文章编号　1004-924X(2022)07-0840-14

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

杨　军[1,2]，李博赞[2*]

(1. 兰州交通大学 测绘与地理信息学院，甘肃 兰州 730070；
2. 兰州交通大学 自动化与电气工程学院，甘肃 兰州 730070)

摘要：针对现有算法忽略点云数据全局单点特征和局部几何特征的深层关系，导致捕获的局部几何信息缺乏鉴别性且难以有效识别复杂形状的问题，提出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算法。首先，设计轻量化网络框架的代理点图卷积提取点云局部几何特征，并加入组卷积操作减少计算量和复杂度，以较少的冗余信息增强特征的丰富性；其次，通过Transformer模块进行不同分支间特征信息的交流，使全局特征和局部几何特征相互补偿，增强特征的完备性；然后，将点云底层语义特征与原始点云融合以扩大局部邻域感受野，获得高级上下文语义信息；最后，将特征输入到分割模块完成细粒度语义分割。实验结果表明，该算法在S3DIS数据集和SemanticKITTI数据集上的分割精度分别达到79.3%和56.6%，能够提取三维点云的关键特征信息，网络参数量较少且具有较高的语义分割鲁棒性。

关　键　词：三维点云；语义分割；图卷积；组卷积

中图分类号：TP391　　**文献标识码**：A　　**doi**：10.37188/OPE.20223007.0840

Semantic segmentation of 3D point cloud based on self-attention feature fusion group convolutional neural network

YANG Jun[1,2], LI Bozan[2*]

(1. *Faculty of Geomatics, Lanzhou Jiaotong University, Lanzhou* 730070, *China*;
2. *School of Automation and Electrical Engineering, Lanzhou Jiaotong University, Lanzhou* 730070, *China*)
* *Corresponding author, E-mail*: 763311755@qq.com

Abstract: The existing algorithms ignore the profound relationship between global single point features and local geometric features. This results in the lack of discriminative captured local geometric information and increases the difficulty of effectively identifying complex shape categories. This paper proposes a semantic segmentation algorithm for three-dimensional point clouds based on a self-attention feature fusion group convolutional neural network. First, the proxy point graph convolution of lightweight network is designed to extract the local geometric features of the point cloud. Then, the group convolution operation is

收稿日期：2021-10-08；**修订日期**：2021-11-19.

基金项目：国家自然科学基金资助项目(No. 61862039)；甘肃省科技计划资助项目(No. 20JR5RA429)；2021年度中央引导地方科技发展资金资助项目(No. 2021-51)；兰州市人才创新创业项目(No. 2020-RC-22)；兰州交通大学天佑创新团队(No. TY202002)

added to reduce the amount of calculation and complexity and enhance the richness of features with less redundant information. Second, the feature information exchange between different branches is carried out through the Transformer module to ensure mutual compensation between the global and local geometric features and to enhance the completeness of features. Then, the underlying semantic features of the point cloud are fused with the original point cloud to expand the local neighborhood perception field and obtain high-level context semantic information. Finally, the features are input into the segmentation module to complete fine-grained semantic segmentation. The experimental results show that the segmentation accuracy reaches 79.3% and 56.6% in the S3DIS and SemanticKITTI datasets, respectively. This algorithm can extract the key feature information from a 3D point cloud using fewer network parameters and exhibits high robustness of semantic segmentation.

Key words: three-dimensional point cloud; semantic segmentation; graph convolution; group convolution

1　引　言

近年来,3D扫描技术的发展促进了智能驾驶[1-2]和增强现实[3]等新技术的应用,对场景的准确理解已成为人工智能领域的主要研究方向。为结合三维模型表面细节信息从而提高分割精度,研究人员利用二维图像分割算法处理规则数据的优势,将一组点云投影为二维图像便于学习点云特征,并将像素级语义标签反投影到点云获得分割结果[4]。但是,多视图方法会不可避免地丢失某些具有鉴别力的几何信息,并且投影视角的选择也需要丰富的先验知识。直接处理点云数据的方法能够利用点云固有信息且不增加额外操作,可以充分获取点云所有信息。然而,原始点云具有不规则、稀疏和无序结构等特点,需要构建局部邻域图或转化为规则结构才能直接利用。基于体素[5]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。此外,该结构通常受到存储器的严格限制,高分辨率会消耗巨大的计算和存储成本,低分辨率则容易出现严重的信息丢失问题。稀疏卷积[6]虽然能够减少内存占用,但为了获得更大的感受野,在低分辨率操作下多个类别会合并到一个网格从而影响分割结果。基于逐点的方法[7-9]虽然便于获取局部几何信息,但只有部分几何信息对物体整体结构具有判别性,点的绝对位置信息和点对间的相对位置信息缺乏描述物体高级全局几何结构的能力,而且网络运行消耗大量时间用于构建局部点云数据,导致时间成本上升。

针对上述问题,本文提出了基于自注意力特征融合组卷积神经网络(Self-attention Feature Fusion Group Convolutional Neural Network, SAFFGCNN)的点云细粒度分析方法。引入Transformer模块将全局单点特征和局部几何特征进行融合,提高特征的丰富性。提出了一种轻量级的图卷积运算——代理点图卷积,获得深层细粒度的几何特征,能够简化边缘卷积操作降低内存消耗,对语义特征和局部几何特征进行编码,增强特征局部的上下文信息。通过多尺度策略不断扩大局部邻域感受野以学习局部几何特征,增强网络泛化能力,有利于捕获高级语义的上下文细粒度特征。此外,多尺度点云特征拼接后输入到分割模块,可以提高网络分割精度。

2　研究现状

目前,三维模型语义分割主要有基于投影、基于体素和基于点云三类方法。投影方法利用多视图表示场景物体表面信息,为提高分割效率,基于距离图像的球面投影方法被提出。体素方法将点云转化为密集体素网格表示,为了适应点云稀疏性和密度变化,用稀疏体素网格表示点云场景。点云方法直接对点云进行卷积操作,可以有效获取点云数据的本征属性,主要有基于递归神经网络、构建点云卷积核和基于图网络三类方法。

2.1　基于投影的方法

由于点云的不规则性,许多研究首先将点云投影为鸟瞰图像或距离图像,再用二维卷积操作

进行学习。Lawin等[4]首先从多个虚拟视角将点云投影到2D平面上，然后使用全连接层进行像素级语义分割，并将每张图像的分割结果反投影到点云进行融合得到点的语义标签。Milioto等[12]利用球面投影方法将点云转换为距离图像，并在图像上进行二维全卷积操作；为修正反投影后物体边缘部位的分割结果，在点云上利用高效的k近邻搜索解决遮挡问题。徐等[13]在Squeeze-Seg模型[14]结构基础上设计空间自适应卷积，它具有空间适应性和内容感知的能力，解决了标准卷积应用于LiDAR图像导致的网络性能下降的问题。

基于投影的方法的核心是将点云数据转化为规则的二维图像，利用现有成熟的二维卷积算法提取三维模型的表面细节信息。但该类方法主要存在两点缺陷：一是模型的部分表面细节信息会由于物体遮挡而消失；二是经投影后产生的图像中物体可能会出现扭曲现象，从而影响模型表面细节信息的获取。

2.2 基于体素的方法

体素化的方法通常将点云转变为密集网格，然后利用标准的3D卷积处理。黄等[5]在网络训练时将点云生成为一组占位体素网格，其标签由周围单元类别决定，然后将它输入到3D CNN进行体素分割，将推断的体素结果映射回原始点云产生逐点标签。Graham等[6]提出了子流形稀疏卷积网络，通过哈希表构建稀疏矩阵的索引关系，卷积的输出只与被占用的体素相关，内存占用和计算成本大大减少，并且能够确保卷积网络的空间稀疏性不会消失，避免出现子流行膨胀问题。Choy等[15]提出一种用于时空三维点云数据的4D稀疏卷积网络，并创建了稀疏张量自动微分的开源库。所提出的广义稀疏卷积能够有效处理高维数据，显著降低传统3D卷积核计算成本，且该卷积核对于立方体结构的物体识别能力更强。

体素表示一定程度上保留了点云的邻域结构，其数据格式能够直接运用标准3D卷积进行学习。然而，体素化不可避免地丢失了细粒度几何信息。为了解决信息丢失等问题，需要提高体素分辨率，而此操作易导致计算成本高和内存占用大等问题。虽然稀疏卷积能够处理更小的网格结构且具有良好的性能，但是依然需要进行计算效率和体素比例的权衡。

2.3 基于点云的方法

PointNet[16]和PointNet++[17]开创了基于多层感知机对点云直接进行操作的先例。蒋等[18]将编码-解码结构引入3D点云分割网络中，在解码器部分建立边分支以提供上下文信息，通过分层图设计使特征信息由粗糙到细致。党等[19]提出分层并行组卷积，可以同时捕捉点云的区分性独立单点特征和局部几何特征，以较少的冗余信息增强特征的丰富性，提高网络识别复杂类别的能力。胡等[20]提出了一种高效、轻量级的Rand-LA-Net网络，通过局部特征聚集模块扩大k近邻点搜索范围来减少信息损失，并利用随机采样降低了存储成本，提高了计算效率。Landrieu等[21]将点云通过一系列相互联系的简单形状构成超点，其属性有向图能够捕获丰富的上下文信息和几何信息，同时超点能够大大减少点云中点的数目，使网络应用于大规模点云数据集。

直接处理和分析点云的方法需要获取更精细的点云特征，才能达到细粒度点云分割任务的要求，但现有方法缺乏分辨相似物体几何特征和局部细节结构的能力，对于具有抽象语义识别能力的高级全局结构信息缺乏考虑。此外，没有考虑全局单点特征和低级局部几何特征的联系。

3 自注意力特征融合组卷积神经网络

在自注意力特征融合组卷积神经网络中，通过学习全局特征和局部几何特征的深层隐含关系，获得具有抽象语义识别能力的高级全局单点特征，提高了网络在复杂环境下的物体分割能力。首先，通过MLP和代理点图卷积获得全局特征和局部几何特征，加入组卷积操作减少冗余特征信息，获得具有鉴别性的特征。其次，利用Transformer特征融合模块增强不同特征间的联系，获得细粒度上下文信息。最后，通过多尺度特征融合扩大感受野获得全局高级单点特征。

3.1 全局-局部组卷积

本文的全局-局部组卷积由两部分组成：MLP组卷积和代理点图组卷积。

MLP组卷积在减少计算复杂度和网络参数量的同时，特征丰富性会因为组卷积产生的分组操作而降低。为了加强组间信息交流，将不同分组特征进行融合，以保证MLP组卷积层输出特征的有效性。

组卷积操作先将每层的MLP分为N组，表示为$M^l=\{M_1^l,M_2^l,\cdots,M_N^l\}$，其中$l$为第$l$个卷积层。再对输入特征进行MLP组卷积提取各个分组特征。第一组特征是第一组原始特征经过组卷积后的新特征，其余组特征为前一组新特征和自身经过组卷积后的新特征融合得到的结果。将所有分组的全局特征进行拼接操作得到MLP组卷积模块在该层的输出。MLP组卷积第l层的输出结果如下：

$$f_M^l=\sum_{n=1}^{N}x_n^l, \tag{1}$$

$$x_n^l=M_n^l x_n^l+M_{n-1}^l x_{n-1}^l, \tag{2}$$

式中：x_n^l为第l层各组的全局单点特征，f_M^l为MLP组卷积在第l层输出的全局单点特征。

MLP组卷积虽然能够捕获独立的单点特征，但对几何信息的获取存在局限性。局部几何信息包含点的位置信息以及点的相对位置，对于物体细粒度分割起到至关重要的作用。

本文以边缘卷积为出发点设计代理点图组卷积，将特征空间上的k近邻搜索转变为在原始点云空间中的k近邻搜索。原始点云空间中点的位置是固定不变的，k近邻图能更好地表征物体的空间结构信息，获得更具鉴别性的局部几何特征信息。同时，由于原始点云位置是固定的，在特征空间上构造k近邻图无需重新计算，解决了计算代价大的问题。k近邻图的邻域点在空间内接近，特征的丰富性差异小，为了保留关键几何特征信息，将k近邻点特征进行平均操作赋值到代理点，使用代理点和中心点进行几何信息学习。通过对全部卷积层共享空间邻接矩阵以减少内存消耗和计算开销，能够使特征映射的内存消耗从$O(n\times h\times d)$减少到$O(n\times d)$，大大提高了图卷积提取几何特征的效率。边缘卷积与代理点图组卷积的网络结构如图1所示。

为了在原始点云空间进行k近邻搜索，首先要计算图的空间邻接矩阵$\boldsymbol{G}\in\mathbf{R}^{N\times N}$，其元素表示一组点在图中是否相邻。为计算邻接矩阵$\boldsymbol{G}$，需

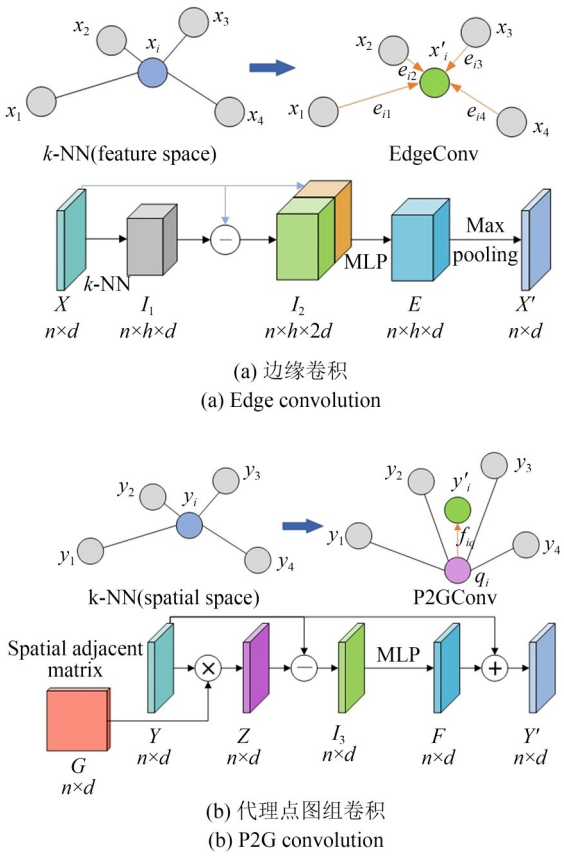


图1　边缘卷积与代理点图组卷积网络结构

Fig. 1　Network structures of edge convolution and proxy point graph group convolution

要计算点i和点j之间的欧氏距离$D_{i,j}$：

$$D_{i,j}=\left\| \boldsymbol{p}_i-\boldsymbol{p}_j \right\|^2, \tag{3}$$

式中$\boldsymbol{p}_i\in\mathbf{R}^3$和$\boldsymbol{p}_j\in\mathbf{R}^3$是两个坐标向量。将$\boldsymbol{G}$中每一行的元素进行二值化，$k$个最小的元素设为1，其余元素设为0，以此得到空间邻接矩阵$\boldsymbol{G}\in\mathbf{R}^{N\times N}$。

其次，通过矩阵乘法求得局部邻域的特征平均值，并将该特征值视为代理点特征，公式如下：

$$\boldsymbol{Z}=\frac{\boldsymbol{G}\times\boldsymbol{y}}{k}, \tag{4}$$

式中：$\boldsymbol{y}\in\mathbf{R}^{N\times d}$是由MLP组卷积获得的点云全局特征，$k$为中心点$i$的邻域点数目，$\boldsymbol{Z}$为生成的代理点特征，其中$\boldsymbol{Z}_i$为第$i$个代理点的特征。

然后，使用中心点和代理点来计算局部几何信息得到新的聚合特征，定义如下：

$$\boldsymbol{f}_i=\mathrm{ReLU}(g_\Theta(\boldsymbol{Z}_i-\boldsymbol{y}_i)), \tag{5}$$

式中：$\boldsymbol{f}_i$为生成的第i个点几何特征，$\boldsymbol{y}_i$为第i个点

的全局单点特征，ReLU 为激活函数，$g_\Theta: \mathbf{R}^d \to \mathbf{R}^d$ 是具有可学习参数 Θ 的非线性函数。最后，通过在生成的几何特征上融合输入点的全局特征来定义局部几何特征，即：

$$\boldsymbol{Y}_i = \boldsymbol{y}_i + \boldsymbol{f}_{\mathrm{G}}^l, \tag{6}$$

式中 $\boldsymbol{Y}_i$ 为第 i 个点最终的局部几何特征。

同样，将组卷积引入到代理点图卷积中，可以表示为 $G^l = \{G_1^l, G_2^l, \cdots, G_N^l\}$，其中 N 表示有 N 组，l 表示第 l 个卷积层。将不同组特征融合，以保证该卷积层最终输出的局部几何特征 $\boldsymbol{f}_{\mathrm{G}}^l$ 的丰富性。输出结果如下：

$$\boldsymbol{f}_{\mathrm{G}}^l = \sum_{n=1}^{N} \boldsymbol{f}_{\mathrm{in}}^l, \tag{7}$$

$$\boldsymbol{f}_{\mathrm{in}}^l = \boldsymbol{G}_i^l \boldsymbol{f}_{\mathrm{in}}^l + \boldsymbol{G}_{i-1}^l \boldsymbol{f}_{i(n-1)}^l, \tag{8}$$

式中：$\boldsymbol{f}_{\mathrm{in}}^l$ 为第 l 层各组的局部几何特征，$\boldsymbol{f}_{\mathrm{G}}^l$ 为代理点图组卷积在第 l 层输出的局部几何特征。

3.2 Transformer 特征融合模块

经过全局-局部组卷积模块后，全局上下文特征和局部几何特征的丰富性得到了增强，但是组卷积内部同层不同组之间缺乏信息交流，而且不同组卷积模块之间没有信息传播，缺乏具有高级语义的局部上下文信息。因此，本文通过 Transformer 的自注意力机制获得具有高级语义识别能力的特征。由于自注意力机制输入为离散标记组成的序列，各分支特征被视为集合，其中每个 $1\times1\times C$ 维特征等同于集合中的元素，并视为一个标记。分支以不同的关注方向对场景进行编码，根据特征间的自注意力系数融合其他组的特征，使更新后的每组特征包含来自其他组的特征，利用不同特征的互补性促进模块之间的信息交流，加强特征间的语义联系。全局-局部特征的 Transformer 自注意力融合操作如图 2 所示。

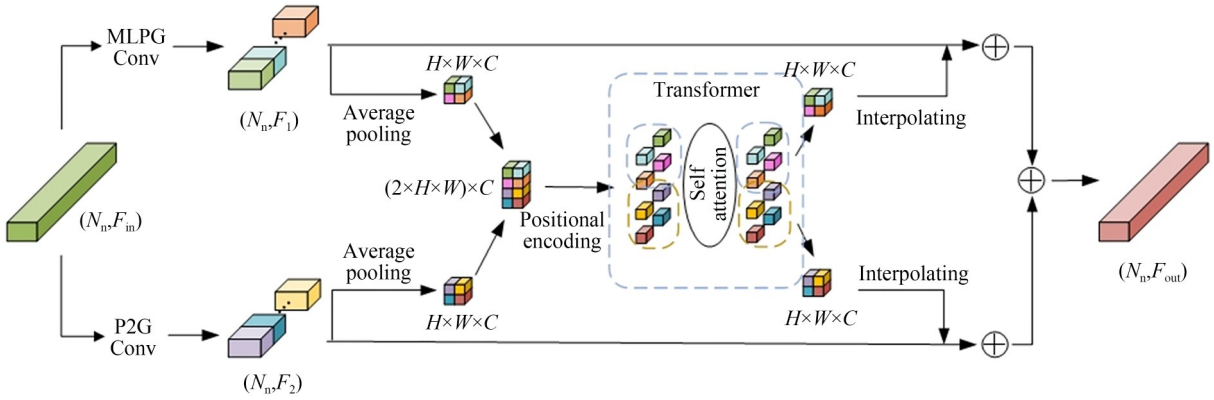


图 2　全局-局部特征的 Transformer 自注意力融合

Fig. 2　Transformer self-attention fusion of global-local features

为了减轻 Transformer 网络计算代价，将较高分辨率的分支特征做平均池化下采样处理为 $H\times W\times C$ 的三维张量，再将两者叠加形成维度为 $(2\times H\times W)\times C$ 的输入张量，并嵌入一个维度一致的可训练位置参数，使网络在训练时能够理解不同标记之间的空间位置关系。自注意力输出特征根据输入张量的位置关系重新划分为两个 $H\times W\times C$ 的特征图，并通过双线性插值上采样到原始分辨率，再与原始分支特征逐元素求和。多次实验结果表明，特征图分辨率为 $H=W=8$ 时效果最佳。

特征图上的自注意力操作类似于将 Transformer 应用于图像的工作[10-11]。设输入序列表示为 $F_{\mathrm{in}} \in \mathbf{R}^{N\times D_f}$，其中 N 是序列中的标记数，每个标记由维数为 D_f 的特征向量表示。首先，Transformer 模块使用线性投影来计算出每个标记的一组查询向量 $\boldsymbol{Q}$、关键向量 $\boldsymbol{K}$ 和值向量 $\boldsymbol{V}$，计算公式为：

$$\boldsymbol{Q} = F_{\mathrm{in}} B^Q, \boldsymbol{K} = F_{\mathrm{in}} B^K, \boldsymbol{V} = F_{\mathrm{in}} B^V, \tag{9}$$

式中：$B^Q \in \mathbf{R}^{D_f\times d_k}$，$B^K \in \mathbf{R}^{D_f\times d_k}$ 和 $B^V \in \mathbf{R}^{D_f\times d_v}$ 都是权重矩阵，目的在于将输入特征映射到不同高维空间，增强模型表达能力，更好地捕获 $\boldsymbol{Q}$，$\boldsymbol{K}$ 和 $\boldsymbol{V}$ 之间的语义级别联系。

其次，通过当前查询向量 $\boldsymbol{Q}$ 和所有关键向量

$\boldsymbol{K}$之间的点积计算自注意力权重，将所有值向量和相应权重相乘并求和，得到该特征向量标记最终的自注意力输出结果，计算公式如下：

$$\boldsymbol{A}=\mathrm{Softmax}\left(\frac{\boldsymbol{QK}^{\mathrm{T}}}{\sqrt{D_K}}\right)\boldsymbol{V}, \tag{10}$$

式中：$\sqrt{D_K}$用于在训练过程中保持梯度值稳定，防止Softmax($\boldsymbol{QK}^{\mathrm{T}}$)结果过大，导致梯度变小不利于反向传播；Softmax函数用于确保所有自注意力权重的和为1。

最后，Transformer模块使用MLP将自注意结果映射到与F_{in}同一维度，并计算输出特征，即：

$$F_{\mathrm{out}}=MLP(A)+F_{\mathrm{in}}. \tag{11}$$

输出特征F_{out}与输入特征F_{in}具有相同的维度。

3.3　自注意力特征融合组卷积神经网络

本文构建的自注意力特征融合组卷积神经网络架构如图3所示，主要由3个模块组成：MLP组卷积、代理点图组卷积和Transformer特征融合模块。点云输入到网络前进行下采样操作处理保证网络训练过程中能够收敛，选择最远点采样(Farthest Point Sampling，FPS)对场景进行均匀采样，保留点云的原始空间结构。在网络学习过程中，为了获取全局单点特征和细粒度的几何特征，通过MLP组卷积和代理点图组卷积分别提取全局特征和局部几何特征。然后，通过Transformer特征融合模块将全局单点特征和局部几何特征进行融合并增强，提高网络识别复杂形状物体的能力。为了提高分割准确率，将上一次下采样后的特征映射结果输入本次下采样后的点云中来增加不同尺度局部区域的感受野，从而获得具有高级语义的上下文细粒度特征。最后，将不同下采样的特征映射进行拼接，对它进行全局平均池化操作加强特征映射和类别之间的关联，使获得的形状级别的全局特征映射更加接近语义类别信息。

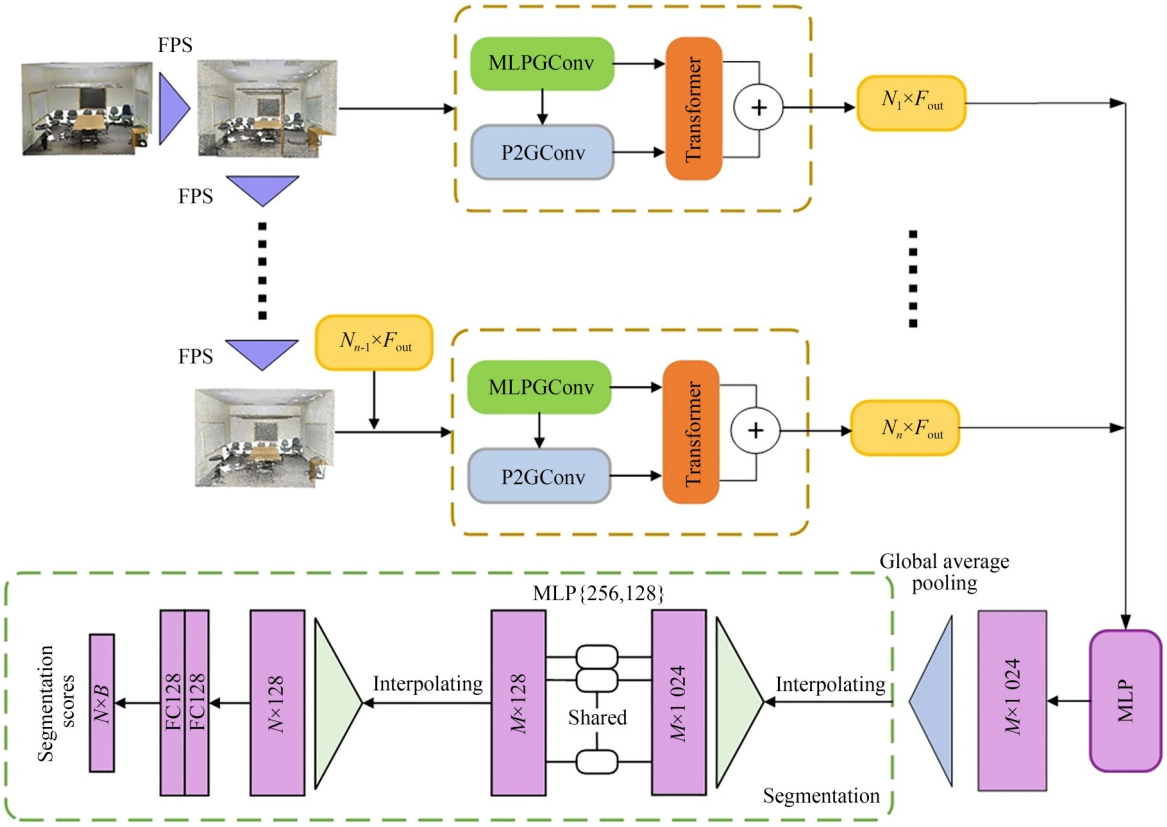


图3　自注意力特征融合组卷积神经网络

Fig. 3　Self-attention feature fusion group convolutional neural network

为了获取每个点的点级别标签，分割模块需将全局特征映射从形状级别传播到点级别。通过第一次插值后的特征与对应点的原始特征相结合获得M个点的点级特征，将点级特征输入到多个MLP层和SeLU层获得降维后点级特征，再通过第二次插值将M个点的点级特征传播到原始点云，得到原始点云空间中所有点的新特征。使用两个叠加的全连接层对点云特征进行分类，输出$N\times B$特征矩阵，其中N为原始空间内所有的点，B为每个点对应于每个类别的分数。每个点选取得分最高的类别作为其语义标签，由此获得点云场景的语义分割结果。

4　实验结果与分析

为了测试SAFFGCNN对点云的细粒度形状分析的有效性，在两个大规模语义分割数据集S3DIS[22]和SemantiKITTI[23]上评估了网络模型性能。实验中，在32 GB内存、Intel i7 8700k CPU和GeForce RTX 2080Ti图形处理器的工作站上通过TensorFlow-GPU训练模型，操作系统为Linux Ubuntu 16.04。SAFFGCNN的训练过程采用基于动量的随机梯度下降(Stochastic Gradient Descent, SGD)优化算法，采用Adam优化算法更新SGD步长。

4.1　实验数据集及评价指标

S3DIS[22]数据集由来自3个不同建筑的6个大型室内区域共计271个房间组成，每个房间都由一个中等大小的密集点云组成(约20 m×15 m×5 m)，共标注了13个类别。实验中使用标准的6重交叉验证。

SemanticKITTI[23]数据集是目前最大的具有点级注释的激光雷达序列数据集，包含了复杂的室外交通场景，由43 552个密集注释激光雷达扫描组成22个序列，共包含19个有效类别。实验中，数据集中序列00～10作为训练集(其中序列08用作验证集)，序列11～21作为测试集。

平均交并比(mean Intersection over Union, mIoU)作为实验结果的主要评估指标，其公式如下：

$$mIoU=\frac{1}{(c+1)}\sum_{i=0}^{c}\frac{TP}{(FN+FP+TP)}. \tag{12}$$

总体准确率(Overall Accuracy, OA)作为实验结果的参考评估指标，用正确预测分类的点数和总体点数的比值表示：

$$OA=\frac{TP+TN}{TP+TN+FP+FN}. \tag{13}$$

4.2　语义分割评估

4.2.1　S3DIS数据集上的评估分析

为了验证本文算法的有效性，在S3DIS数据集上进行了分割对比实验，结果如表1所示。

表1　S3DIS数据集上不同方法的分割精度对比(六重交叉验证)

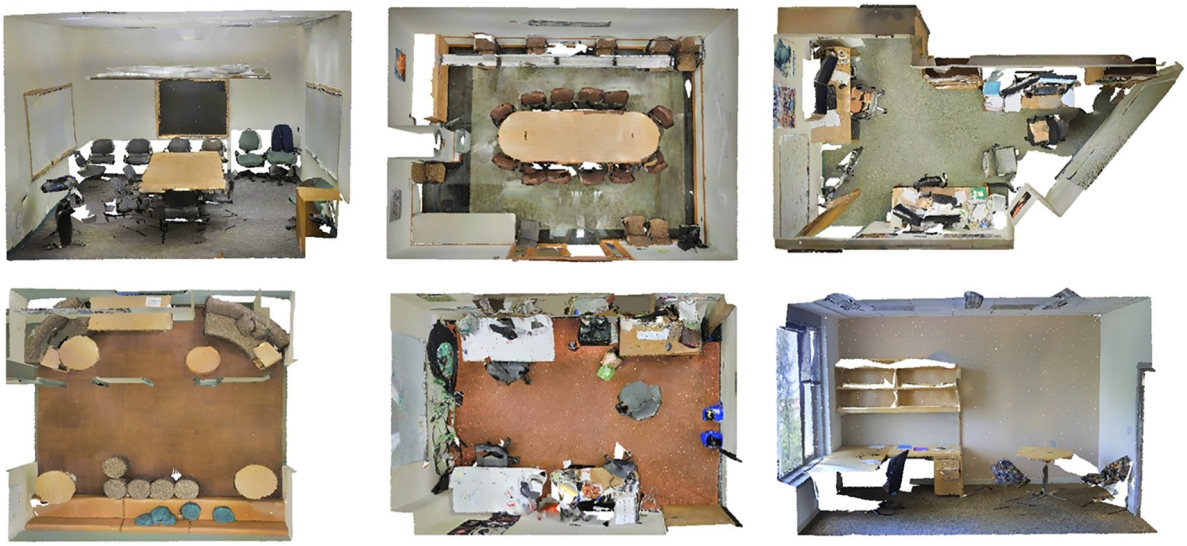
Tab. 1　Comparison of segmentation accuracy of different approaches on S3DIS dataset(6-fold cross-validation)

Methods	OA/%	mIoU/%	Ceil	Floor	Wall	Beam	Colu	Wind	Door	Chair	Table	Book	Sofa	Board	Clut
PointNet[16]	78.5	47.6	88.0	88.7	69.3	42.4	23.1	47.5	51.6	42.0	54.1	38.2	09.6	29.4	35.2
PointCNN[26]	88.1	65.4	94.8	97.3	75.8	63.3	51.7	58.4	57.2	71.6	69.1	39.1	61.2	52.2	58.6
PointWeb[27]	87.3	66.7	93.5	94.2	80.8	52.4	41.3	64.9	68.1	67.1	71.4	62.7	50.3	62.2	58.5
RandLA-Net[20]	88.0	70.0	93.1	96.1	80.6	62.4	48.0	64.4	69.4	76.4	69.4	64.2	60.0	**65.9**	60.1
KPConv[25]	92.9	70.6	93.6	92.4	83.1	63.9	54.3	66.1	76.6	57.8	64	69.3	74.9	61.3	60.3
Point Transformer[24]	90.2	73.5	94.3	97.5	84.7	55.6	58.1	66.1	78.2	74.1	77.6	**71.2**	67.3	65.7	64.8
Ours	93.1	79.3	**96.7**	**97.9**	**86.4**	**82.7**	**61.9**	**78.3**	**80.4**	**79.3**	**87.6**	69.4	**75.9**	61.2	**72.9**

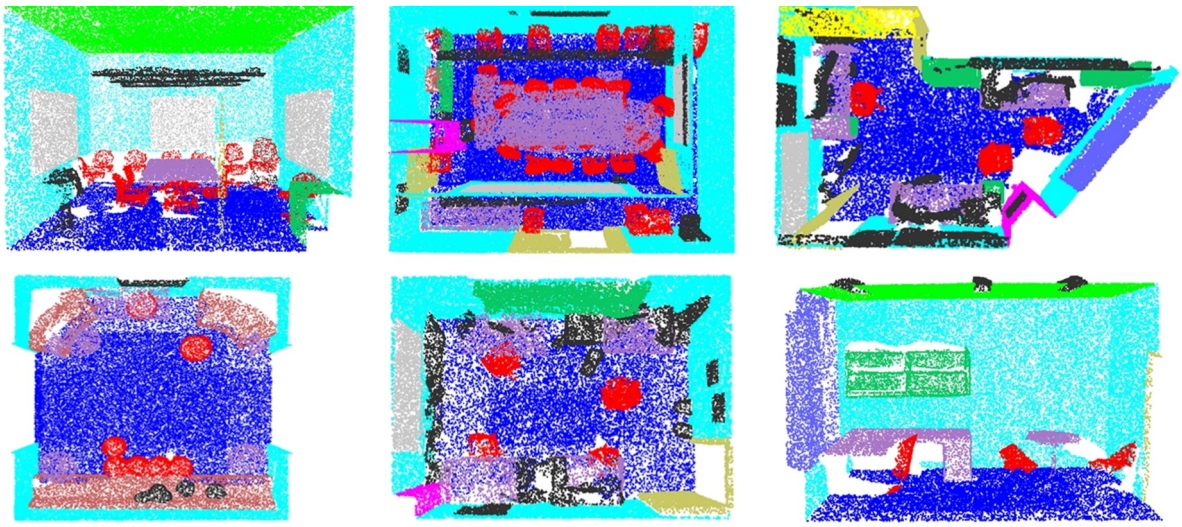
本文算法在13个类别中的11个类别上获得了最佳分割精度结果，尤其在光束、桌子、椅子和杂乱物体等类别上具有更好的分割精度。Point Transformer[24]设计自注意力层提取点云邻域特征，能够获得充分的全局单点特征，但通过MLP获得的位置信息主要用于生成查询向量，仅简单描述点对之间的相对位置关系，缺乏对几何特征的进一步提取，网络捕获高级局部几何特征信息的能力弱。本文通过代理点图组卷积能够获得细粒度的几何特征信息，引入自注意力机制探究

全局特征和局部几何特征之间的联系，使网络具备识别物体全局结构的能力，mIoU 和 OA 分别提高了 5.8% 和 2.9%。KPConv[25] 手工设计固定数目的核心点学习局部邻域点特征，但手工制作的核心点组合并不是最佳的，需要根据数据集或网络架构进行优化。此外，在网络中加入核心点位置偏移训练使球体拟合三维点云局部几何结构，无法从根本上解决卷积缺乏灵活性的问题，不能够模拟复杂三维场景中物体的位置变化。本文利用原始点云构造图结构，能够灵活且高效模拟点云的复杂空间变化和几何结构，而且 Transformer 模块能够通过特征间关联获得局部上下文细粒度的几何结构信息，mIoU 和 OA 分别提高了 8.7% 和 0.2%。

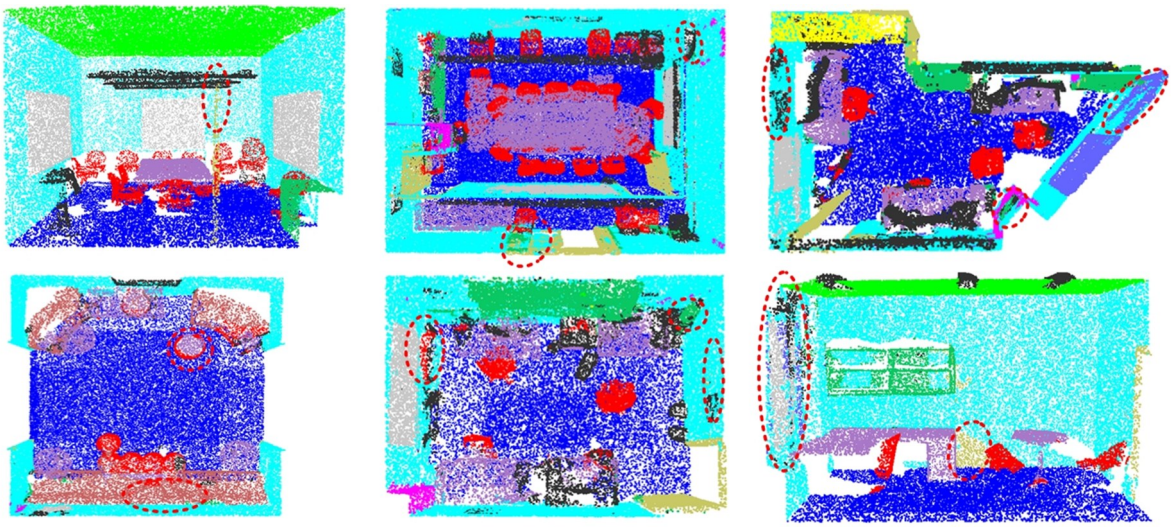
从图 4 分割可视化结果中可以看出，网络增强了识别细节采样点几何信息的能力，能够更加准确地确定物体的边界范围，使本文算法的分割结果接近于真实标签。图 4 中虚线圆圈标记为分割结果不理想的部分，对于错分割问题，网络依旧对物体几何结构信息做出比较准确的判断；对于欠分割问题，网络能够识别物体位置范围，减轻错误分类对正确结果的干扰。



(a) 真实场景
(a) Real scene



(b) 真实标签
(b) Real label



(c) 分割结果
(c) Segmentation results

图 4　S3DIS 数据集分割结果的可视化

Fig. 4　Visualization of segmentation results on S3DIS dataset

4.2.2　SemanticKITTI 数据集上的评估分析

大规模场景分割是一项具有挑战性的任务，为了进一步验证本文算法对于细粒度几何特征分析的有效性，在大规模激光雷达点云数据集 SemanticKITTI 上进行了对比实验，结果如表 2 所示。

RandLA-Net[20]采用随机采样高效处理大规模点云，设计局部特征聚合模块逐步增加点的感受野，防止采样过程丢失关键信息，但在稀疏性较大的激光雷达数据集不可避免地会丢失场景边缘信息。网络会由于边缘物体信息丢失缺乏对物体完整结构的学习，出现错分割或欠分割。本文算法采用最远点采样更能表征场景的整体结构信息，保证网络输入能够获得边缘物体的完

表 2　SemanticKITTI 数据集上不同方法的分割精度对比

Tab. 2　Comparison of segmentation accuracy of different approaches on SemanticKITTI dataset

Method	mIoU/%	Road	Sidewalk	Parking	Other-ground	Building	Car	Truck	Bicycle	Motorcycle
PointNet[16]	14.6	61.6	35.7	15.8	1.4	41.4	46.3	0.1	1.3	0.3
SPG[21]	17.4	45.0	28.5	1.6	0.6	64.3	49.3	0.1	0.2	0.2
PointNet++[17]	20.1	72.0	41.8	18.7	5.6	62.3	53.7	0.9	1.9	0.2
SqueezeSeg[14]	29.5	85.4	54.3	26.9	4.5	57.4	68.8	3.3	16.0	4.1
SqueezeSegV2[29]	39.7	88.6	67.6	45.8	17.7	73.3	81.8	13.4	18.5	17.9
PointASNL[30]	46.8	87.4	74.3	24.3	1.8	83.1	87.9	39.0	0.0	25.1
DarkNet21Seg[23]	47.4	91.4	74.0	57.0	26.4	81.9	85.4	16.6	26.2	26.5
DarkNet53Seg[23]	49.9	**91.8**	74.6	64.8	27.9	84.1	86.4	25.5	24.5	32.7
HPGCNN[19]	50.5	89.5	73.6	58.8	34.6	91.2	93.1	21.0	6.5	17.6
RangeNet++[12]	52.2	**91.8**	**75.2**	**65.0**	27.8	87.4	91.4	25.7	25.7	34.4
RandLA-Net[20]	53.9	90.7	73.7	60.3	20.4	86.9	94.2	**40.1**	26.0	25.8
PolarNet[28]	54.3	90.8	74.4	61.7	21.7	90.0	93.8	22.9	**40.3**	30.1
SqueezeSegv3[13]	55.9	91.7	74.8	63.4	26.4	89.0	92.5	29.6	38.7	**36.5**
Ours	56.6	89.9	73.9	63.5	**35.1**	**91.5**	**95.0**	38.3	33.2	35.1

(续表 2)

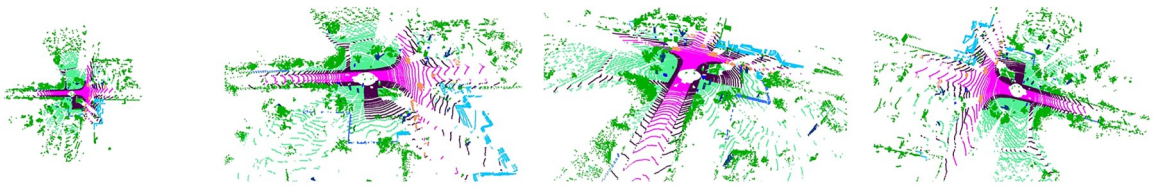
Method	Other-vehicle	Vege-tation	Trunk	Terrain	Person	Bicyclist	Motor-cyclist	Fence	Pole	Traffic-sign
PointNet[16]	0.8	31.0	4.6	17.6	0.2	0.2	0.0	12.9	2.4	3.7
SPG[21]	0.8	48.9	27.2	24.6	0.3	2.7	0.1	20.8	15.9	0.8
PointNet++[17]	0.2	46.5	13.8	30.0	0.9	1.0	0.0	16.9	6.0	8.9
SqueezeSeg[14]	3.6	60.0	24.3	53.7	12.9	13.1	0.9	29.0	17.5	24.5
SqueezeSegV2[29]	14.0	71.8	35.8	60.2	20.1	25.1	3.9	41.1	20.2	36.3
PointASNL[30]	29.2	84.1	52.2	**70.6**	34.2	**57.6**	0.0	43.9	**57.8**	36.9
DarkNet21Seg[23]	15.6	77.6	48.4	63.7	31.8	33.6	4.0	52.3	36.0	50.0
DarkNet53Seg[23]	22.6	78.3	50.1	64.0	36.2	33.6	4.7	55.0	38.9	52.2
HPGCNN[19]	23.3	84.4	65.9	70.0	32.1	30.0	14.7	65.5	45.5	41.5
RangeNet++[12]	23.0	80.5	55.1	64.6	38.3	38.8	4.8	58.6	47.9	55.9
RandLA-Net[20]	**38.9**	81.4	61.3	66.8	**49.2**	48.2	7.2	56.3	49.2	47.7
PolarNet[28]	28.5	84.0	65.5	67.8	43.2	40.2	5.6	61.3	51.8	57.5
SqueezeSegv3[13]	33.0	82.0	58.7	65.4	45.6	46.2	**20.1**	59.4	49.6	**58.9**
Ours	28.7	**84.4**	**67.1**	69.5	45.3	43.5	7.3	**66.1**	54.3	53.7

整结构信息。而且,本文在原始点云构造的 k 近邻图经过最远点采样后,依旧能够保留场景边缘物体的整体几何信息,保证特征的丰富性,在栅栏和其他地面等较稀疏的类别上 mIoU 比 RandLA-Net 分别提高了 9.8% 和 14.7%。PolarNet[28] 设计极化鸟瞰图平衡网格内点数,利用简易 PointNet 将点转换为固定长度表示,将该表示分配到环矩阵中相应的位置,通过环形卷积学习二维特征。虽然极化鸟瞰图解决了点云稀疏性问题,但自上而下的处理方式破坏了物体的几何结构信息,缺乏具有抽象语义识别能力的高级单点特征。而本文通过 MLP 组卷积获取全局单点特征,再利用代理点图卷积获得具有鉴别性的高级单点特征,引入 Transformer 模块学习点对之间的语义关系,获得局部上下文细粒度的几何信息,增强了网络的识别分割能力,在货车、摩托车和骑自行车的人等复杂结构类别的 mIoU 比 PolarNet 分别提高了 15.4%,5% 和 3.3%。

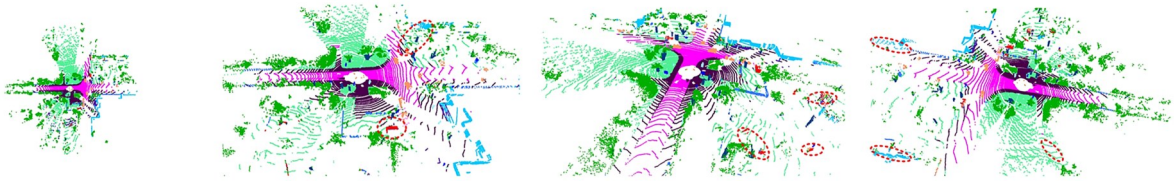
从图 5 可视化分割结果可以看出,本文算法具有提取局部上下文几何信息的能力,在稀疏性较大的大规模激光雷达点云数据中依然有着良好的分割结果。复杂结构类别由于点云的稀疏性导致物体信息不充分,加大了网络提取特征的难度,但本文对复杂类别精度相比其他方法有明显的提升,原因在于特征融合过程中加强了全局信息和局部信息交流,获得的上下文细粒度信息有助于提高网络识别复杂形状物体的能力,增强了语义分割的鲁棒性。

4.3　消融实验

S3DIS 数据集中点云密度一致,物体信息丰富,点云下采样操作对输入信息损失较少,不同配置下的模块性能都能够充分发挥,对比实验更具说服性。因此,在 S3DIS 数据集上进行了消融



(a) 真实标签
(a) Real label



(b) 分割结果
(b) Segmentation results

图 5　SemanticKITTI 数据集分割结果的可视化

Fig. 5　Visualization of segmentation results on SemanticKITTI dataset

实验。考虑网络模型的各种设置，比较了模型在 k 近邻点数不同下的性能，以验证本文算法代理点图组卷积和 Transformer 特征融合模块的有效性。

4.3.1　k 近邻点

邻域点的数目影响网络提取到的几何特征的优劣，较小的邻域点数目 k 使网络无法学习到有效的几何特征，导致分割精度较差；而 k 的数量过大又会引入更多的噪声，影响网络对几何特征的学习。从表 3 中可以看出，当 k 为 8 时，网络总参数量 Params 和 OA 都较小，原因在于邻域图对物体几何信息的描述不完整，网络性能无法充分利用而造成欠分割问题。随着 k 的增加，邻域图能够更好地表征物体的几何结构，网络能够充分挖掘局部上下文的几何信息。但当 k 过大时，对物体的几何结构描述无法带来更大的优势，相反会造成更多冗余的局部几何结构特征，影响具有区分性的局部几何特征的贡献程度，而且增加网络计算量。

表 3　邻域点数量对分割结果影响的对比

Tab. 3　Comparison of influence of number of neighborhood points on segmentation results

k	Params/MB	OA/%
8	0.93	92.4
16	0.95	93.1
32	0.98	92.8

4.3.2　P2GConv

为了验证代理点图组卷积（P2GConv）在保持较少的参数量的同时可以获得与边缘卷积（EdgeConv）相当的结果，对网络分别使用 P2GConv 和 EdgeConv，定量实验结果如表 4 所示。使用 P2GConv 的网络参数量更少，原因在于构建局部邻域图不需要重复计算中心点的邻域点，取消了在特征图上的 k 近邻图构建。此外，代理点是手工设计，计算边缘特征时不会出现 EdgeConv 中添加中心点特征的操作。而在分割精度方面，P2GConv 接近 EdgeConv，原因：一方面在于代理点特征是邻域点特征的平均值，场景中平面结构多且特征差异性小，代理点特征能够表征局部邻域点的特征信息，仅会损失特征的一小部分丰富性；另一方面，由于在原始空间构建的邻域图对物体几何信息的描述更加准确，P2GConv 网络能够获得物体细粒度的几何结构信息。

表 4　边缘卷积和代理点图组卷积对比

Tab. 4　Comparison of EdgeConv and P2GConv

Convolution	Params/MB	mIoU/%
EdgeConv	0.86	78.1
P2GConv	0.79	77.9

4.3.3　MLPGConv

MLP 组卷积将全局单点特征输入代理点图组卷积，获得有助于识别物体的高级全局单点特征，增强了特征的局部上下文信息。当删除 MLP 组卷积操作后，局部几何特征只对自身进行自注意力融合操作，融合后的特征依旧能够充分表达局部区域的细节信息。但由于忽略每个点的绝对位置信息，缺乏从点云空间中学习到的全局单点结构特征，从而降低了特征丰富性，无法获得具备高级语义识别能力的上下文语义信息，导致网络识别能力下降而影响分割

精度。虽然参数量有一定下降,但精度的增长对网络整体性能的提升更大。实验结果如表5所示,其中MLPG-NO表示不引入MLPGConv模块。

表5　MLPGConv模块有效性验证

Tab. 5　Effectiveness verification of MLPGConv module

Module	Params/MB	mIoU/%
MLPG-NO	0.88	74.6
MLPG	0.95	79.3

4.3.4　Transformer

网络加入Transformer模块的自注意力机制,分割精度和网络参数量都有明显增长。实验结果如表6所示,其中Transformer-NO表示不引入Transformer模块。网络参数量增长在于:对特征的额外操作增加了网络计算量。分割精度增长的原因在于点对之间的语义关系和局部细粒度的上下文信息。学习点对之间的语义关系能够提高网络识别复杂环境中物体的能力,减少错分割现象。全局单点特征和局部几何特征融合后获得局部细粒度的上下文信息,获得物体局部的几何结构信息,解决了欠分割或过分割问题,提高了网络细粒度分割精度。

表6　Transformer模块有效性验证

Tab. 6　Effectiveness verification of Transformer module

Module	Params/MB	mIoU/%
Transformer-NO	0.79	77.9
Transformer	0.95	79.3

5　结　论

本文提出了一种自注意力特征融合组卷积神经网络的三维点云语义分割算法。首先,利用MLP组卷积获得全局点云特征;其次,通过代理点图组卷积获得细粒度的几何特征信息;然后,通过Transformer特征融合模块的自注意机制加强全局和局部几何特征之间的联系,挖掘局部上下文几何信息;最后,通过多尺度操作扩大局部邻域感受野,进一步增强捕获细粒度局部上下文几何信息的能力。通过轻量化特征提取网络,以较少的冗余信息增强了特征的丰富性,实现了对点云的高性能处理,在S3DIS数据集和SemanticKITTI数据集上算法的分割精度分别达到79.3%和56.6%。

然而,本文算法仍存在一定的局限性,一方面在于网络分析复杂环境下物体类别时存在不足,具有相似几何结构特征的物体在空间上接近时,网络对物体边界点类别的判断不准确,周围类别影响网络对物体整体结构的判断,出现欠分割或错分现象,网络抗干扰能力有待提高;另一方面在于网络处理稀疏性较强点云数据集时效果不理想,物体远离传感器导致描述同部件几何信息的点云数目减少,影响网络从采样后点云学习物体的几何信息。所以,在非常稀疏数据集下保留更丰富信息和有效处理场景边缘物体是未来研究的重点。

参考文献:

[1] QI C R, LIU W, WU C X, *et al*. Frustum PointNets for 3D object detection from RGB-D data [C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition. June* 18-23, 2018, *Salt Lake City*, *UT*, *USA*. *IEEE*, 2018: 918-927.

[2] LIU Z Z, CHEN H Y, DI H J, *et al*. Real-time 6D lidar SLAM in large scale natural terrains for UGV [C]. 2018 *IEEE Intelligent Vehicles Symposium. June* 26-30, 2018, *Changshu*, *China*. *IEEE*, 2018: 662-667.

[3] GOLOVINSKIY A, KIM V G, FUNKHOUSER T. Shape-based recognition of 3D point clouds in urban environments[C]. 2009 *IEEE* 12*th International Conference on Computer Vision. September* 29 - *October* 2, 2009, *Kyoto*, *Japan*. *IEEE*, 2009: 2154-2161.

[4] LAWIN F J, DANELLJAN M, TOSTEBERG P, *et al*. Deep projective 3D semantic segmentation [C]. *International Conference on Computer Analysis of Images and Patterns* (*CAIP*). *Aug*. 22-24, 2017, *Ystad*, *Sweden*. Cham: Springer, 2017: 95-107.

[5] HUANG J, YOU S Y. Point cloud labeling using 3D Convolutional Neural Network[C]. 2016 23*rd International Conference on Pattern Recognition* (*ICPR*). *December* 4-8, 2016, *Mexico*: *Cancun. IEEE*, 2016: 2670-2675.

[6] GRAHAM B, ENGELCKE M, MAATEN L V D. 3D semantic segmentation with submanifold sparse convolutional networks[C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition. June* 18-23, 2018, *Salt Lake City*, *UT*, *USA. IEEE*, 2018: 9224-9232.

[7] LIU Z J, TANG H T, LIN Y J, *et al*. Point-voxel CNN for efficient 3D deep learning[J]. *CoRR*, 2019, abs/1907.03739.

[8] 杨军, 党吉圣. 采用深度级联卷积神经网络的三维点云识别与分割[J]. 光学 精密工程, 2020, 28(5): 1187-1199.
YANG J, DANG J SH. Recognition and segmentation of three-dimensional point cloud based on deep cascade convolutional neural network[J]. *Opt. Precision Eng.*, 2020, 28(5): 1187-1199. (in Chinese)

[9] 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.
YANG J, DANG J SH. Semantic segmentation of 3D point cloud based on contextual attention CNN[J]. *Journal on Communications*, 2020, 41(7): 195-203. (in Chinese)

[10] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, *et al*. An image is worth 16×16 words: transformers for image recognition at scale[J]. *ArXiv Preprint ArXiv*: 2010.11929, 2020.

[11] QI D, SU L, SONG J, *et al*. Imagebert: Cross-modal pre-training with large-scale weak-supervised image-text data[J]. *ArXiv Preprint ArXiv*: 2001.07966, 2020.

[12] MILIOTO A, VIZZO I, BEHLEY J, *et al*. RangeNet ++: fast and accurate LiDAR semantic segmentation[C]. 2019 *IEEE/RSJ International Conference on Intelligent Robots and Systems* (*IROS*). *November* 3-8, 2019, *Macao*, *China. IEEE*, 2019: 4213-4220.

[13] XU C F, WU B C, WANG Z N, *et al*. SqueezeSegV3: spatially-adaptive convolution for efficient point-cloud segmentation[C]. *European Conference on Computer Vision* (*ECCV*). Aug. 23-28, 2020, Online. Cham: Springer, 2020: 1-19..

[14] 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. May* 21-25, 2018, *Brisbane*, *QLD*, *Australia. IEEE*, 2018: 1887-1893.

[15] CHOY C, GWAK J, SAVARESE S. 4D spatio-temporal ConvNets: minkowski convolutional neural networks[C]. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *June* 15-20, 2019, *Long Beach*, *CA*, *USA. IEEE*, 2019: 3070-3079.

[16] 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. July* 21-26, 2017, *Honolulu*, *HI*, *USA. IEEE*, 2017: 77-85.

[17] QI C R, YI L, SU H, *et al*. PointNet++: deep hierarchical feature learning on point sets in a metric space[EB/OL]. 2017: *arXiv*: 1706.02413[cs.CV]. https://arxiv.org/abs/1706.02413

[18] 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*, *Korea* (*South*). *IEEE*, 2019: 10432-10440.

[19] DANG J S, YANG J. *HPGCNN: Hierarchical Parallel Group Convolutional Neural Networks for Point Clouds Processing*[M]. Computer Vision-ACCV 2020. Cham: Springer International Publishing, 2021: 20-37.

[20] HU Q Y, YANG B, XIE L H, *et al*. RandLA-net: efficient semantic segmentation of large-scale point clouds[C]. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *June* 13-19, 2020, *Seattle*, *WA*, *USA. IEEE*, 2020: 11105-11114.

[21] 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. IEEE*, 2018: 4558-4567.

[22] 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. June* 27-30, 2016, *Las Vegas*, *NV*, *USA. IEEE*, 2016: 1534-1543.

[23] BEHLEY J, GARBADE M, MILIOTO A, *et al*. SemanticKITTI: a dataset for semantic scene understanding of LiDAR sequences [C]. 2019 *IEEE/CVF International Conference on Computer Vision* (*ICCV*). *October* 27-*November* 2, 2019, *Seoul*, *Korea* (*South*). *IEEE*, 2019: 9296-9306.

[24] ZHAO H, JIANG L, JIA J, *et al*. Point transformer[C]. *Proceedings of the IEEE/CVF International Conference on Computer Vision*. October 11-17, 2021, Online. IEEE, 2021: 16259-16268.

[25] 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*, *Korea* (*South*). *IEEE*, 2019: 6410-6419.

[26] LI Y, BU R, SUN M, *et al*. PointCNN: convolution on x-transformed points[J]. *Advances in Neural Information Processing Systems*, 2018, 31: 820-830.

[27] ZHAO H S, JIANG L, FU C W, *et al*. PointWeb: enhancing local neighborhood features for point cloud processing[C]. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *June* 15-20, 2019, *Long Beach*, *CA*, *USA. IEEE*, 2019: 5560-5568.

[28] ZHANG Y, ZHOU Z X, DAVID P, *et al*. PolarNet: an improved grid representation for online LiDAR point clouds semantic segmentation [C]. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *June* 13-19, 2020, *Seattle*, *WA*, *USA. IEEE*, 2020: 9598-9607.

[29] 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]. 2019 *International Conference on Robotics and Automation* (*ICRA*). *May* 20-24, 2019, *Montreal*, *QC*, *Canada. IEEE*, 2019: 4376-4382.

[30] 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. IEEE*, 2020: 5588-5597.

作者简介:

杨　军(1973—),男,宁夏吴忠人,博士,教授,博士生导师,1995年于西北师范大学获得学士学位,2002年于兰州交通大学获得硕士学位,2007年于西南交通大学获得博士学位,主要从事三维模型的空间分析、遥感影像的分析与处理、模式识别等方面的研究。E-mail: yangj@mail.lzjtu.cn

通讯作者:

李博赞(1995—),男,河南洛阳人,硕士研究生,2018年于河南理工大学获得学士学位,主要从事计算机视觉、模式识别等方向的研究。E-mail: 763311755@qq.com