E-mail: jig@aircas.ac.cn
Website: www.cjig.cn
Tel: 010-58887035
中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS

中图法分类号:TP391　　**文献标识码**: A　　**文章编号**: 1006-8961(2024)08-2388-11

论文引用格式:Zhu Z J, Zhang R, Bai Y Q, Wang Y E and Sun J M. 2024. Bilateral cross enhancement with self-attention compensation for semantic segmentation of point clouds. Journal of Image and Graphics, 29(08):2388-2398(朱仲杰, 张荣, 白永强, 王玉儿, 孙嘉敏. 2024. 结合双边交叉增强与自注意力补偿的点云语义分割. 中国图象图形学报, 29(08):2388-2398)[DOI:10.11834/jig.230430]

结合双边交叉增强与自注意力补偿的点云语义分割

朱仲杰[1,2*],张荣[1],白永强[1],王玉儿[1],孙嘉敏[1,2]

1. 浙江万里学院宁波市DSP重点实验室,宁波　315000; 2. 中国海洋大学信息科学与工程学院,青岛　266000

摘　要:**目的**　针对现有点云语义分割方法对几何与语义特征信息利用不充分,导致分割性能不佳,特别是局部细粒度分割精度不足的问题,提出一种结合双边交叉增强与自注意力补偿的充分融合几何与语义上下文信息的点云语义分割新算法以提升分割性能。**方法**　首先,设计基于双边交叉增强的空间聚合模块,将局部几何与语义上下文信息映射到同一空间进行交叉学习增强后聚合为局部上下文信息。然后,基于自注意力机制提取全局上下文信息与增强后的局部上下文信息进行融合,补偿局部上下文信息的单一性,得到完备特征图。最后,将空间聚合模块各阶段输出的多分辨率特征输入特征融合模块进行多尺度特征融合,得到最终的综合特征图以实现高性能语义分割。**结果**　实验结果表明,在S3DIS(Stanford 3D indoor spaces dataset)数据集上,本文算法的平均交并比(mean intersection over union,mIoU)、平均类别精度(mean class accuracy,mAcc)和总体精度(overall accuracy,OA)分别为70.2%、81.7%和88.3%,与现有优秀算法RandLA-Net相比,分别提高2.4%、2.0%和1.0%。同时,对S3DIS数据集Area 5单独测试,本文算法的mIoU为66.2%,较RandLA-Net提高5.0%。**结论**　空间聚合模块不仅能够充分利用局部几何与语义上下文信息增强局部上下文信息,而且基于自注意力机制融合局部与全局上下文信息,增强了特征的完备性以及局部与全局的关联性,可以有效提升点云局部细粒度的分割精度。在可视化分析中,相较于对比算法,本文算法对点云场景的局部细粒度分割效果明显提升,验证了本文算法的有效性。

关键词:点云;语义分割;双边交叉增强;自注意力机制;特征融合

Bilateral cross enhancement with self-attention compensation for semantic segmentation of point clouds

Zhu Zhongjie[1,2*], Zhang Rong[1], Bai Yongqiang[1], Wang Yuer[1], Sun Jiamin[1,2]

1. Ningbo Key Laboratory of DSP, Zhejiang Wanli University, Ningbo　315000, China;

2. Faculty of Information Science and Engineering, Ocean University of China, Qingdao　266000, China

Abstract: **Objective** Point cloud semantic segmentation is a computer vision task that aims to segment 3D point cloud data and assign corresponding semantic labels to each point. Specifically, according to the location and other attributes of the point, point cloud semantic segmentation involves assigning each point to predefined semantic categories, such as ground, buildings, vehicles, and pedestrians. Existing methods for point cloud semantic segmentation can be broadly categorized into three types: projection-, voxel-, and raw point cloud-based methods. Projection-based methods project the 3D point

收稿日期:2023-07-11;**修回日期**:2023-11-09;**预印本日期**:2023-11-16

*通信作者:朱仲杰　zhongjiezhu@yeah.net

基金项目:国家自然科学基金项目(61671412);浙江省自然科学基金项目(LY21F010014);宁波市科技创新2025重大专项(2022Z076)

Supported by: National Natural Science Foundation of China (61671412); Natural Science Foundation of Zhejiang Province, China (LY21F010014); Ningbo Municipal Major Project of Science and Technology Innovation 2025 (2022Z076)

cloud onto a 2D plane (e. g. , an image) and then apply standard image-based segmentation techniques. Voxel-based methods divide the point cloud space into regular voxel grids and assign semantic labels to each voxel. Both methods require data transformation, which inevitably leads to some loss of feature information. By contrast, raw point cloud-based methods directly process the point cloud without any transformation, which ensures the integrity of the input algorithm network with the original point cloud data. The geometric and semantic feature information of each point in the point cloud scene needs to be fully considered and utilized to achieve accurate semantic segmentation tasks. Existing methods for point cloud semantic segmentation generally extract, process, and utilize geometric and semantic feature information separately, without considering their correlation. This approach leads to less precise local fine-grained segmentation. Therefore, this study proposes a new algorithm for point cloud semantic segmentation based on bilateral cross-enhancement and self-attention compensation. It not only fully utilizes the geometric and semantic feature information of the point cloud but also constructs offsets between them as a medium for information interaction. In addition, the fusion of local and global feature information is achieved, which enhances feature completeness and overall segmentation performance. This fusion process enhances the integrity of features and ensures the full representation and utilization of local and global contexts during the segmentation process. By considering the overall information of the point cloud scene, this algorithm demonstrates better performance in segmenting local fine-grained details and larger-scale structures. **Method** First, the original input point cloud data are preprocessed to extract geometric contextual information and initial semantic contextual information. The geometric contextual information is represented by the original coordinates of the point cloud in 3D space, while the initial semantic contextual information is extracted using a multilayer perceptron. Next, a spatial aggregation module is designed, which consists of bilateral cross-enhancement and self-attention mechanism units. In the bilateral cross-enhancement units, local geometric and semantic contextual information is preliminarily extracted by constructing local neighborhoods for the preprocessed geometric contextual information and initial semantic contextual information. Then, offsets are constructed to facilitate cross-learning and enhancement of the local geometric and semantic contextual information by mapping it onto a common space. Finally, the enhanced local geometric and semantic contextual information is aggregated to local contextual information. Next, using the self-attention mechanism, global contextual information is extracted and fused with the local contextual information to compensate for the singularity of the local contextual information, which results in a comprehensive feature map. Finally, the multi-resolution feature maps obtained at different stages of the spatial aggregation module are fed into the feature fusion module for multi-scale feature fusion, which produces the final comprehensive feature map. Thus, high-performance semantic segmentation is achieved. **Result** Experimental results on the Stanford 3D indoor spaces dataset (S3DIS) show a mean intersection over union (mIoU) of 70. 2%, a mean class accuracy of (mAcc) 81. 7%, and an overall accuracy (OA) of 88. 3%, which are 2. 4%, 2. 0%, and 1. 0% higher than those of the existing representative algorithm RandLA-Net. Meanwhile, for Area 5 of the S3DIS, the mIoU is 66. 2%, which is 5. 0% higher than that of RandLA-Net. In addition, visualizations of the segmentation results are achieved on the Semantic3D dataset. **Conclusion** By utilizing the spatial aggregation module, the proposed algorithm maximizes the utilization of geometric and semantic contextual information, which enhances the details of local contextual information. In addition, the integration of local and global contextual information through self-attention mechanism ensures comprehensive feature representation. As a result, the proposed algorithm achieves a significant improvement in the segmentation accuracy of fine-grained details in point clouds. Visual analysis further validates the effectiveness of the algorithm. Compared with baseline algorithms, the proposed algorithm demonstrates clear superiority in the fine-grained segmentation of local regions in point cloud scenes. This result serves as partial evidence, which confirms the effectiveness of the proposed algorithm in addressing challenges related to point cloud segmentation tasks. In conclusion, the spatial aggregation module and its fusion of local and global contextual information significantly improve the segmentation accuracy of local details in point clouds. This approach offers a promising solution to enhance the segmentation accuracy of fine-grained details in point cloud local regions.

Key words: point cloud; semantic segmentation; bilateral cross enhancement; self-attention mechanism; feature fusion

0 引 言

点云语义分割在自动驾驶(郑阳 等,2021)、计算机视觉(杜静和蔡国榕,2021)、机器人(Giang和Ryoo,2023)等多个领域具有广泛应用。然而,由于点云分布具有稀疏性、不规则性和无序性等特征(刘盛 等,2023),点云语义分割相较于二维图像语义分割更具挑战性。

根据点云预处理方式不同,早期点云语义分割主要分为基于投影和基于体素化两类方法。基于投影的方法(Boulch等,2018;Alonso等,2020)首先将点云转化为二维图像进行分割,然后将像素级语义标签反投影至点云空间完成分割,算法复杂度相对较低,但会丢失点云内部的几何结构信息,从而导致分割精度不高;基于体素化的方法(Poux和Billen,2019;Luo等,2020)利用三维卷积将点云正则化为网格结构后进行分割,可以有效保留点云内部几何结构信息,但难以对目标边界的几何形状进行精细分割。上述两类方法均对原始输入点云进行预处理,导致部分几何信息丢失,因此其语义分割精度通常不高,难以满足实际应用需求。

为了有效保留并利用点云空间几何信息提升分割性能,一些学者专注于研究基于原始点云的语义分割方法。其中,PointNet(Qi等,2017a)是第1个直接使用原始点云作为输入的神经网络,通过共享的多层感知机逐点学习特征,解决了点云稀疏性、无序性以及置换不变性的问题,但不能很好地提取点云局部特征信息。为此,PointNet++(Qi等,2017b)在PointNet的基础上逐点邻域迭代性特征提取,解决局部特征缺失问题,提升分割效果。

RSNet(recurrent slice network)(Huang等,2018)在PointNet++的基础上引入了轻量级的局部依赖模块,有效模拟点云中的局部结构。Pointweb(Zhao等,2019)通过自适应特征调整模块寻找点之间的相互作用,较好地反映局部区域的上下文信息。然而,这两种方法都缺乏对全局特征的充分考虑,因此分割精度不高。为了克服这一问题,MNFEAM(multi-scale neighborhood feature extraction and aggregation model)(Li等,2021)通过多尺度特征聚合方法,从编码器和解码器中捕获点云的全局特征,但没有充分利用点云位置和方向等重要空间信息,影响了分割精度。为了更好地学习空间几何信息,RandLA-Net(Hu等,2022)通过逐渐增大每个点的感受野,以感知学习复杂的空间几何结构,取得较好分割结果,但在复杂点云场景中容易丢失局部特征。CSANet(cross self-attention network)(Wang等,2022)通过交叉自注意力模块学习点云坐标和特征信息,融合点云几何和语义信息,取得了较好的局部分割效果,但缺乏对局部和全局信息的综合考虑。PA-Net(point attention network)(Ren等,2022)利用注意力模块将语义相互依赖和远程依赖相结合,充分考虑局部和全局信息之间的关联性以提升点云语义分割性能,但对几何信息利用不足。GeoSegNet(Chen等,2023)通过引入多几何特征编码器和边界引导解码器,高效利用与处理几何信息,增强了物体边界的分割性能。

虽然点云分割技术近年来取得了积极进展,但在实际应用中仍存在分割不够精准、分割效率不高等问题,为此,提出一种结合双边交叉增强与自注意力补偿的点云语义分割新算法,通过充分融合几何与语义上下文信息提升点云分割性能。本文贡献包括:1)针对局部细粒度分割精度不高的问题,在局部和全局角度充分利用点云的几何与语义特征信息提升分割性能。2)从局部角度,设计了基于双边交叉增强的空间聚合模块,在局部邻域中构建偏移量充当有效、可识别的参数,将局部几何与语义上下文信息映射到同一空间进行交叉学习与聚合增强。3)从全局角度,为了补偿局部上下文信息的单一性,基于自注意力机制提取全局上下文信息,并与增强后的局部上下文信息进行高效融合,增强信息完备性,提高点云局部细粒度分割精度。

1 所提算法

提出的点云语义分割算法流程如图1所示,主要分为预处理、空间聚合和多尺度特征融合3个模块。首先,对输入原始点云数据进行预处理,提取几何与初始语义上下文信息。然后,在空间聚合模块中构建局部邻域,利用偏移量分别增强局部几何与语义上下文信息。在此基础上,基于自注意力机制提取全局上下文信息,进一步对特征信息完善与补充,增强特征的完备性以及局部与全局的关联性。最后,将不同分辨率的特征图输出至多尺度特征融合模块,进行上采样和融合,最终通过全连接层实现

点云的语义分割。

1.1 预处理模块

给定点云集合 $\boldsymbol{X}=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_i,\cdots,\boldsymbol{x}_N\}$，其中 $\boldsymbol{x}_i(1\leqslant i\leqslant N)$ 表示点云集合中的点，N 为点云集合中的点数。首先，利用 $\boldsymbol{X}$ 的坐标信息得到坐标集合 $\boldsymbol{P}=\{\boldsymbol{p}_{x1},\boldsymbol{p}_{x2},\cdots,\boldsymbol{p}_{xi},\cdots,\boldsymbol{p}_{xN}\}$ 作为几何上下文信息，其中 $\boldsymbol{p}_{xi}$ 表示点 $\boldsymbol{x}_i$ 在三维空间中的坐标矢量。然后，提取 $\boldsymbol{X}$ 的特征向量集合 $\boldsymbol{F}=\{\boldsymbol{f}_{x1},\boldsymbol{f}_{x2},\cdots,\boldsymbol{f}_{xi},\cdots,\boldsymbol{f}_{xN}\}$ 作为初始语义上下文信息，其中 $\boldsymbol{f}_{xi}$ 表示点 $\boldsymbol{x}_i$ 在语义特征空间的高维特征向量，具体为

$$\boldsymbol{f}_{xi}=M(\boldsymbol{x}_i)=\delta_\theta(BN(\eta^d_{1\times1}(\boldsymbol{x}_i))) \tag{1}$$

式中，M 表示多层感知机对高维特征向量的提取操作，δ_θ 表示带有参数 θ 的 ReLU（rectified linear unit）非线性激活函数，BN（batch normalization）表示批量归一化，η 表示卷积操作，1×1 为卷积核大小，d 为输出的特征维数，此处 $d=8$。预处理模块得到的几何上下文信息 $\boldsymbol{P}$ 以及初始语义上下文信息 $\boldsymbol{F}$ 将会输入到空间聚合模块。

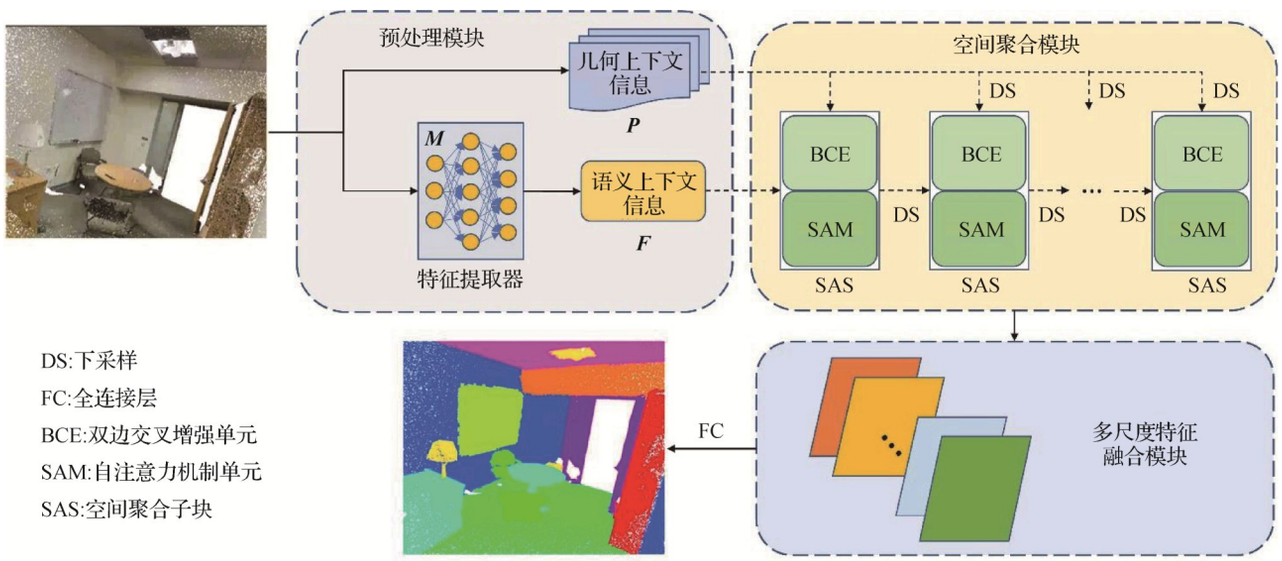


图 1　本文算法流程

Fig. 1　Proposed algorithm flow

1.2 空间聚合模块

空间聚合模块包含 5 个子块，每个子块包含双边交叉增强（bilateral cross enhancement，BCE）和自注意力机制（self-attention mechanism，SAM）两个单元，结构如图 2 所示。首先将 $\boldsymbol{P}$ 和 $\boldsymbol{F}$ 输入至双边交叉增强单元中，构建点云局部邻域，提取初步局部几何与语义上下文信息。然后交叉学习偏移量，对二者进行增强后聚合生成局部上下文信息。同时在自注意力机制单元中提取点云全局上下文信息，将其与聚合生成的局部上下文信息进行融合，得到完备特征图。

空间聚合子块采用级联的方式输出不同分辨率的特征图，每个子块输出的特征图作为下一个子块的输入，最终形成多层次的特征图，以更全面地表征点云信息。

1.2.1 双边交叉增强单元

为了充分考虑并利用点云的局部几何与语义上下文信息，设计双边交叉增强单元。首先，构建局部邻域提取局部几何与语义上下文信息。对于给定点云集合 $\boldsymbol{X}=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_i,\cdots,\boldsymbol{x}_N\}$，使用基于欧氏距离的最远点采样方法，选择 $\boldsymbol{X}$ 中的一组点作为中心点组，并记为 $\boldsymbol{Y}=\{\boldsymbol{y}_1,\boldsymbol{y}_2,\cdots,\boldsymbol{y}_i,\cdots,\boldsymbol{y}_n\}(1\leqslant i\leqslant n)$。对于每个中心点 $\boldsymbol{y}_i$，使用 KNN（K-nearest neighbors）近邻搜索近邻点 $\boldsymbol{y}_i^k=\{\boldsymbol{y}_i^1,\boldsymbol{y}_i^2,\cdots,\boldsymbol{y}_i^{K-1},\boldsymbol{y}_i^K\}$，其中，$K=16$，$\forall\boldsymbol{y}_i^k\in\boldsymbol{X}$。将中心点 $\boldsymbol{y}_i$ 和近邻点 $\boldsymbol{y}_i^k$ 的绝对坐标分别表示为 $\boldsymbol{p}_i$、$\boldsymbol{p}_i^k$，$\boldsymbol{y}_i$、$\boldsymbol{y}_i^k$ 在语义特征空间中所对应的语义特征信息分别表示为 $\boldsymbol{f}_i$、$\boldsymbol{f}_i^k$。将中心点的绝对位置和近邻点相对于中心点的相对位置组合为局部上下文信息 $\boldsymbol{\varphi}$，具体为

$$\boldsymbol{\varphi}(\boldsymbol{p}_i)=\boldsymbol{p}_i\oplus(\boldsymbol{p}_i^k-\boldsymbol{p}_i),\boldsymbol{\varphi}(\boldsymbol{p}_i)\in\mathbf{R}^{K\times6} \tag{2}$$

$$\boldsymbol{\varphi}(\boldsymbol{f}_i)=\boldsymbol{f}_i\oplus(\boldsymbol{f}_i^k-\boldsymbol{f}_i),\boldsymbol{\varphi}(\boldsymbol{f}_i)\in\mathbf{R}^{K\times2d} \tag{3}$$

式中，$\boldsymbol{\varphi}(\boldsymbol{p}_i)$ 和 $\boldsymbol{\varphi}(\boldsymbol{f}_i)$ 分别表示 3D 空间中局部几何与语义上下文信息；$\boldsymbol{p}_i^k-\boldsymbol{p}_i$ 表示近邻点相对于中心点

的相对坐标，$\boldsymbol{f}_i^k - \boldsymbol{f}_i$表示近邻点相对于中心点的语义特征信息，$d$是语义特征信息$\boldsymbol{f}_i$，$\boldsymbol{f}_i^k$的维度；⊕表示特征拼接操作。

由于点云密度在不同区域存在不均匀性，不仅影响点云的特征提取，也影响对中心点邻域特征的表征。为此，构建交叉增强网络，通过引入偏移量增加局部区域内点云密度，从而增强局部信息。

首先，为了充分考虑局部几何与语义上下文信息之间的关联，基于$\boldsymbol{\varphi}(\boldsymbol{f}_i)$，$\boldsymbol{\varphi}(\boldsymbol{p}_i)$分别计算$\boldsymbol{p}_i^k$，$\boldsymbol{f}_i^k$的偏移量，计算式为

$$\widetilde{\boldsymbol{p}_i^k} = M(\boldsymbol{\varphi}(\boldsymbol{f}_i)) + \boldsymbol{p}_i^k,\ \widetilde{\boldsymbol{p}_i^k} \in \mathbf{R}^{K\times 3} \tag{4}$$

$$\widetilde{\boldsymbol{f}_i^k} = M(\boldsymbol{\varphi}(\boldsymbol{p}_i)) + \boldsymbol{f}_i^k,\ \widetilde{\boldsymbol{f}_i^k} \in \mathbf{R}^{K\times d} \tag{5}$$

将得到的两个偏移量以交叉方式添加至局部几何与语义上下文信息中，使得$\boldsymbol{\varphi}(\boldsymbol{p}_i)$与$\boldsymbol{\varphi}(\boldsymbol{f}_i)$能够进一步学习偏移量的特征信息并进行增强，得到增强后的局部几何与语义上下文信息$\boldsymbol{A}$、$\boldsymbol{B}$，具体为

$$\boldsymbol{A} = \boldsymbol{\varphi}(\boldsymbol{p}_i)\oplus\widetilde{\boldsymbol{p}_i^k} \tag{6}$$

$$\boldsymbol{B} = \boldsymbol{\varphi}(\boldsymbol{f}_i)\oplus\widetilde{\boldsymbol{f}_i^k} \tag{7}$$

在此基础上，通过M对$\boldsymbol{A}$、$\boldsymbol{B}$进行维度调节，使得$\boldsymbol{A}$与$\boldsymbol{B}$维度一致。最后将两者聚合生成增强后局部上下文信息$\boldsymbol{Le}$，具体为

$$\boldsymbol{Le} = concat(M(\boldsymbol{A}), M(\boldsymbol{B})),\ \boldsymbol{Le} \in \mathbf{R}^{K\times d_{\text{out}}} \tag{8}$$

式中，d_{out}表示空间聚合子块输出的特征向量维度。本文根据经验分别设定为[32，128，256，512，1 024]（Qiu等，2021）。

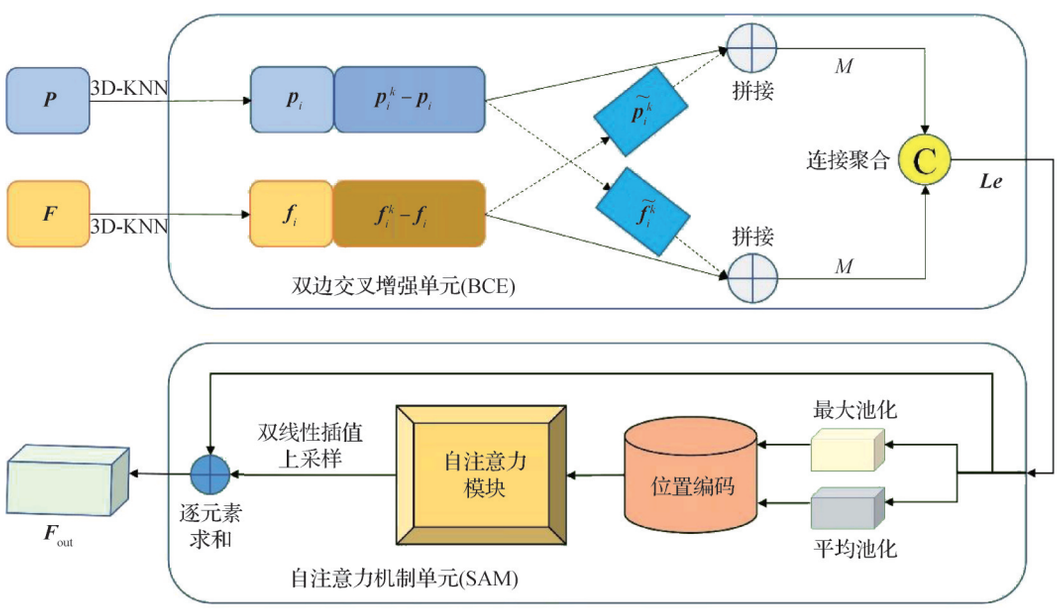


图2　空间聚合子块

Fig. 2　Spatial aggregation subblock

1.2.2　自注意力机制单元

为了增强特征完备性，构建自注意力机制单元，提取点云的全局上下文信息，并与局部上下文信息进行融合，补偿局部上下文信息的单一性。

对于每个中心点提取的局部上下文信息$\boldsymbol{Le} \in \mathbf{R}^{K\times d_{\text{out}}}$，首先，在特征维度上对$K$个近邻点进行局部平均和最大池化下采样，得到两个形状为$K\times\frac{1}{2}d_{\text{out}}$的特征向量并进行拼接，形成$K\times d_{\text{out}}$的特征向量。然后，对拼接后的特征向量嵌入维度为$d_{\text{out}}$的位置编码，进而得到形状为$K\times 2d_{\text{out}}$的特征向量，用于捕捉不同近邻点之间的空间位置关系。最后，构建自注意力模块从$K\times 2d_{\text{out}}$特征向量中提取全局上下文信息，具体如下：

首先，将嵌有位置编码的特征向量表示为$\boldsymbol{Z} \in \mathbf{R}^{K\times 2d_{\text{out}}}$，通过线性变换将$\boldsymbol{Z}$映射到查询向量$\boldsymbol{Q}_I$、关键向量$\boldsymbol{K}_I$和值向量$\boldsymbol{V}_I$，并通过点积计算$\boldsymbol{Q}_I$与$\boldsymbol{K}_I$之间的相似度，得到每个点的注意力权重$\Lambda$，具体为

$$\boldsymbol{Q}_I = \boldsymbol{Z}\times\boldsymbol{W}_q,\ \boldsymbol{K}_I = \boldsymbol{Z}\times\boldsymbol{W}_k,\ \boldsymbol{V}_I = \boldsymbol{Z}\times\boldsymbol{W}_v \tag{9}$$

$$\Lambda = f_{\text{softmax}}\left(\frac{\boldsymbol{Q}_I\cdot\boldsymbol{K}_I^{\mathrm{T}}}{\sqrt{D_k}}\right) \tag{10}$$

式中，$\boldsymbol{W}_q \in \mathbf{R}^{2d_{\text{out}}\times D_q}$，$\boldsymbol{W}_k \in \mathbf{R}^{2d_{\text{out}}\times D_k}$，$\boldsymbol{W}_v \in \mathbf{R}^{2d_{\text{out}}\times D_v}$为权重矩阵。$\sqrt{D_k}$用于缩放注意力权重，保持训练过程中梯度值稳定。

然后，使用Λ对$\boldsymbol{V}_l$进行加权求和，得到每个点的自注意力聚合结果$\boldsymbol{SAG}$，并对其进行双线性插值上采样，得到与$\boldsymbol{Le}$相同的维度，具体为

$$\boldsymbol{SAG}=\Lambda\cdot\boldsymbol{V}_l,\boldsymbol{SAG}\in\mathbf{R}^{K\times D_v} \tag{11}$$

$$\boldsymbol{Gci}=\boldsymbol{SAG}\times\boldsymbol{W}_O,\boldsymbol{Gci}\in\mathbf{R}^{K\times d_{out}} \tag{12}$$

式中，$\boldsymbol{W}_O\in\mathbf{R}^{D_v\times d_{out}}$表示输出映射的权重矩阵。$\boldsymbol{Gci}$即为提取的全局上下文信息，表示$K$个近邻点在全局上下文中的特征向量。

最后，将$\boldsymbol{Gci}$与$\boldsymbol{Le}$合并，实现局部与全局上下文信息的融合，具体为

$$\boldsymbol{Mix}=\boldsymbol{Le}+\boldsymbol{Gci},\boldsymbol{Mix}\in\mathbf{R}^{K\times d_{out}} \tag{13}$$

式中，$\boldsymbol{Mix}$表示融合后的完备特征信息。由于中心点组$\boldsymbol{Y}$有n个中心点，从而得到由n个$\boldsymbol{Mix}$组成的特征图。完成5个空间聚合子块编码后，得到多分辨率特征图$\boldsymbol{F}_{out}$。

1.2.3　损失函数

在整个编码模块中，引入损失函数以调节双边交叉增强的学习过程。为了充分学习点云的局部特征信息，将近邻区域作为一个整体考虑，而不是考虑单个近邻点。首先，添加偏移量以增强局部几何上下文信息并保持密集邻域的几何完整性。然后，通过最小化l_2距离鼓励近邻区域的几何中心点靠近局部区域的中心点位置，具体为

$$L(\boldsymbol{p}_i)=\left\|\frac{1}{K}\sum_{k=1}^{K}(\tilde{\boldsymbol{p}}_i^k-\boldsymbol{p}_i)\right\|_2 \tag{14}$$

1.3　多尺度特征融合模块

为了提高算法对不同尺度物体的感知能力，通过多尺度特征融合模块整合多分辨率特征图，具体流程如图3所示。

对于$\boldsymbol{F}_{out}$，设$\{\boldsymbol{F}_1,\boldsymbol{F}_2,\cdots,\boldsymbol{F}_t,\cdots,\boldsymbol{F}_T\}$表示所有空间聚合子块输出的特征图集合，$\boldsymbol{F}_t$表示第$t$个子块的输出，$T$表示空间聚合子块数量，本文$T=5$。对每个特征图首先进行上采样，然后将上采样后的特征图与上一层的特征图进行融合，最后得到不同尺度的特征图，记为$\{\widetilde{\boldsymbol{F}}_1,\widetilde{\boldsymbol{F}}_2,\cdots,\widetilde{\boldsymbol{F}}_t,\cdots,\widetilde{\boldsymbol{F}}_T\}$。

为了整合不同尺度的特征图和提炼有用的上下文信息，进一步将$\{\widetilde{\boldsymbol{F}}_1,\widetilde{\boldsymbol{F}}_2,\cdots,\widetilde{\boldsymbol{F}}_t,\cdots,\widetilde{\boldsymbol{F}}_T\}$输入至$M$获得点级信息，并对点级信息归一化，使用softmax得到$\{\phi_1,\phi_2,\cdots,\phi_t,\cdots,\phi_T\}$。然后融合归一化的点级信息和不同尺度特征图，得到用于语义分割的综合特征图$\widetilde{\boldsymbol{F}}_{out}$。最后通过全连接层实现点云的语义分割，具体为

$$\widetilde{\boldsymbol{F}}_{out}=\sum_{t=1}^{T}\phi_t\times\widetilde{\boldsymbol{F}}_t \tag{15}$$

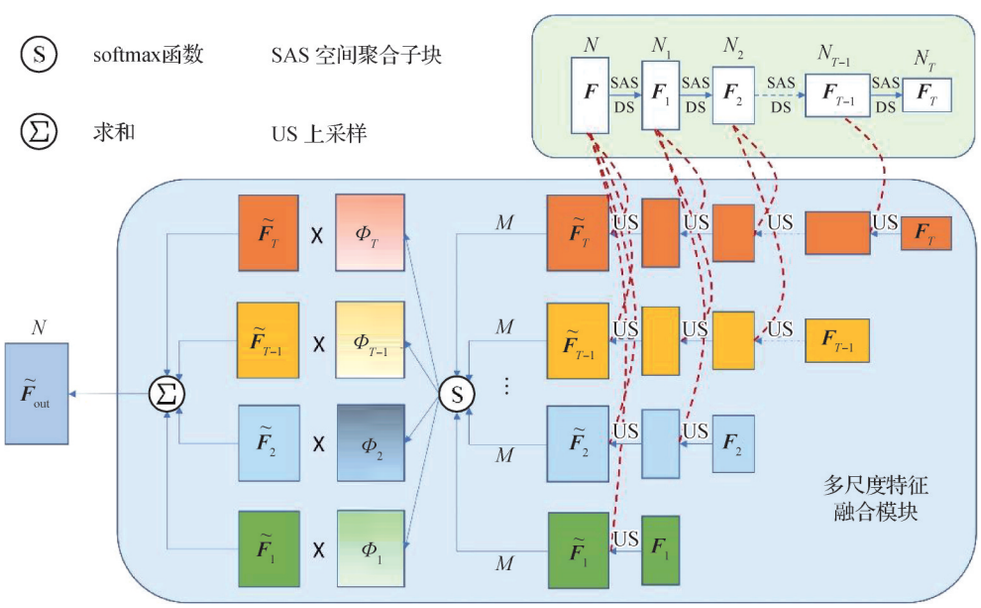


图3　多尺度特征融合模块

Fig. 3　Multi-scale feature fusion module

2　实验结果与分析

为了评估所提算法的性能，在公共数据集S3DIS(Stanford 3D indoor spaces dataset)(Armeni等，2017)和Semantic3D(Hackel等，2017)上进行了实验分析。基于Ubuntu 16.04操作系统，采用Python与Tensorflow实验平台进行算法实现。在单个NVIDIA

TITAN RTX GPU上训练100轮，使用Adam优化算法以端到端的方式进行训练，批量大小设为16，学习率从0.01开始，每10轮后以0.5的速率衰减。

2.1 数据集及评价指标

S3DIS是室内点云数据集，标注了13个语义类别，分为6个大型室内区域，包含271个房间，每个房间都由密集点云组成，每个点包含三维坐标信息和RGB信息。Semantic3D是室外点云数据集，标注了8个语义类别，总计超过40亿个点，每个点都包含3维坐标、RGB颜色和强度值信息。

使用平均交并比（mean intersection over union，mIoU）作为算法性能主要评估指标，mIoU是数据集所有语义类别IoU的平均值。总体精度（overall accuracy，OA）和平均类别精度（mean class accuracy，mAcc）作为参考评估指标。

2.2 S3DIS数据集上的室内场景语义分割

2.2.1 评价指标对比

表1和表2给出了S3DIS数据集上的对比实验结果，其中，RandLA-Net采用重新训练的实验数据。

由表1的6折交叉验证可以看出本文算法在mIoU、mAcc、OA这3个指标上都取得了很好的性能，分别达到70.2%、81.7%和88.3%，较RandLA-Net分别提升了2.4%、2.0%和1.0%。同时，相比于最新算法PASIFTNet，本文算法在mIoU和OA两个指标上分别提升了1.9%和1.8%。

S3DIS数据集Area 5的点云场景较为复杂，对其精细高效分割更具有挑战性。表2给出了对Area 5的独立测试结果，可以看出本文算法对于复杂场景有着较好的分割精度，mIoU达到了66.2%，较RandLA-Net算法提升了5.0%。相比于其他最新算法，如LGFF-Net（local-global feature fusing）、LG-Net（local dual features and global correlations）、GeoSeg-Net（geometric encoder-decoder model）等，mIoU和OA也有不同程度的提升。

2.2.2 可视化分析

图4给出了S3DIS数据集部分场景的可视化结果。可以看出，即使在复杂场景中，例如办公室、会议室以及走廊，本文算法相较于RandLA-Net的分割结果更准确，特别是对于墙壁、黑板和杂物等类别。从场景1中可以明显看出，本文算法能准确分割墙壁以及横梁部分，但RandLA-Net将部分横梁错误地分割成了墙壁。通过测试，RandLA-Net对墙壁与横

表1 S3DIS数据集6折交叉验证结果

Table 1 The 6-fold cross-validation results on S3DIS dataset

/%

方法	OA	mAcc	mIoU
PointNet（Qi等，2017a）	78.6	66.2	47.6
3P-RNN（Ye等，2018）	86.9	–	56.3
RSNet（Huang等，2018）	–	66.5	56.5
SPG（Landrieu和Simonovsky，2018）	86.4	73	62.1
PointCNN（Li等，2018）	88.1	75.6	65.4
PointWeb（Zhao等，2019）	87.3	76.2	66.7
Fuzzy3DSeg（Zhong等，2020）	86.3	–	56.4
MNFEAM（Li等，2021）	85	–	56.8
DPFA-Net（Chen等，2022）	**89**	–	61.6
PASIFTNet（Wang等，2023）	86.5	–	68.3
RandLA-Net*（Hu等，2022）	87.3	79.7	67.8
本文	88.3	**81.7**	**70.2**

注：“*”表示该方法重新训练的结果，加粗字体表示各列最优结果，“–”表示原文中未给出具体值。3P-RNN：recurrent neural network with pointwise pyramid pooling；SPG：superpoint graph；Fuzzy3DSeg：fuzzy neighborhood learning for 3D segmentation；DPFA-Net：dynamic point feature aggregation network；PASIFTNet：point-atrous sift network。

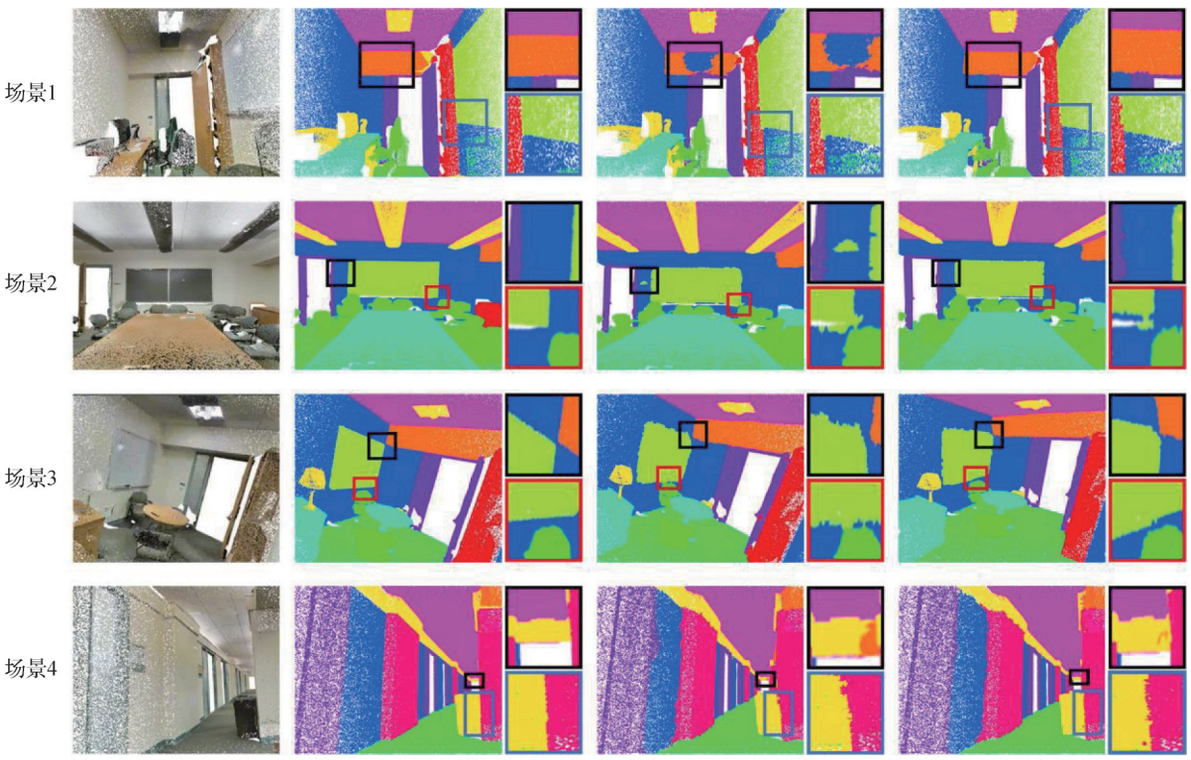
表2 S3DIS数据集Area 5测试结果

Table 2 Test results of Area 5 on S3DIS dataset

/%

方法	OA	mIoU
PointNet++（Qi等，2017b）	–	50
CSANet（Wang等，2022）	–	53.3
DPFA-Net（Chen等，2022）	88	55.2
PointCNN（Li等，2018）	85.9	–
PointWeb（Zhao等，2019）	87	60.3
PointASNL（Yan等，2020）	87.7	62.6
GACNet（Wang等，2019）	87.8	62.9
GA-Net（Deng和Dong，2021）	87.6	63.7
LGFF-Net（Bi等，2023）	87	63.9
LG-Net（Zhao等，2023）	88.1	64.9
GeoSegNet（Chen等，2023）	88.5	64.9
RandLA-Net*（Hu等，2022）	87.4	61.2
本文	**88.6**	**66.2**

注：“*”表示该方法重新训练的结果，加粗字体表示各列最优结果，“–”表示原文中未给出具体值。PointASNL：nonlocal neural network with adaptive sampling；GACNet：graph attention convolution network；GA-Net：global attention network。



(a) 原始点云　(b) 真实标签　(c) RandLA-Net　(d) 本文

图4　语义分割结果在S3DIS数据集上的可视化

Fig. 4　Visualization of semantic segmentation results on S3DIS dataset

((a)raw point cloud;(b)ground truth;(c)RandLA-Net;(d)ours)

梁的 mIoU 分别为 80. 02%、69. 77%，而本文算法分别为 82. 60%、71. 89%。场景 2—场景 4 给出了物体边界处的分割细节，相较于 RandLA-Net，本文算法表现出更好的分割效果，局部细节分割更精确。

2. 2. 3　消融实验

消融实验在 S3DIS 数据集上进行，使用 6 折交叉验证进行评估，结果如表 3 所示。其中，方法 1 是基础模型，仅使用几何信息作为网络输入，未使用语义信息和双边交叉增强、自注意力机制和多尺度特征融合，采用最大池化。方法 2—方法 4 在方法 1 的基础上分别测试了双边交叉增强、多尺度特征融合和自注意力机制的性能。实验结果显示，双边交叉增强的性能提升最多，也验证了本文算法双边交叉增强单元的有效性。在双边交叉增强单元中，通过构建偏移量对局部几何与语义上下文信息进行增强，可以增强点云局部上下文信息，有效提升了点云分割精度。

方法 5 将最大池化替换成最大池化与平均池化相结合的混合池化。结果显示，mIoU 相比于方法 1 略有提升。方法 6 在方法 5 的基础上使用了双边交叉增强，在几何信息输入的同时增加了语义信息，并通过双边交叉学习增强点云的几何和语义信息。结果显示，相比于方法 5 性能有明显提升。方法 7 每次上采样都结合上一层特征图，最后融合得到多尺度特征。实验结果显示，相比于方法 6，mIoU 提升了 2. 4%。方法 8 在此基础上进一步增加了自注意力

表 3　消融实验结果

Table 3　Results of ablation experiments

方法	池化方式	双边交叉增强	多尺度特征融合	自注意力机制	mIoU/%
1	最大池化	×	×	×	55.6
2	最大池化	√	×	×	65.8
3	最大池化	×	√	×	62.6
4	最大池化	×	×	√	59.4
5	混合池化	×	×	×	57.4
6	混合池化	√	×	×	66.3
7	混合池化	√	√	×	68.7
8	混合池化	√	√	√	**70.2**

注：加粗字体表示最优结果。“√”和“×”分别表示使用和未使用该模块。

模块，是本文的完整网络架构，通过全局上下文信息补偿局部上下文信息，进一步增强特征的完备性，得到了最好的分割效果。

2.3 Semantic3D数据集上的室外场景语义分割

为了进一步分析所提算法对室外场景点云分割的可行性，在Semantic3D数据集上进行了实验分析。表4给出了本文算法与RandLA-Net方法的对比训练结果。可以看出，本文算法在mIoU、OA两个指标上分别提升了1.3%和1.2%。测试实验采用具有在线评估的reduced-8测试集，针对4个特定场景，使用0.01 m的均匀下采样进行评估。

可视化结果如图5所示，其中图5(a)表示原始点云，图5(b)(c)分别表示RandLA-Net和本文算法的分割效果。可以看出，本文算法可以较好地提取室外场景1—场景3中的地面、建筑等地图要素以及场景4中的草地和道路。总之，在Semantic3D数据集上的训练与测试可视化结果都显示本文算法可以有效用于室外场景点云分割。

表4 Semantic3D数据集训练结果

Table 4 Training results on Semantic3D dataset

/%

方法	OA	mIoU
RandLA-Net*（Hu等，2022）	89.3	68.1
本文	**90.5**	**69.4**

注：“*”表示该方法重新训练的结果，加粗字体表示各列最优结果。

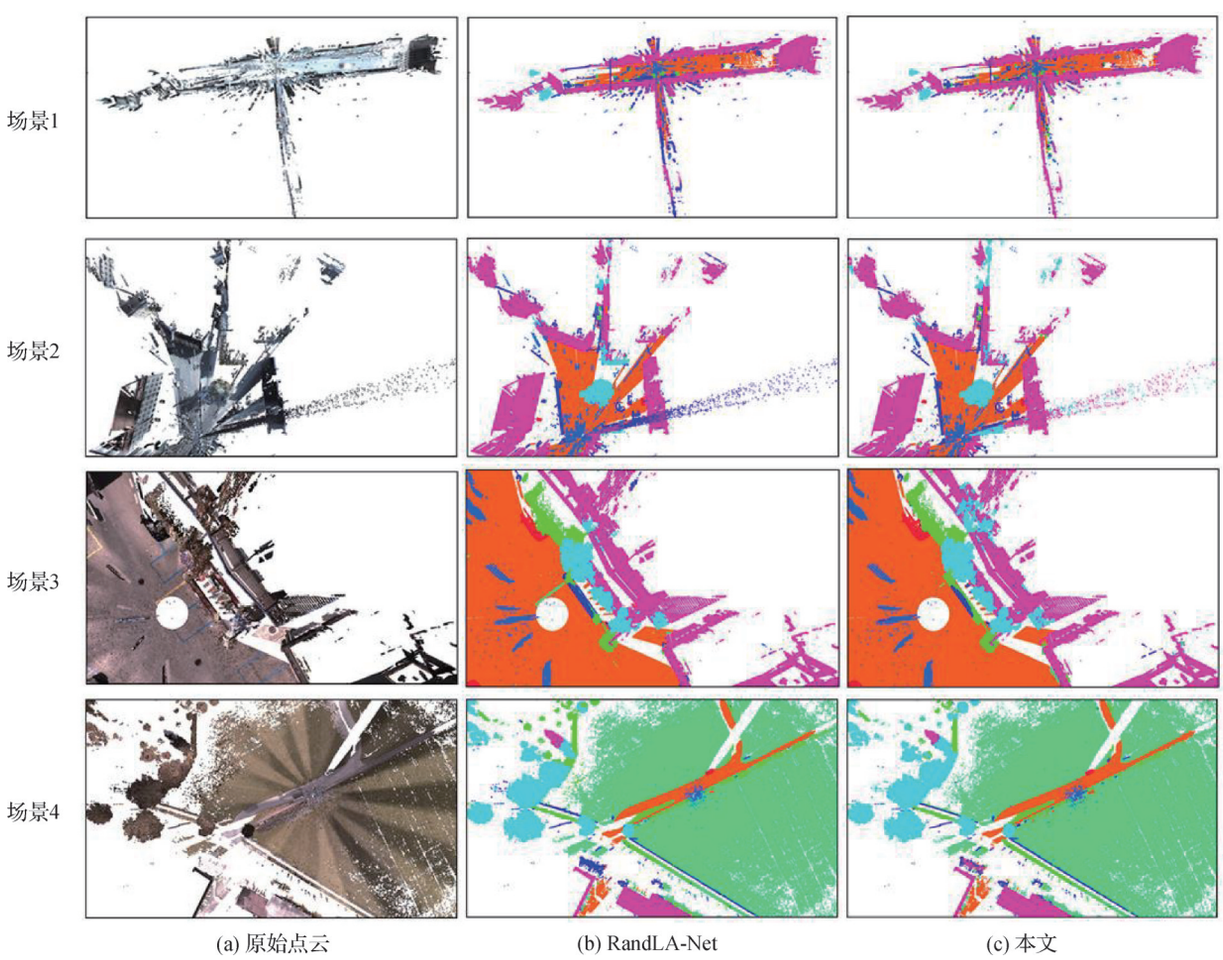


图5 语义分割结果在Semantic3D数据集上的可视化

Fig. 5 Visualization of semantic segmentation results on Semantic3D dataset ((a) raw point cloud; (b) RandLA-Net; (c) ours)

3 结 论

本文提出了一种结合双边交叉增强与自注意力补偿的点云语义分割新算法，通过局部几何与语义上下文信息的交互增强、局部和全局特征的相互补偿，提升点云细粒度分割精度。首先，对输入原始点云数据进行预处理，提取几何与初始语义上下文信

息。然后，在空间聚合模块中构建局部邻域，利用偏移量分别增强局部几何与语义上下文信息。在此基础上基于自注意力机制提取全局上下文信息，进一步对特征信息完善与补充，增强特征的完备性以及局部与全局的关联性。最后，将不同分辨率的特征图输出至多尺度特征融合模块，进行上采样和融合，最终通过全连接层实现点云的语义分割。在S3DIS和Semantic3D数据集上的实验结果表明，相较于对比算法，本文算法在多项指标上表现出更优性能。但由于点云数据的无序性、稀疏性和数据量巨大等特点，在实际应用中，精准、高效的点云语义分割仍面临巨大挑战。为此，下一步拟在努力提升实验平台和算力资源的基础上，进一步深入研究更精准、更高效的深度学习模型与算法，提升点云语义分割的性能。

参考文献(References)

Alonso I, Riazuelo L, Montesano L and Murillo A C. 2020. 3D-mininet: learning a 2D representation from point clouds for fast and efficient 3D LIDAR semantic segmentation. IEEE Robotics and Automation Letters, 5(4): 5432-5439 [DOI: 10.1109/LRA.2020.3007440]

Armeni I, Sax S, Zamir A R and Savarese S. 2017. Joint 2D-3D-semantic data for indoor scene understanding [EB/OL]. [2023-06-25]. https://arxiv.org/pdf/1702.01105.pdf

Bi Y W, Zhang L J, Liu Y W, Huang Y S and Liu H. 2023. A local-global feature fusing method for point clouds semantic segmentation. IEEE Access, 11: 68776-68790 [DOI: 10.1109/ACCESS.2023.3293161]

Boulch A, Guerry J, Le Saux B and Audebert N. 2018. SnapNet: 3D point cloud semantic labeling with 2D deep segmentation networks. Computers and Graphics, 71: 189-198 [DOI: 10.1016/j.cag.2017.11.010]

Chen C, Wang Y S, Chen H H, Yan X F, Ren D Y, Guo Y W, Xie H R, Wang F L and Wei M Q. 2023. GeoSegNet: point cloud semantic segmentation via geometric encoder-decoder modeling. The Visual Computer, 40: 5107-5121 [DOI: 10.1007/s00371-023-02853-7]

Chen J J, Kakillioglu B and Velipasalar S. 2022. Background-aware 3-D point cloud segmentation with dynamic point feature aggregation. IEEE Transactions on Geoscience and Remote Sensing, 60: #5703112 [DOI: 10.1109/TGRS.2022.3168555]

Deng S and Dong Q L. 2021. GA-NET: global attention network for point cloud semantic segmentation. IEEE Signal Processing Letters, 28: 1300-1304 [DOI: 10.1109/LSP.2021.3082851]

Du J and Cai G R. 2021. Point cloud semantic segmentation method based on multi-feature fusion and residual optimization. Journal of Image and Graphics, 26(5): 1105-1116 (杜静, 蔡国榕. 2021. 多特征融合与残差优化的点云语义分割方法. 中国图象图形学报, 26(5): 1105-1116) [DOI: 10.11834/jig.200374]

Giang T T H and Ryoo Y J. 2023. Pruning points detection of sweet pepper plants using 3D point clouds and semantic segmentation neural network. Sensors, 23(8): #4040 [DOI: 10.3390/s23084040]

Hackel T, Savinov N, Ladicky L, Wegner J D, Schindler K and Pollefeys M. 2017. Semantic3D. net: a new large-scale point cloud classification benchmark [EB/OL]. [2023-06-25]. https://arxiv.org/pdf/1704.03847.pdf

Hu Q Y, Yang B, Xie L H, Rosa S, Guo Y L, Wang Z H, Trigoni N and Markham A. 2022. Learning semantic segmentation of large-scale point clouds with random sampling. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(11): 8338-8354 [DOI: 10.1109/TPAMI.2021.3083288]

Huang Q G, Wang W Y and Neumann U. 2018. Recurrent slice networks for 3D segmentation of point clouds//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 2626-2635 [DOI: 10.1109/CVPR.2018.00278]

Landrieu L and Simonovsky M. 2018. Large-scale point cloud semantic segmentation with superpoint graphs//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 4558-4567 [DOI: 10.1109/CVPR.2018.00479]

Li D W, Shi G L, Wu Y H, Yang Y P and Zhao M B. 2021. Multi-scale neighborhood feature extraction and aggregation for point cloud segmentation. IEEE Transactions on Circuits Systems for Video Technology, 31(6): 2175-2191 [DOI: 10.1109/TCSVT.2020.3023051]

Li Y Y, Bu R, Sun M C, Wu W, Di X H and Chen B Q. 2018. PointCNN: convolution on *X*-transformed points//Proceedings of the 32nd Conference on Neural Information Processing Systems. Montréal, Canada: Curran Associates Inc.: 828-838 [DOI: 10.5555/3326943.3327020]

Liu S, Cao Y F, Huang W H and Li D D. 2023. LiDAR point cloud semantic segmentation combined with sparse attention and instance enhancement. Journal of Image and Graphics, 28(2): 483-494 (刘盛, 曹益烽, 黄文豪, 李丁达. 2023. 融合稀疏注意力和实例增强的雷达点云分割. 中国图象图形学报, 28(2): 483-494) [DOI: 10.11834/jig.210787]

Luo H F, Chen C C, Fang L N, Khoshelham K and Shen G X. 2020. MS-RRFSegNet: multiscale regional relation feature segmentation network for semantic segmentation of urban scene point clouds. IEEE Transactions on Geoscience and Remote Sensing, 58(12): 8301-8315 [DOI: 10.1109/TGRS.2020.2985695]

Poux F and Billen R. 2019. Voxel-based 3D point cloud semantic segmentation: unsupervised geometric and relationship featuring vs deep learning methods. ISPRS International Journal of Geo-

Information, 8(5): #213 [DOI: 10.3390/IJGI8050213]

Qi C R, Su H, Mo K C and Guibas L J. 2017a. PointNet: deep learning on point sets for 3D classification and segmentation//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 652-660 [DOI: 10.1109/CVPR.2017.16]

Qi C R, Yi L, Su H and Guibas L J. 2017b. PointNet++: deep hierarchical feature learning on point sets in a metric space//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc.: 5105-5114 [DOI: 10.5555/3295222.3295263]

Qiu S, Anwar S and Barnes N. 2021. Semantic segmentation for real point cloud scenes via bilateral augmentation and adaptive fusion//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 1757-1767 [DOI: 10.1109/CVPR46437.2021.00180]

Ren D Y, Wu Z Y, Li J W, Yu P P, Guo J, Wei M Q and Guo Y W. 2022. Point attention network for point cloud semantic segmentation. Science China Information Sciences, 65(9): #192104 [DOI: 10.1007/s11432-021-3387-7]

Wang G H, Zhai Q Y and Liu H. 2022. Cross self-attention network for 3D point cloud. Knowledge-Based Systems, 247: #108769 [DOI: 10.1016/j.knosys.2022.108769]

Wang L, Huang Y C, Hou Y L, Zhang S M and Shan J. 2019. Graph attention convolution for point cloud semantic segmentation//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 10296-10305 [DOI: 10.1109/CVPR.2019.01054]

Wang S F, Liu Y, Wang L C, Sun Y F and Yin B C. 2023. PASIFTNet: scale-and-directional-aware semantic segmentation of point clouds. Computer-Aided Design, 156: #103462 [DOI: 10.1016/j. cad.2022.103462]

Yan X, Zheng C D, Li Z, Wang S and Cui S G. 2020. PointASNL: robust point clouds processing using nonlocal neural networks with adaptive sampling//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 5588-5597 [DOI: 10.1109/CVPR42600.2020.00563]

Ye X Q, Li J M, Huang H X, Du L and Zhang X L. 2018. 3D recurrent neural networks with context fusion for point cloud semantic segmentation//Proceedings of the 15th European Conference on Computer Vision-ECCV 2018. Munich, Germany: Springer: 415-430 [DOI: 10.1007/978-3-030-01234-2_25]

Zhao H S, Jiang L, Fu C W and Jia J Y. 2019. PointWeb: enhancing local neighborhood features for point cloud processing//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 5565-5573 [DOI: 10.1109/CVPR.2019.00571]

Zhao Y Q, Ma X Y, Hu B, Zhang Q, Ye M and Zhou G Q. 2023. A large-scale point cloud semantic segmentation network via local dual features and global correlations. Computers and Graphics, 111: 133-144 [DOI: 10.1016/j.cag.2023.01.011]

Zheng Y, Lin C Y, Liao K, Zhao Y and Xue S. 2021. LiDAR point cloud segmentation through scene viewpoint offset. Journal of Image and Graphics, 26(10): 2514-2523 (郑阳, 林春雨, 廖康, 赵耀, 薛松. 2021. 场景视点偏移的激光雷达点云分割. 中国图象图形学报, 26(10): 2514-2523) [DOI: 10.11834/jig.200424]

Zhong M Y, Li C J, Liu L C, Wen J H, Ma J W and Yu X H. 2020. Fuzzy neighborhood learning for deep 3-D segmentation of point cloud. IEEE Transactions on Fuzzy Systems, 28(12): 3181-3192 [DOI: 10.1109/TFUZZ.2020.2992611]

作者简介

朱仲杰,通信作者,男,教授,主要研究方向为2D和3D视频编码与传输、高动态图像视频编码及处理。
E-mail:zhongjiezhu@yeah.net
张荣,男,硕士研究生,主要研究方向为计算机视觉与点云语义分割。E-mail:rongzhang0105@163.com
白永强,男,副教授,主要研究方向为信息隐藏、图像处理和人工智能。E-mail:byq-163@163.com
王玉儿,女,助理研究员,主要研究方向为视频编码、信息隐藏。E-mail:365401628@qq.com
孙嘉敏,女,硕士研究生,主要研究方向为高动态图像视频处理及编码研究。E-mail:jiaminsun0717@163.com