol{x}_{ik}\}$。然后,以 $\boldsymbol{x}_i$ 为球心,以 $\boldsymbol{x}_i$ 到 $N(\boldsymbol{x}_i)$ 邻居点的最远距离为半径 r 构建一个球。在球体内,找 d 个点作为核心点 $\tilde{\boldsymbol{x}}_d$,核心点并不是点云中的点,而是通过球心与各核心点之间的引力和斥力得到的。对于每个核心点 $\tilde{\boldsymbol{x}}_d$,分别有一个权重矩阵 $\boldsymbol{W}_d$ 与之对应。对于任意的 $\boldsymbol{y}_{ij}$,定义核函数 $\boldsymbol{g}$ 为

$$\boldsymbol{g}\left(\boldsymbol{y}_{ij}\right)=\sum_{d<D}h\left(\boldsymbol{y}_{ij},\tilde{\boldsymbol{x}}_d\right)\boldsymbol{W}_d \tag{3}$$

式中:D 为核心点总数。

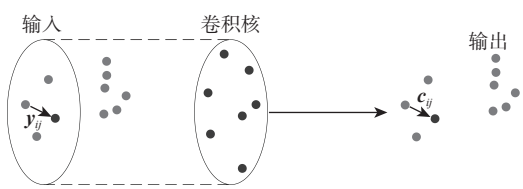


图2 KPConv模块结构示意图

Fig. 2 Structure of KPConv module

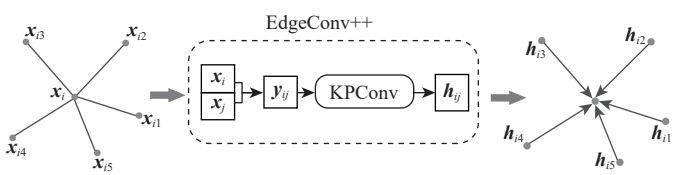


图3 EdgeConv++模块结构示意图

Fig. 3 Structure of EdgeConv++ module

可以看出, $\boldsymbol{g}(\boldsymbol{y}_{ij})$ 相当于对 d 个权重矩阵加权求和。权重系数 $h(\boldsymbol{y}_{ij},\tilde{\boldsymbol{x}}_d)$ 定义为

$$h\left(\boldsymbol{y}_{ij},\tilde{\boldsymbol{x}}_d\right)=\max\left(0,\ 1-\frac{\left\|\boldsymbol{y}_{ij}-\tilde{\boldsymbol{x}}_d\right\|}{\beta}\right) \tag{4}$$

式中: β 为核心点的影响距离, 用来确保每个核心点的影响区域之间有重叠, 其大小由点云的输入密度确定。

每个权重矩阵 $\boldsymbol{W}_d$ 的系数由 $\boldsymbol{W}_d$ 对应的核心点 $\tilde{\boldsymbol{x}}_d$ 到 $\boldsymbol{y}_{ij}$ 的相对距离来确定。相对距离越小, 权重越大, 最大值为 1; 相对距离越大, 权重越小, 最小值为 0。即距离越近, 相关性越大; 反之亦然。$\boldsymbol{g}(\boldsymbol{y}_{ij})$ 可以看作专门针对每个 x_{ij} 计算出来一个权重矩阵, 利用该权重矩阵对局部特征 $\boldsymbol{y}_{ij}$ 进行特征变换得到 $\boldsymbol{c}_{ij}$, 此时得到的 $\boldsymbol{c}_{ij}$ 既包含了中心点与邻域点的局部几何信息, 也包含了邻域点与核心点之间的位置信息。

$$\boldsymbol{c}_{ij}=\boldsymbol{g}\left(\boldsymbol{y}_{ij}\right)\boldsymbol{y}_{ij} \tag{5}$$

$\boldsymbol{g}(\boldsymbol{y}_{ij})$ 可通过邻域点与核心点之间的相对位置关系计算得到, 因此, 在 KPConv 中直接将 $\boldsymbol{c}_{ij}$ 当作边缘特征而没有考虑中心点的全局特征, 核心点虽然由中心点生成, 可在一定程度上保留原始中心点的全局信息, 但还是会不可避免地导致中心点的描述能力下降, 进而丢失部分全局特征。

2.2　SAM 模块

现有网络中的注意力机制[17,19] 侧重于采样点与邻域内的局部信息交互, 普遍忽略了邻域点相互结构关系的学习, 而空间注意力的引入可以建立邻域点之间的依赖关系, 无论距离远近, 只要特征相似就会相互关联, 从而在局部邻域内提取到更深层次的邻域上下文信息。SAM 模块结构如图 4 所示。

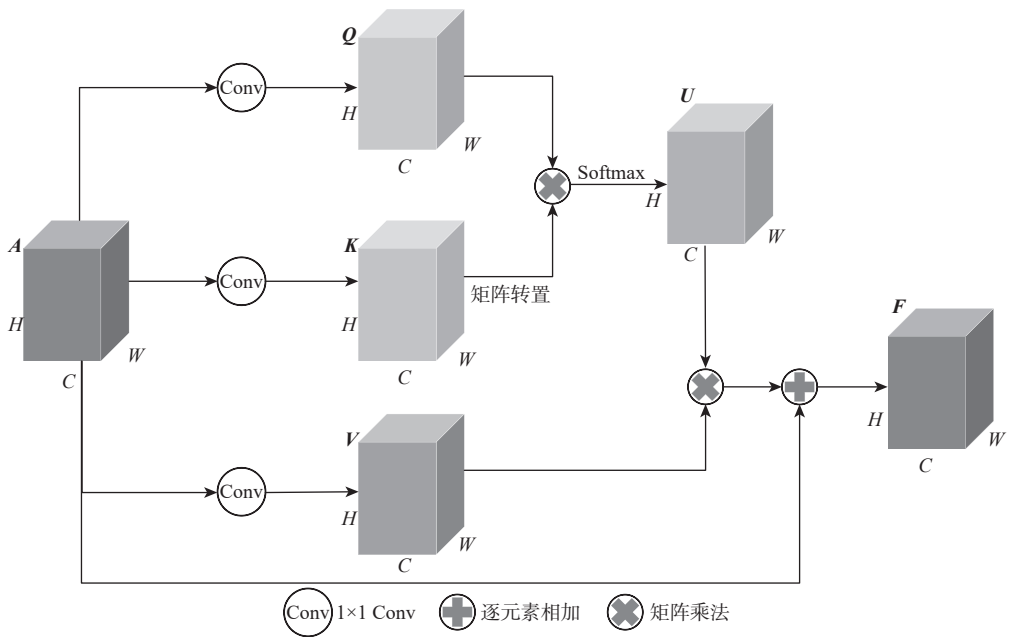


图 4　SAM 模块结构示意图

Fig. 4　Structure of SAM module

受 DANet[21] 启发, 对于给定的特征向量 $\boldsymbol{A}\in\mathbf{R}^{C\times H\times W}$, 分别使用 2 个独立的卷积操作生成 2 个新的特征向量 $\boldsymbol{Q}\in\mathbf{R}^{C\times H\times W}$ 和 $\boldsymbol{K}\in\mathbf{R}^{C\times H\times W}$。对 $\boldsymbol{Q}$ 和 $\boldsymbol{K}$ 进行矩阵相乘, 并应用 Softmax 函数得到归一化的空间注意力权重矩阵 $\boldsymbol{U}_{ij}\in\mathbf{R}^{N\times N}$:

$$\boldsymbol{U}_{ij}=\mathrm{Softmax}\left(\boldsymbol{Q}_i\boldsymbol{K}_j\right)=\frac{\exp\left(\boldsymbol{Q}_i\boldsymbol{K}_j^{\mathrm{T}}\right)}{\sum_{i=1}^{N}\sum_{j=1}^{N}\exp\left(\boldsymbol{Q}_i\boldsymbol{K}_j^{\mathrm{T}}\right)} \tag{6}$$

式中: N 为特征图大小, 即 $N=H\times W$; $\boldsymbol{Q}_i$ 和 $\boldsymbol{K}_j$ 分别为点 i 和点 j 在 $\boldsymbol{Q}$ 和 $\boldsymbol{K}$ 中的特征表示; $\boldsymbol{U}_{ij}$ 衡量第 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。对特征向量 $\boldsymbol{A}$ 应用卷积操作得到新的特征向量 $\boldsymbol{V}\in\mathbf{R}^{C\times H\times W}$, 并将其与注意力权重 $\boldsymbol{U}_{ij}$ 进行矩阵相乘操作, 最终与尺度参数 α 相乘, 并与输入特征 $\boldsymbol{A}$ 进行逐元素求和操作, 即可得到增强的输出特征 $\boldsymbol{F}$。

$$\boldsymbol{F}=\alpha\sum_{j=1}^{N}\left(\boldsymbol{U}_{ij}\boldsymbol{V}_j\right)+\boldsymbol{A}_i \tag{7}$$

式中: α 为初始值为 0 的可学习的尺度参数。

2.3　CAM 模块

目前, 大多数方法[22-23] 在提取到点云的局部邻域特征 $\boldsymbol{f}_i$ 时, 通常选择局部邻域特征中的最大值特

征来代表该局部邻域特征，定义为$\boldsymbol{F}_i=\max(\boldsymbol{f}_i)$。但是，简单地对局部邻域进行最大值池化操作，难免会丢失局部邻域内的其他特征。因此，在聚合特征时，有必要选择具有鉴别性和有效性的特征。最近的工作[24-25]在二维图像分割领域取得了不错的效果，受此启发，本文在对局部邻域特征的聚合过程中设计了一个针对三维点云数据的CAM模块，通过给不同通道赋予不同的权值以达到强调有用通道同时抑制无用通道的目的，从而提高网络训练的效果。CAM模块结构如图5所示。

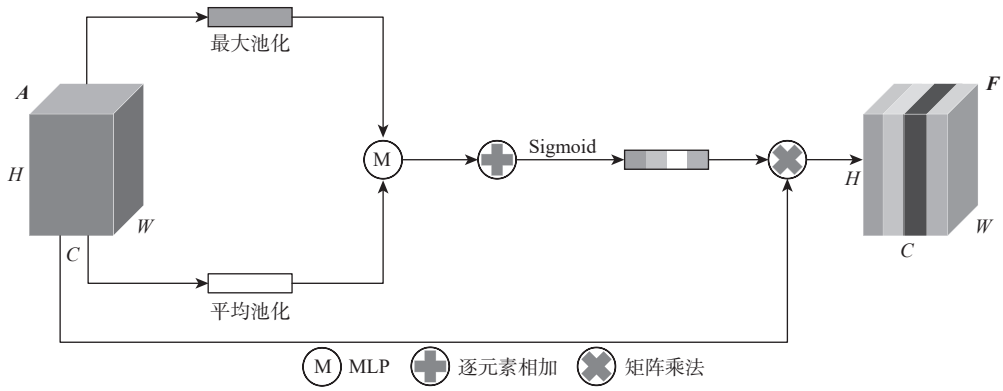


图5 CAM模块结构示意图

Fig. 5 Structure of CAM

首先，对输入特征向量$\boldsymbol{A}$在空间维度分别进行最大池化和平均池化，使其只保留通道的维度，其他维度为1，得到一个一维的矢量。然后，分别经过一个共享权重的MLP，将两者的输出做逐元素相加，再经过Sigmoid激活函数，得到通道注意力权重。最后，将该权重和输入特征向量$\boldsymbol{A}$做逐元素乘法，得到经过通道注意力机制后的特征向量$\boldsymbol{F}$：

$$\boldsymbol{F}=\sigma\left(\mathrm{MLP}(\mathrm{MaxPool}(\boldsymbol{A}))+\mathrm{MLP}(\mathrm{AvgPool}(\boldsymbol{A}))\right)\boldsymbol{A} \tag{8}$$

式中：σ表示Sigmoid激活函数。

3 实 验

3.1 评估指标及参数设置

为评估本文所构建的融合双注意力机制和DGCNN架构的有效性，选用流行的大规模室内场景数据集S3DIS[26]和大规模室外场景数据集SemanticKITTI[27]进行实验。评估指标采用总体准确率(overall accuracy, OA)和平均交并比(mean intersection-over-union, mIoU)，分别表示为

$$A=\frac{T_{\mathrm{P}}+T_{\mathrm{N}}}{T_{\mathrm{P}}+T_{\mathrm{N}}+F_{\mathrm{P}}+F_{\mathrm{N}}} \tag{9}$$

$$M=\frac{1}{c+1}\sum_{i=0}^{c}\frac{T_{\mathrm{P}}}{F_{\mathrm{N}}+F_{\mathrm{P}}+T_{\mathrm{P}}} \tag{10}$$

式中：T_{P}表示被模型预测为正类的正样本；T_{N}表示被模型预测为负类的负样本；F_{P}表示被模型预测为正类的负样本；F_{N}表示被模型预测为负类的正样本；c为数据集中的类别数。

本文的实验环境配置为：硬件为Intel(R) Core(TM) i9-10900K CPU@3.70GHz和NVIDIA RTX3090 GPU(24 GB)，软件为Linux Ubuntu 18.04和深度学习框架PyTorch1.11.0。实验采用了动量因子为0.99的基于动量的随机梯度下降(stochastic gradient descent, SGD)优化算法，使用Adam算法更新SGD的步长。

3.2 室内场景语义分割

S3DIS数据集是从3个不同建筑物的6个区域所包含的271个房间中采样生成的RGB-D点云数据。将房间内的点云数据划分为1 m×1 m的块，再对每个块预测其中每个点的语义标签。每个点由一个9维向量表示，分别为：$\boldsymbol{X}$、$\boldsymbol{Y}$、$\boldsymbol{Z}$、$\boldsymbol{R}$、$\boldsymbol{G}$、$\boldsymbol{B}$、$\boldsymbol{X}'$、$\boldsymbol{Y}'$、$\boldsymbol{Z}'$。其中，$\boldsymbol{X}$、$\boldsymbol{Y}$、$\boldsymbol{Z}$表示位置坐标，$\boldsymbol{R}$、$\boldsymbol{G}$、$\boldsymbol{B}$表示颜色信息，$\boldsymbol{X}'$、$\boldsymbol{Y}'$、$\boldsymbol{Z}'$为标准化后的每个块相对于所在房间的位置坐标(值为0～1)。

为验证本文算法的有效性，与经典算法在S3DIS数据集上进行了分割对比实验，结果如表1所示。可以看出，本文算法获得了66.0%的mIoU和92.8%的OA，在OA方面表现最优，虽然mIoU略逊于KPConv的可变形核，但在3个类别(ceiling、wall、table)上实现了最佳性能。由于本文算法使用的是KPConv的刚性核而非可变形核，避免了学习局部偏移量而修改核点位置的操作，这也导致了KPConv的可变形核对于大小各异的物体(如chair、bookcase)的分割效果优于本文算法。与BAAF-Net相比，本文算法的mIoU和OA这2个分割精度评价指标均高于该算法，主要由于BAAF-Net在提取局部邻域特征时，并未考虑邻域点间的深层次语义信息。与常规的6折交叉验证方法不

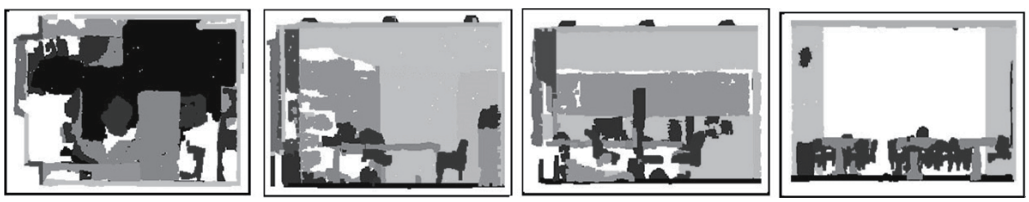
表 1　不同算法在 S3DIS 数据集上的分割精度对比（Area 5 作为测试）

Table 1　Comparison of segmentation accuracy of different algorithms on S3DIS dataset (Area 5 as a test)　%

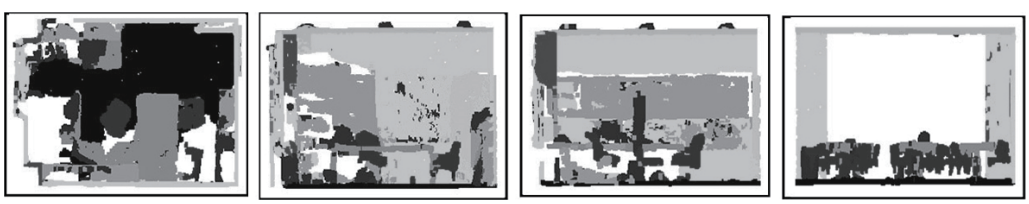
算法	OA	mIoU	IoU												
			ceiling	floor	wall	beam	column	window	door	table	chair	sofa	bookcase	board	clutter
PointNet[6]	79.3	41.1	88.8	97.3	69.8	0.1	3.9	46.3	10.8	59.0	52.6	5.9	40.3	26.4	33.2
TangentConv[8]	82.5	52.6	90.5	97.7	74.0	0	20.7	39.0	31.3	77.5	69.4	57.3	38.5	48.8	39.8
PointCNN[28]	85.9	57.3	92.3	98.2	79.4	0	17.6	22.8	62.1	74.4	80.6	31.7	66.7	62.1	56.7
SPG[29]	86.4	58.0	89.4	96.9	78.1	0	42.8	48.9	61.6	84.7	75.4	69.8	52.6	2.1	52.2
PCCN[30]		58.3	92.3	96.2	75.9	0.3	6.0	69.5	63.5	66.9	65.6	47.3	68.9	59.1	46.2
PointWeb[18]	87.0	60.3	92.0	98.5	79.4	0	21.1	59.7	34.8	76.3	88.3	46.9	69.3	64.9	52.5
HPEIN[31]	87.2	61.9	91.5	98.2	81.4	0	23.3	65.3	40.0	75.5	87.7	58.5	67.8	65.6	49.4
RandLA-Net[32]	87.2	62.4	91.1	95.6	80.2	0	24.7	62.3	47.7	76.2	83.7	60.2	71.1	65.7	53.8
GACNet[19]	87.8	62.8	92.3	98.3	81.9	0	20.3	59.1	40.8	78.5	85.8	61.7	70.7	74.7	52.8
BAAF-Net[33]	88.9	65.4	92.9	97.9	82.3	0	23.1	65.5	64.9	78.5	87.5	61.4	70.7	68.7	57.2
KPConv[20]		67. 1	92.8	97.3	82.4	0	23.9	58.0	69.0	81.5	91.0	75.4	75.3	66.7	58.9
本文	92.8	66.0	92.9	98.3	82.8	0	21.2	56.6	68.5	91.3	81.3	74.4	64.2	68.9	57.8

同, 本文将 S3DIS 数据集的 Area 5 区域单独作为测试区域, 而将 Area 1～Area 4 和 Area 6 区域用于训练。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, 对网络模型的要求更高。通过引入 SAM 模块, 本文算法提取到了局部邻域内任意点对之间更深层次的特征, 同时引入 CAM 模块, 在聚合局部邻域特征时能够根据不同通道的重要程度产生更具代表性的特征, 从而提高点云语义分割的精度。

此外, 为定性分析语义分割结果, 对 Area 5 的测试结果进行了可视化, 如图 6 所示。可以看出, 本文算法可以高效、准确地提取到点云特征, 在某些场景下得到的分割结果已经很接近真实标签。



(a) 真实标签



(b) 分割结果

图 6　S3DIS 数据集分割结果的可视化

Fig. 6　Visualization of segmentation results on S3DIS dataset

3.3　室外场景语义分割

SemanticKITTI 数据集包含了德国卡尔斯鲁厄附近的市内交通、居民区、高速公路和乡村道路场景, 是目前最大的汽车激光雷达点云数据集。该数据集由 22 个点云序列组成, 00～10 序列作为训练集(08 序列用作验证集并生成可视化结果), 共包含 23 201 帧三维点云场景; 11～21 序列作为测试集, 共包含 20 351 帧三维点云场景。训练集中每点都有自己的标签, 人为进行了语义标注, 共有 19 种类别标签。表 2 为本文算法在 SemanticKITTI 数据集上与其他经典算法的场景语义分割定量结果对比。可以看出, 与目前主流算法相比, 本文算法的 mIoU 为 59.4%, 仅次于 BAAF-Net, 并且在其中 6 个类别上也取得了最好的分割结果。而 BAAF-Net 的输入与目前主流算法不同, 不仅包含点云的位置信息, 还包含一些额外的特征(如法线、颜色等), 因此, BAAF-Net 对一些小物体 (如 other-vehicle 和 motorcyclist) 进行分割时, 能够利用这些额外的特

表2 不同算法在SemanticKITTTI数据集上的分割精度对比

Table 2 Comparison of segmentation accuracy of different algorithms on SemanticKITTI dataset %

算法	mIoU	IoU								
		road	sidewalk	parking	other-ground	building	car	truck	bicycle	motorcycle
PointNet[6]	14.6	61.6	35.7	15.8	1.4	41.4	46.3	0.1	1.3	0.3
SPG[29]	17.4	45.0	28.5	1.6	0.6	64.3	49.3	0.1	0.2	0.2
PointNet++[16]	20.1	72.0	41.8	18.7	5.6	62.3	53.7	0.9	1.9	0.2
TangentConv[8]	40.9	83.9	63.9	33.4	15.4	83.4	90.8	15.2	2.7	16.5
SpSequenceNet[34]	43.1	90.1	73.9	57.6	27.1	91.2	88.5	29.2	24.0	0
PointASNL[35]	46.8	87.4	74.3	24.3	1.8	83.1	87.9	39.0	0	25.1
HPGCNN[36]	50.5	89.5	73.6	58.8	34.6	91.2	93.1	21.0	6.5	17.6
RangeNet++[11]	52.2	91.8	75.2	65.0	27.8	87.4	91.4	25.7	25.7	34.4
RandLA-Net[32]	53.9	90.7	73.7	60.3	20.4	86.9	94.2	40.1	26.0	25.8
PolarNet[37]	54.3	90.8	74.4	61.7	21.7	90.0	93.8	22.9	40.3	30.1
3D-MiniNet[38]	55.8	91.6	74.5	64.2	25.4	89.4	90.5	28.5	42.3	42.1
SAFFGCNN[39]	56.6	89.9	73.9	63.5	35.1	91.5	95	38.3	33.2	35.1
KPConv[20]	58.8	88.8	72.7	61.3	31.6	90.5	96.0	33.4	30.2	42.5
BAAF-Net[33]	59.9	90.9	74.4	62.2	23.6	89.8	95.4	48.7	31.8	35.5
本文	59.4	92.0	77.4	70.1	35.2	89.7	94.4	38.4	43.7	40.3

算法	IoU									
	other-vehicle	vegetation	trunk	terrain	person	bicyclist	motorcyclist	fense	pole	traffic-sign
PointNet[6]	0.8	31.0	4.6	17.6	0.2	0.2	0	12.9	2.4	3.7
SPG[29]	0.8	48.9	27.2	24.6	0.3	2.7	0.1	20.8	15.9	0.8
PointNet++[16]	0.2	46.5	13.8	30.0	0.9	1.0	0	16.9	6.0	8.9
TangentConv[8]	12.1	79.5	49.3	58.1	23.0	28.4	8.1	49.0	35.8	28.5
SpSequenceNet[34]	22.7	84.0	66.0	65.7	6.3	0	0	67.7	50.8	48.7
PointASNL[35]	29.2	84.1	52.2	70.6	34.2	57.6	0	43.9	57.8	36.9
HPGCNN[36]	23.3	84.4	65.9	70.0	32.1	30.0	14.7	65.5	45.5	41.5
RangeNet++[11]	23.0	80.5	55.1	64.6	38.3	38.8	4.8	58.6	47.9	55.9
RandLA-Net[32]	38.9	81.4	61.3	66.8	49.2	48.2	7.2	56.3	49.2	47.7
PolarNet[37]	28.5	84.0	65.5	67.8	43.2	40.2	5.6	61.3	51.8	57.5
3D-MiniNet[38]	29.4	82.8	60.8	66.7	47.8	44.1	14.5	60.8	48.0	56.6
SAFFGCNN[39]	28.7	84.4	67.1	69.5	45.3	43.5	7.3	66.1	54.3	53.7
KPConv[20]	31.6	84.8	69.2	69.1	61.5	61.6	11.8	64.2	56.4	48.4
BAAF-Net[33]	46.7	82.7	63.4	67.9	49.5	55.7	53.0	60.8	53.7	52.0
本文	30.3	84.3	64.9	70.0	60.1	47.4	7.6	66.9	53.1	62.4

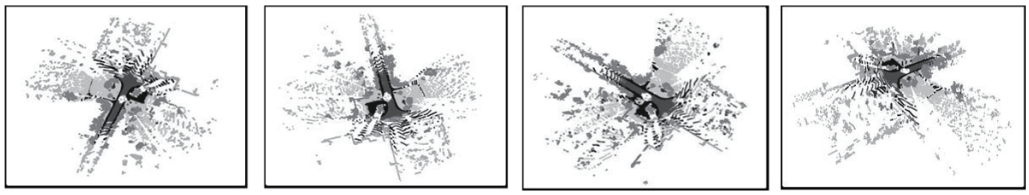
征,从而取得较好的结果。本文在提取局部特征时使用了KPConv的核点方法,但由于加入了SAM模块与CAM模块,得到的大部分场景mIoU均优于KPConv算法。图7为本文算法在SemanticKITTI数据集上的语义分割可视化结果。可以看出,由于本文算法能够提取到更深层次的局部邻域信息,即使在处理大规模复杂室外点云数据时,仍然能够产生较好的分割效果。

3.4 消融实验

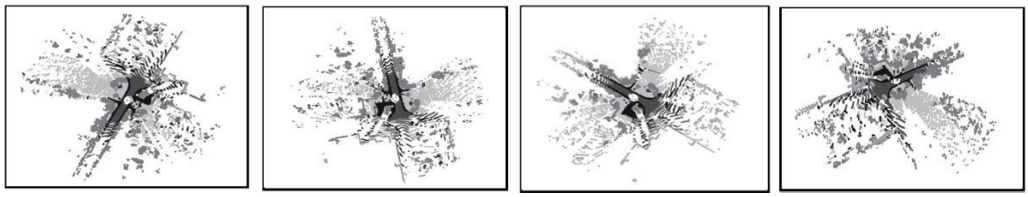
3.4.1 双注意力机制模块

为进一步说明本文引入的2种注意力模块的有效性,在S3DIS数据集上分别对这2个模块进行消融实验,结果如表3所示。其中,网络模型Net Model-3为完整的融合双注意力机制和动态图卷积神经网络,而网络模型Net Model-1是在所构建网络的基础上移除了SAM模块,网络模型Net Model-2则是移除了CAM模块。

从表3中结果可以看出,网络模型Net Model-3在S3DIS数据集上的mIoU比网络模型Net Model-1高出3.2%,验证了空间注意力的有效性。原因在于:SAM模块能够在局部邻域内学习到任意2个特征相似的点的深层次语义信息,从而使每个点的特征都得到一定程度的增强。而网络模型Net Model-3在S3DIS数据集上的mIoU比网络模型Net Model-2



(a) 真实标签



(b) 分割结果

图 7　SemanticKITTI 数据集分割结果的可视化

Fig. 7　Visualization of segmentation results on SemanticKITTI dataset

表 3　不同模块在 S3DIS 数据集上的消融实验

Table 3　Ablation experiments of different modules on S3DIS dataset

网络模型	模块名	mIoU/%
Net Model-1	CAM	62.8
Net Model-2	SAM	65.2
Net Model-3	SAM+CAM	66.0

只高出 0.8%, 因为在加入 CAM 模块后, 虽然会增强有用通道的权重, 但还是会通过最大池化操作来聚合局部邻域特征, 所以, 对于点云的语义分割效果提升相比于 SAM 模块并不是很大。

3.4.2　改进的动态图卷积模块

为进一步说明本文改进的动态图卷积模块的有效性, 在 S3DIS 数据集上分别对 EdgeConv++和 KPConv 这 2 个模块进行消融实验, 结果如表 4 所示。其中, 网络模型 Net Model-6 为完整的融合双注意力机制和 DGCNN, 而网络模型 Net Model-4 是在所构建网络的基础上移除了刚性核的核点卷积(KPConv rigid)模块, 网络模型 Net Model-5 则是移除了动态图卷积模块, 网络模型 Net Model-7 为使用 KPConv 原网络模型的刚性核。

从表 4 中结果可以看出, 虽然在单独使用动态图卷积或刚性核的核点卷积进行边缘特征提取时, 本文的结果不如 KPConv 原网络模型的刚性核, 但将二者相结合(即改进的动态图卷积)得到的结果优于 KPConv 原网络模型的刚性核, 这也证明了本文所提出的改进的动态图卷积模块的有效性。

表 4　改进的动态图卷积模块有效性验证

Table 4　Effectiveness verification of improved dynamic graph convolution module

网络模型	模块名	mIoU/%
Net Model-4	EdgeConv++	64.2
Net Model-5	KPConv	64.7
Net Model-6	KPConv+EdgeConv++	66.0
Net Model-7	KPConv rigid	65.4

3.4.3　核点卷积模块中的核心点数量

核心点的数量会影响网络提取到的局部几何特征的优劣, 较少的核心点数量 d 使网络无法充分学习到有效的局部特征, 导致分割精度较差; 而 d 的数量过大反而会造成更多冗余的局部特征, 增加网络计算量。从表 5 可以看出, 当核心点数量 d<15 时, 随着 d 的增大, 分割结果也在逐步提升; 而当 d>15 时, 由于核心点数量过多, 难免会学习到冗余的特征, 反而影响了分割的结果。因此, 本文在实验中的参数设置为 d=15。

表 5　核心点数量对分割结果的影响

Table 5　Influence of number of kernel points on segmentation results

d	mIoU/%
11	64.8
13	65.3
15	66.0
17	65.7

4　结　论

本文针对三维点云语义分割算法在提取局部特征时忽略邻域点间深层次语义信息, 聚合局部邻域特征时忽略其他邻域特征中有用信息的问题, 提出融合双注意力机制和 DGCNN 的三维点云语义分割算法。

1) 采用动态图采样边缘特征, 以及使用通道、空间注意力有效提升语义分割的精度, 在 S3DIS 和 SemanticKITTI 数据集上, 语义分割精度分别达到了 66.0% 和 59.4%。

2）由于网络采用KNN算法构建局部邻域图，会导致网络鲁棒性较差，从而影响点云的分割结果。因此，如何处理稀疏数据集中具有相似几何结构特征的物体仍是需要研究的重点。

参考文献（References）

[1] ZHU Y K, MOTTAGHI R, KOLVE E, et al. Target-driven visual navigation in indoor scenes using deep reinforcement learning[C]// Proceedings of the IEEE International Conference on Robotics and Automation. Piscataway: IEEE Press, 2017: 3357-3364.

[2] ZHANG K G, XIONG C H, ZHANG W, et al. Environmental features recognition for lower limb prostheses toward predictive walking[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2019, 27(3): 465-476.

[3] QI C R, LIU W, WU C X, et al. Frustum PointNets for 3D object detection from RGB-D data[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 918-927.

[4] ZHANG G C, YU J, WANG Z H, et al. Visual 3D reconstruction system based on RGBD camera[C]//Proceedings of the IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference. Piscataway: IEEE Press, 2020: 908-911.

[5] 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 (in Chinese).

[6] CHARLES R Q, HAO S, MO K C, 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. Piscataway: IEEE Press, 2017: 77-85.

[7] 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.

[8] TATARCHENKO M, PARK J, KOLTUN V, et al. Tangent convolutions for dense prediction in 3D[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 3887-3896.

[9] 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]//Proceedings of the IEEE International Conference on Robotics and Automation. Piscataway: IEEE Press, 2018: 1887-1893.

[10] WU B C, ZHOU X Y, ZHAO S C, et al. SqueezeSegV2: Improved model structure and unsupervised domain adaptation for road-object segmentation from a LiDAR point cloud[C]//Proceedings of the International Conference on Robotics and Automation. Piscataway: IEEE Press, 2019: 4376-4382.

[11] MILIOTO A, VIZZO I, BEHLEY J, et al. RangeNet++: Fast and accurate LiDAR semantic segmentation[C]//Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway: IEEE Press, 2019: 4213-4220.

[12] MATURANA D, SCHERER S. VoxNet: A 3D convolutional neural network for real-time object recognition[C]//Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway: IEEE Press, 2015: 922-928.

[13] TCHAPMI L, CHOY C, ARMENI I, et al. SEGCloud: Semantic segmentation of 3D point clouds[C]//Proceedings of the International Conference on 3D Vision. Piscataway: IEEE Press, 2017: 537-547.

[14] RIEGLER G, ULUSOY A O, GEIGER A. OctNet: Learning deep 3D representations at high resolutions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 6620-6629.

[15] ZENG W, GEVERS T. 3DContextNet: k-d tree guided hierarchical learning of point clouds using local and global contextual cues[C]// Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2018: 314-330.

[16] 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. Cambridge: MIT Press, 2017: 5099-5108.

[17] 党吉圣, 杨军. 多特征融合的三维模型识别与分割[J]. 西安电子科技大学学报, 2020, 47(4): 149-157.
DANG J S, YANG J. 3D model recognition and segmentation based on multi-feature fusion[J]. Journal of Xidian University, 2020, 47(4): 149-157(in Chinese).

[18] ZHAO H S, 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. Piscataway: IEEE Press, 2019: 5560-5568.

[19] WANG L, HUANG Y C, HOU Y L, et al. Graph attention convolution for point cloud semantic segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 10288-10297.

[20] 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. Piscataway: IEEE Press, 2019: 6410-6419.

[21] FU J, LIU J, TIAN H J, et al. Dual attention network for scene segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 3141-3149.

[22] LE T, DUAN Y. PointGrid: A deep network for 3D shape understanding[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 9204-9214.

[23] ZHANG Z Y, HUA B S, YEUNG S K. ShellNet: Efficient point cloud convolutional neural networks using concentric shells statistics [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 1607-1616.

[24] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.

[25] WOO S, PARK J C, LEE J Y, et al. CBAM: Convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2018: 3-19.

[26] 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. Piscataway: IEEE Press, 2016: 1534-1543.

[27] BEHLEY J, GARBADE M, MILIOTO A, et al. SemanticKITTI: A dataset for semantic scene understanding of LiDAR sequences[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 9296-9306.

[28] LI Y Y, BU R, SUN M C, et al. PointCNN: Convolution on X-transformed points[C]//Proceedings of the Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2018, 31: 828-838.

[29] 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. Piscataway: IEEE Press, 2018: 4558-4567.

[30] WANG S L, SUO S, MA W C, et al. Deep parametric continuous convolutional neural networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 2589-2597.

[31] JIANG L, ZHAO H S, LIU S, et al. Hierarchical point-edge interaction network for point cloud semantic segmentation[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 10432-10440.

[32] HU Q Y, YANG B, XIE L H, et al. RandLA-Net: Efficient semantic segmentation of large-scale point clouds[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 11105-11114.

[33] QIU S, ANWAR S, BARNES N. Semantic segmentation for real point cloud scenes via bilateral augmentation and adaptive fusion[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 1757-1767.

[34] SHI H Y, LIN G S, WANG H, et al. SpSequenceNet: Semantic segmentation network on 4D point clouds[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 4573-4582.

[35] YAN X, ZHENG C D, LI Z, et al. PointASNL: Robust point clouds processing using nonlocal neural networks with adaptive sampling[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 5588-5597.

[36] DANG J S, YANG J. HPGCNN: Hierarchical parallel group convolutional neural networks for point clouds processing[C]//Proceedings of the Asian Conference on Computer Vision. Berlin: Springer, 2020: 20-37.

[37] ZHANG Y, ZHOU Z X, DAVID P, et al. PolarNet: An improved grid representation for online LiDAR point clouds semantic segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 9598-9607.

[38] 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 Robotics and Automation Letters, 2020, 5(4): 5432-5439.

[39] 杨军, 李博赞. 基于自注意力特征融合组卷积神经网络的三维点云语义分割[J]. 光学 精密工程, 2022, 30(7): 840-853.
YANG J, LI B Z. Semantic segmentation of 3D point cloud based on self-attention feature fusion group convolutional neural network[J]. Optics and Precision Engineering, 2022, 30(7): 840-853(in Chinese).

Semantic segmentation of point clouds by fusing dual attention mechanism and dynamic graph convolution

YANG Jun[1, 2, *], ZHANG Chen[1]

(1. School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China;
2. Faculty of Geomatics, Lanzhou Jiaotong University, Lanzhou 730070, China)

Abstract: The existing semantic segmentation methods of 3D point clouds based on deep learning usually ignore the profound semantic information between neighboring points when extracting local features and fail to consider the useful information in other neighboring features when aggregating local neighboring features. To solve these problems, a semantic segmentation algorithm of 3D point clouds fusing dual attention mechanism and dynamic graph convolution neural network (DGCNN) was proposed. Firstly, edge features were constructed by dynamic graph convolution operation, and the relative distance between the center point and the neighboring points was input to the kernel point convolution operation to obtain enhanced edge features, further strengthening the relationship between the center point and the neighboring points. Secondly, the spatial attention module was introduced to establish the dependence between neighboring points, and the similar feature points were intercorrelated, so as to extract profound context information in the local neighborhood and enrich the geometric features of neighboring points. Finally, the channel attention module was introduced when local neighboring features were aggregated. By giving different weights to different channels, the purpose of enhancing useful channels and suppressing useless channels was achieved, so as to improve the accuracy of semantic segmentation. The experimental results on the S3DIS dataset and SemanticKITTI dataset show that the semantic segmentation accuracy of this algorithm has reached 66.0% and 59.4%, respectively. Compared with other classical network models, this algorithm has achieved a better point cloud segmentation effect.

Keywords: 3D point cloud; semantic segmentation; deep learning; attention mechanism; dynamic graph convolution

Received: 2022-09-14; **Accepted**: 2023-01-02; **Published Online**: 2023-01-10 13:51

URL: link.cnki.net/urlid/11.2625.V.20230109.1713.007

Foundation items: National Natural Science Foundation of China (42261067,61862039); Talent Innovation and Entrepreneurship Project of Lanzhou City (2020-RC-22); Tianyou Innovation Team of Lanzhou Jiaotong University (TY202002)

* **Corresponding author**. E-mail: yangj@mail.lzjtu.cn