文章编号:1001-5078(2023)11-1785-08

·图像与信号处理·

融合注意力门控机制的大场景点云语义分割

王　蕾[1,2],朱芬芬[1],李金萍[1],刘　华[3]
(1. 东华理工大学 信息工程学院,江西 南昌 330013;2. 江西省放射性地学大数据技术工程实验室,东华理工大学,江西 南昌 330013;
3. 东华理工大学测绘工程学院,江西 南昌 330013)

摘　要:室外大场景激光点云语义分割已成为 3D 场景理解、环境感知的关键性技术,在自动驾驶、智能机器人和增强现实(AR)等领域应用广泛。然而大场景的激光点云具有多目标、几何结构复杂,不同地物尺度变化大等特点,使得在稀疏的小目标点云(例如行人、摩托车等)上的分割性能较低。针对上述问题,本文提出一种融合注意力门控机制的室外点云语义分割算法,设计由注意力机制和多尺度上下文特征融合组成的注意力门控单元,提高对激光点云细粒度特征的表达,降低随机降采样过程中点云几何结构特征丢失程度,从而增强了网络对弱小目标的特征获取能力;同时设计基于共享 MLP 的平均池化单元,进一步简化自注意力局部特征聚合模块,有效地加速网络收敛,能高效地实现大场景点云的语义分割。本文方法在自动驾驶场景室外激光点云数据集 SemanticKITTI 上的实验表明,与文献 RandLA-Net 相比,收敛速度提升 48.3 %,平均交并比(mIoU)由 53.9 %提升至 54.5 %,提高 0.6 %,尤其是在小目标上交并比(IoU)均有明显提高,person 类和 motorcycle 类的交并比分别提高 0.8 %和 5.4 %。

关键词:大场景激光点云;语义分割;随机降采样;平均池化单元;注意力门控单元;多尺度特征融合
中图分类号:TN249;TP391.41　　**文献标识码:**A　　**DOI:**10.3969/j.issn.1001-5078.2023.11.024

The semantic segmentation algorithm for large scene point cloud based on attention gating mechanism

WANG Lei[1,2], ZHU Fen-fen[1], LI Jin-Ping[1], LIU Hua[3]
(1. School of Information Engineering, East China University of Technology, Nanchang 330013, China; 2. Jiangxi Engineering Laboratory on Radioactive Geoscience and Big Data Technology, East China University of Technology, Nanchang 330013, China;
3. School of Surveying and Mapping Engineering, East China University of Technology, Nanchang 330013, China)

Abstract: Semantic segmentation for outdoor large-scale point cloud has become a key technology for 3D scene understanding and environmental awareness and is widely used in fields such as autonomic driving, intelligent robotic and augmented reality (AR). However, laser point clouds of large scenes are characterized by multi-targets, complex geometrical structures, and large variations in the scales of different features, making the segmentation performance on sparse point clouds of small targets (e.g., pedestrians, motorcycles, etc.) low. To address the above problems, an outdoor point cloud semantic segmentation algorithm incorporating an attentive gating mechanism is proposed in this paper. An attentive Gating Unit based on attention mechanism and multi-scale feature fusion method is designed to improve the expression of fine-grained features of laser point clouds and significantly reduce the information loss during

基金项目:江西省核地学数据科学与系统工程技术研究中心基金项目(No. JELRGBDT202202);江西省放射性地学大数据技术工程实验室开放基金项目(No. JELRGBDT202103);江西省自然科学基金项目(No. 20202BABL212014);东华理工大学江西省数字国土重点实验室开放研究基金项目(No. DLLJ202004);国家自然科学基金项目(No. 42001411)资助。
作者简介:王　蕾(1979 -),女,博士,教授,主要研究方向为计算机图形图像,计算机视觉技术。E-mail:wangl@ecut.edu.cn
通讯作者:朱芬芬(1996 -),女,硕士研究生,主要研究方向为计算机三维视觉。E-mail:2020110184@ecut.edu.cn
收稿日期:2023-01-05;**修订日期:**2023-02-24

the random downsampling process, thus enhancing the feature extraction performance for weak targets. At the same time, anaverage pooling unit based on shared MLP is designed to further simplify the self-attention local feature aggregation module, which effectively accelerates the network convergence speed and can efficiently realize the semantic segmentation of point clouds in large scenes. The experiments on outdoor driving dataset semanticKITTI show that the convergence speed is increased by 48.3 %, and the mean intersection-over-Union (mIoU) of all classes is improved-from 53.9 % to 54.5 %, an increase of 0.6 %, compared with the literature RandLA-Net. Especially, the Intersection-over-Union (IoU) of small-scale class is significantlyimproved, for example, the IoU score of person and motorcycle are increased by 0.8 % and 5.4 %, respectively.

Keywords: large-scale point cloud; semantic segmentation; random sampling; average pooling unit; attentive gating unit; multi-scale feature fusion

1 引 言

随着激光雷达、RGB－D 相机等 3D 传感器技术的迅速发展,激光点云数据作为基础的 3D 数据表达,包含真实世界丰富的信息,受到越来越多的关注。面向激光点云语义的高效分割可以更好地自动理解场景,已成为解决 3D 场景理解、环境感知的关键性技术,并在智能驾驶,机器人视觉等领域中发挥着关键的作用。

随着深度学习技术的兴起,利用数据驱动的方式对点云处理取得较好成果,通常可分为三类:基于投影的方法,基于体素的方法和基于点的方法。CHEN[1] 和 MILIOTO[2] 把点云投影成多视角的二维图像,使用二维卷积神经网络对图像进行处理,图像分割结果被反投影回三维点云上,实现对三维激光点云的间接处理。MENG[3] 和 RIEGLER[4] 将三维点云体素化到稠密的三维网格,由体素网格上二进制变量的概率分布表示,然后使用三维卷积等规则化数据处理方法。以上方法解决了点云数据非结构化的问题,但在投影或体素化的过程中容易损失原始点云的几何信息。PointNet[5] 为代表的直接处理点云数据方法,通过输入原始点云的几何坐标和 RGB 特征,用共享的多层感知机(MLP)独立地学习每个点的特征,然而这种方法使得点与点之间的局部关系表达不够。刘[6] 提出在利用点云三维坐标信息的基础上,增加了点云 RGB 信息和归一化坐标信息,进一步提高了模型的分割精度。AC-Net[7] 提出图注意力卷积自适应地学习局部区域特征,能够有效捕获目标形状和几何模式,但不能直接处理大场景点云(覆盖 200 m × 200 m 的场景,包含百万甚至上亿个点)。

近年来,研究者们提出了许多面向室外大场景点云的深度学习算法。MVP-Net[8] 提出一种新颖的点排序方法和多次旋转输入点云,实现多视角点云局部特征聚合和感受野扩张。RandLA-Net[9] 是直接处理点云的先进标准模型,采用基于注意力的点云局部特征聚合模块和随机降采样方法。MSAA-Net[10] 基于 RandLA-Net,在编码与解码层的特征跳层连接处中增加了注意力机制,并从编码层和解码层中捕获点云的全局特征。然而上述方法在小尺度目标上的分割精度较低。

本文提出改进的大场景点云语义分割算法 gRandLA-Net,主干网络基于 RandLA-Net[9],首先,设计注意力门控单元,利用自注意力机制自适应地学习点云局部几何特征,同时利用多尺度局部特征融合将不同尺度邻域的点云特征相加,增强模块的几何特征表达能力,有利于网络学习细粒度的点云特征;其次,受 pointMixer[11] 的启发,设计平均池化单元,仅利用共享多层感知机(MLP)学习局部点云特征,计算简单,使得网络更容易收敛。本文方法在保证高效架构的同时,训练速度提高近一倍,分割更加准确,尤其是对小尺度目标的分割精度有明显提高。

2 本文方法

面对稀疏的室外大场景点云,本文方法 gRandLA-Net 采用随机降采样(Random Sampling, RS)策略逐层减小点云,以提高计算效率,设计平均池化单元和注意力门控单元为局部特征聚合模块(Local Feature Aggregation, LFA),融合多尺度领域点云局部特征,并逐层扩大每个点的感受野,以增强网络对复杂点云模式的感知能力,如图 1 所示。

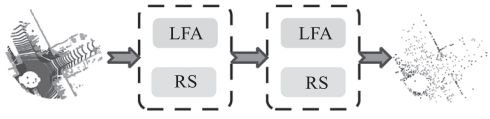

图 1 局部特征聚合与随机下采样模块示意图

Fig. 1 The illustration of local feature aggregation and random sampling

2.1 平均池化单元(Average Pooling Unit)

该模块输入点云形式为 $(N \times (3 + d_{in}))$,输出点云形式为 $(N \times d_{out})$,N 是点云中点的数量,d_{in} 是输入点云的特征通道数,d_{out} 是输出局部特征的通

道数，如图 2 所示。点云 $P = \{p_1 \cdots p_i \cdots p_n\}$，首先由局部空间编码模块（Local spatial encoding，LocSE）学习点云的局部空间几何特征，用于增强输入点云的其他特征，经过增强后的点云语义特征输入均值池化模块（Average Pooling），聚合每个点的局部特征，得到更细粒度的语义特征 $\hat{F} = \{\hat{f}_1 \cdots \hat{f}_i \cdots \hat{f}_n\}$。

2.1.1 局部空间编码（Local spatial encoding，LocSE）

利用 KNN 算法得到第 i 个中心点 p_i 的邻域点三维坐标 $\{p_i^1 \cdots p_i^k \cdots p_i^K\} \subset \mathbb{R}^3$ 及特征 $\{f_i^1 \cdots f_i^k \cdots f_i^K\} \subset \mathbb{R}^{d_{\text{in}}}$，由局部空间编码模块对点云的空间位置关系进行编码，并用编码的空间特征增强其他语义特征，输出邻域特征 $\{\tilde{f}_i^1 \cdots \tilde{f}_i^k \cdots \tilde{f}_i^K\}, \tilde{f}_i^k \in \mathbb{R}^{2d_{\text{in}}}$。

编码空间位置关系：

$$r_i^k = \text{mlp}((p_i \oplus p_i^k \oplus (p_i - p_i^k) \oplus \|p_i - p_i^k\|), W)\#(\text{AUTONUM}\backslash * \text{Arabic})$$

增强语义特征：

$$\tilde{f}_i^k = r_i^k \oplus f_i^k \#(AUTONUM\backslash * Arabic)$$

其中，$(p_i - p_i^k)$ 是邻域点与中心点的相对坐标；$\|p_i - p_i^k\|$ 是邻域点与中心点的欧几里德距离；W 是共享多层感知机的可学习参数；⊕是特征连接操作；$r_i^k, f_i^k \in \mathbb{R}^{d_{\text{in}}}$。

2.1.2 均值池化模块（Average Pooling）

增强后的邻域点特征 $\{\tilde{f}_i^1 \cdots \tilde{f}_i^k \cdots \tilde{f}_i^K\}$，张量形式为 $(K \times 2d_{in})$），将其进行均值池化。点特征通过一个多层感知机后，由均值池化函数聚合特征，最后通过一个共享多层感知机调整输出通道数，得到中心点的邻域聚合特征 $\hat{f}_i$，张量形式为 $(1 \times d_{\text{out}})$：

$$a_i^k = bn(\text{mlp}(\tilde{f}_i^k, W))\#(\text{AUTONUM}\backslash * \text{Arabic})$$

$$\hat{f}_i = \text{mlp}(\frac{1}{K}\sum_{k=1}^{K} a_i^k, W)\#(\text{AUTONUM}\backslash * \text{Arabic})$$

其中，mlp 是共享的多层感知机；W 为可学习参数，bn 为批归一化处理，$a_i^k \in \mathbb{R}^{2d_{\text{in}}}, \hat{f}_i \in \mathbb{R}^{d_{\text{out}}}, d_{\text{out}}$ 为输出通道数。

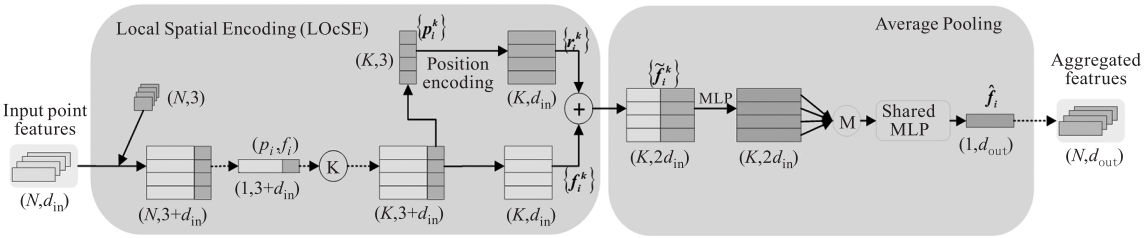

图 2 平均池化单元

Fig. 2 Average Pooling Unit

2.2 注意力门控单元（Attentive Gating Unit）

该模块输入点云形式为 $(N \times (3 + d_{\text{in}}))$，输出高级特征形式为 $(N \times d_{\text{out}})$，N 是点的数量，$d_{\text{in}}$ 是点云特征通道数，d_{out} 是输出局部特征通道数，如图 3 所示。对于输入点云 $P = \{p_1 \cdots p_i \cdots p_n\}$，首先通过局部空间编码模块增强语义特征，其次通过注意力池化模块聚合邻域特征得到 $\{\hat{f}_1 \cdots \hat{f}_i \cdots \hat{f}_n\}$，张量为 $(N \times d_{\text{out}})$，最后，输入的点云特征与聚合的特征进行残差连接（Skipping Connection），融合多尺度局部特征，输出 $\bar{\bar{F}} = \{\bar{\bar{f}}_1 \cdots \bar{\bar{f}}_i \cdots \bar{\bar{f}}_n\}$，张量为 $(N \times d_{\text{out}})$，能够表达点云之间的细微差异。

2.2.1 局部空间编码模块（Local spatial encoding，LocSE）

该模块的计算步骤同 3.1.1，输入中心点坐标和特征 $p_i \in \mathbb{R}^3, f_i \in \mathbb{R}^{d_{\text{in}}}$，利用 KNN 等算法输出增强的邻域特征 $\{\tilde{f}_i^1 \cdots \tilde{f}_i^k \cdots \tilde{f}_i^K\}, \tilde{f}_i^k \in \mathbb{R}^{2d_{\text{in}}}$，包含了丰富的空间几何信息和语义信息。

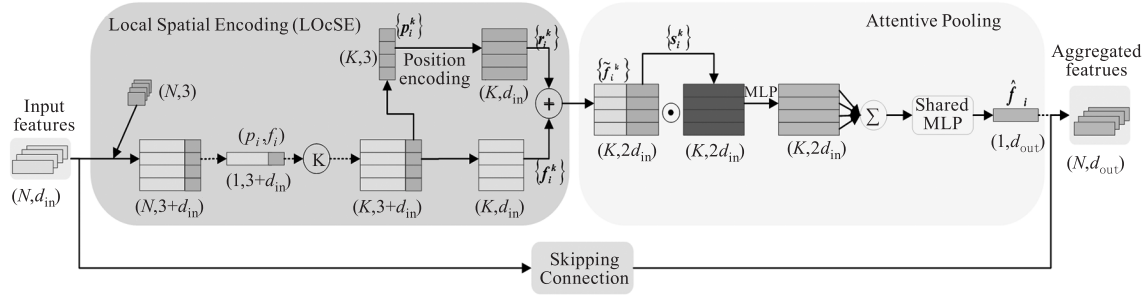

图 3 注意力门控单元

Fig. 3 Attentive Gating Unit

2.2.2 注意力池化模块(Attentive Pooling)

输入增强特征 $\{\tilde{f}_i^1 \cdots \tilde{f}_i^k \cdots \tilde{f}_i^K\}$,张量形式为 $(K \times 2d_{\text{in}})$,采用自注意力机制自适应地选择重要的特征,并利用通道求和函数聚合邻域特征,最后通过共享的多层感知机调整特征通道数为 d_{out},输出 $\hat{f}_i$,张量形式为 $(1 \times d_{\text{out}})$:

$$s_i^k = \text{softmax}(\text{mlp}(\tilde{f}_i^k, W))\#(\text{AUTONUM}\backslash * \text{Arabic})$$

$$a_i = \sum_{k=1}^{K} (\tilde{f}_i^k \odot s_i^k)\#(\text{AUTONUM}\backslash * \text{Arabic})$$

$$\hat{f}_i = mlp(a_i, W)\#(\text{AUTONUM}\backslash * \text{Arabic})$$

其中,W 为共享多层感知机的可学习参数;softmax 是非线性激活函数;$s_i^k \in \mathbb{R}^{2d_{\text{in}}}$ 学习每个点特征独有的注意力分数,自适应地选择重要的特征;$\odot$ 表示逐元素相乘;$a_i \in \mathbb{R}^{2d_{\text{in}}}$ 是聚合特征,用共享的多层感知机调整通道数,输出 $\hat{f}_i \in \mathbb{R}^{d_{\text{out}}}$。

2.2.3 多尺度特征融合

输入特征和局部聚合特征通过残差连接(Skipping Connection)相融合。用共享多层感知机调整输入特征 f_i 通道数,由 d_{in} 变为 d_{out},并与局部聚合特征相加,得到多尺度局部特征:

$$\bar{\bar{f}}_i = \text{LeakyRELU}(\text{mlp}(f_i, W) + \hat{f}_i)\#(\text{AUTONUM}\backslash * \text{Arabic})$$

其中 W 是共享多层感知机的可学习参数,+ 表示逐元素相加,LeakyRELU 是非线性激活函数,用于解决高维特征空间线性不可分问题,$\hat{f}_i, \bar{\bar{f}}_i \in \mathbb{R}^{d_{\text{out}}}, f_i \in \mathbb{R}^{d_{\text{in}}}$。

2.3 扩张残差模块(Dilated Residual Block)

该模块将平均池化单元和注意力门控单元堆叠,更高效地学习点云局部特征,如图 4 所示。该模块扩大每个点的特征感受野至 $K \times K$,并将最初的输入特征与第二层的输出特征相连接,融合低级、丰富的原始空间信息和高级的语义信息,得到更细粒度的局部特征,能更精准地表达相似点云模式之间的差异性。

2.4 网络结构

本文方法的网络结构主要采用基于残差连接的编码 - 解码结构,如图 5 所示。网络首先利用共享 MLP 学习每个点的特征,其次用四个编码层和四个解码层学习每个点的特征,最后利用三个全连接层和一个 Dropout 层用来预测每个点的语义类别。

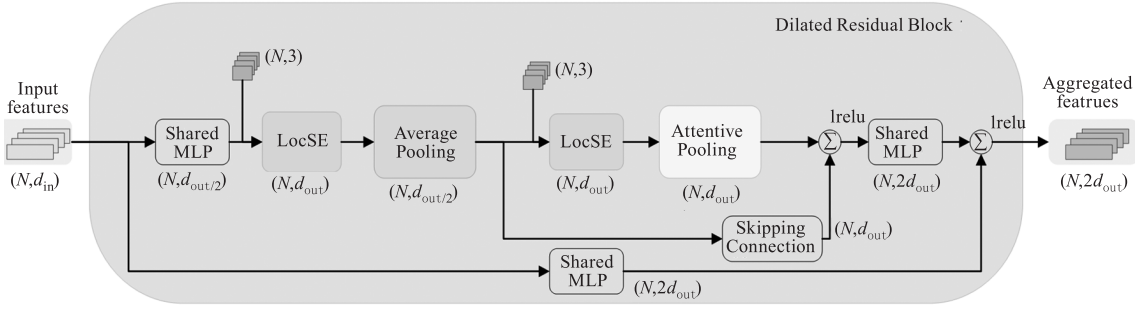

图 4 扩张残差模块

Fig. 4 Dilated residual block

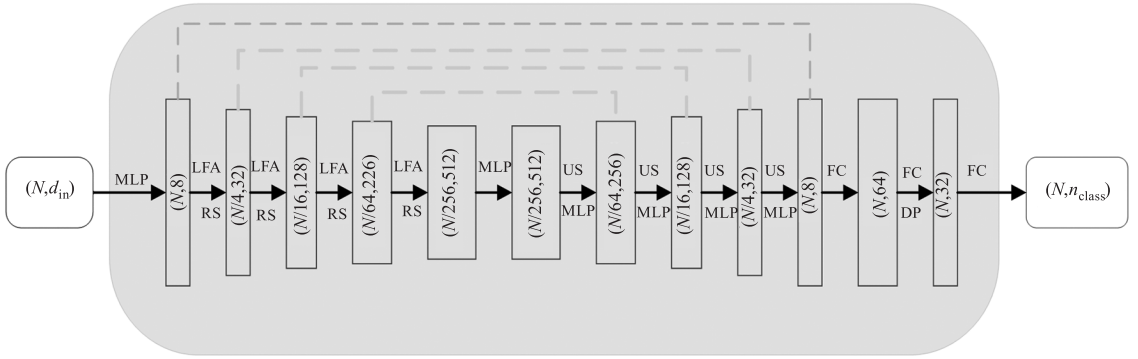

图 5 本文网络结构图

Fig. 5 The illustration of network architecture

网络输入:输入点云形式为 $(N \times d_{in})$,N 是输入点数量,d_{in} 是输入点特征,在 SemanticKITTI[12] 中是三维坐标 $x-y-z$。

编码层:每个编码层由局部特征聚合模块和随机降采样步骤组成,在经过四个编码层之后,点云越来越小,每个点的维度越来越高。点云的下采样率

为25 %,每一层下采样后只保留25 %的点,因此每层的点数为 $N \to \frac{N}{4} \to \frac{N}{16} \to \frac{N}{64} \to \frac{N}{256}$,每层点的特征通道数逐步提升,如8→32→128→256→512。

解码层:在每一个解码层,应用最近邻插值法从小点云中得到大点云的语义特征:在编码层中降采样后,原始点暂存起来,降采样得到的每个中心点都用KNN算法查找距离其最近的前一层中的点,将最近点的特征复制给中心点。随后将上采样的特征图与解码层中对应大小的特征图连接,得到多级融合的特征,增强网络的特征提取能力。

语义预测:最后三个全连接层和Dropout层推理得到每个点的语义预测。三个共享全连接层的输出特征张量形式为 $(N \times 64) \to (N \times 32) \to (N \times n_{\text{class}})$,Dropout参数为0.5。

网络输出:网络输出所有点的语义预测结果,张量形式为 $(N \times n_{\text{class}})$,其中 n_{class} 是类别数。

3 实验与分析

本文提出的方法在室外自动驾驶场景数据集SemanticKITTI[12]上进行实验,并将实验结果与其他方法进行对比分析,验证gRandLA-Net的有效性。

3.1 实 现

本文方法实验配置如表1所示。

表1 实验环境

Tab.1 The environment of experiment

CPU	GPU	Ubuntu	Tensorflow	Python
E5-2678	NVIDIA RTX 2080Ti	18	1.11	3.6

其他设置:本文方法使用Adam优化函数及其默认参数,初始化学习率设置为0.01,每轮学习率衰减5 %,采用反密度加权的交叉熵损失函数应对类别不平衡问题,用KNN算法查找领域点,邻域点数量K为16,网络训练100轮。测试期间,所有的原始点云直接输入训练好的网络进行推理,不需要切块或体素化等预处理步骤,也不需要任何后处理步骤。

3.2 评估指标

本实验以均交并比(mIoU)为评估指标,均交并比首先计算每个类别的交并比,再计算所有类别交并比平均值,能较好地评估模型整体分割性能:

$$mIoU = \frac{1}{k}\sum_{i=0}^{k-1} \cdot \frac{p_{ii}}{\sum_{j=0}^{k-1} p_{ij} + \sum_{j=0}^{k-1} p_{ji} - p_{ii}}$$

#(AUTONUM\ * Arabic)

其中,k 表示类别数;i 表示真实值;j 表示预测值;p_{ii} 是正确预测的正例;p_{ij} 是将 i 误分为 j 的集合;p_{ji} 是将 j 误分为 i 的集合。

交并比(IoU)是真实值和预测值的交集与并集之比,$IoU = \frac{TP}{FN + TP + FP}$。集合A是真实值,集合B是预测值,FN是错误预测的负例,TP是正确预测的正例,FP是错误预测的正例。显然,当IoU达到50 %以上就算是比较成功的预测。

3.3 量化分析和分割效果可视化

实验于室外大场景数据集SemanticKITTI[12]上进行。SemanticKITTI[12]由21个序列共43552帧标注的雷达点云组成,每帧包含 $8 \times 10^4 \sim 1.2 \times 10^5$ 个点,覆盖160 m×160 m×20 m的三维空间,规定序列00~07和09~10作为训练集(19130帧),08(4071帧)作为验证集,序列11~21(20351帧)用于线上测试,原始三维点云只有三维坐标没有颜色信息。网络在08序列上推理时间为189 s(4017帧),约22 fps。

3.3.1 本文方法与其他先进方法的量化分析

本文将实验计算精度结果与一些先进的网络结果进行了比较,如表2所示。第一类是基于点的方法,第二类是基于规则化数据的方法。本文的方法较大幅度地超过了PointNet[13],SPG[14],SPLAT-Net[15],pointnet++[16],TangentConv[17],RandLA-Net[9],FG-Net[18]等基于点的方法;并且超过了SqueezeSegV2[19],RangeNet53++[5],PolarNet[20],LatticeNet[21]等先进的基于结构化数据的方法。

3.3.2 本文方法的分割效果

gRandLA-Net的分割效果展示如图6所示,(a)中将人造地带terrain误分为植被vegetation;(b)中将卡车truck误分为汽车car;(c)中将其他地物other-ground误分为人person。

3.3.3 改进前后模型在各类目标上的性能分析

改进前后方法在各类目标上的性能分析,如图7所示。纵轴是改进前后方法在各类别上的IoU分数,横轴是19个类别由左向右按样本数量从小到大排列。前5个小目标类上IoU均有较大提升,如motorcyclist的IoU由7.2 %至11.4 %,提升了4.2 %,bicyclist的IoU由48.2 %至51.2 %,提升了2 %,bicycle的IoU由26 %至28 %,提升了2 %,motorcycle的IoU由25.8 %至31.2 %,提升了5.4 %,person的IoU由49.2 %至50 %,提升了0.8 %。

表 2 多种方法在 SemanticKITTI[13] 上的量化比较

Tab. 2 Quantitative results of different approaches on SemanticKITTI[13]

Method	mIoU/%	road	side-walk	park-ing	other-ground	buil-ding	car	truck	bicy-cle	motor-cycle	other-vehicle	veget-ation	trunk	terrain	person	bicy-clist	motor-cyclist	fence	pole	traffic-sign
PointNet[13]	14.6	61.6	35.7	15.8	1.4	41.4	46.3	0.1	1.3	0.3	0.8	31.0	4.6	17.6	0.2	0.2	0.0	12.9	2.4	3.7
SPG[14]	17.4	45.0	28.5	0.6	0.6	64.3	49.3	0.1	0.2	0.2	0.8	48.9	27.2	24.6	0.3	2.7	0.1	20.8	15.9	0.8
SPLATNet[15]	18.4	64.6	39.1	0.4	0.0	58.3	58.2	0.0	0.0	0.0	0.0	71.1	9.9	19.3	0.0	0.0	0.0	23.1	5.6	0.0
PointNet ++[16]	20.1	72.0	41.8	18.7	5.6	62.3	53.7	0.9	1.9	0.2	0.2	46.5	13.8	30.0	0.9	1.0	0.0	16.9	6.0	8.9
TangentConv[17]	40.9	83.9	63.9	33.4	15.4	83.4	90.8	15.2	2.7	16.5	12.1	79.5	49.3	58.1	23.0	28.4	8.1	49.0	35.8	28.5
RandLA-Net[9]	53.9	90.7	73.7	60.3	20.4	86.9	**94.2**	**40.1**	26.0	25.8	**38.9**	81.4	61.3	66.8	49.2	48.2	7.2	56.3	49.2	47.7
SqueezeSegV2[19]	39.7	88.6	67.6	45.8	17.7	73.7	81.8	13.4	18.5	17.9	14.0	71.8	35.8	60.2	20.1	25.1	3.9	41.1	20.2	36.3
DarkNet53Seg[19]	49.9	**91.8**	74.6	64.8	**27.9**	84.1	86.4	25.5	24.5	32.7	22.6	78.3	50.1	64.0	36.2	33.6	4.7	55.0	38.9	52.2
RangeNet53 ++[5]	52.2	**91.8**	**75.2**	**65.0**	27.8	87.4	91.4	25.7	25.7	**34.4**	23.0	80.5	55.1	64.6	38.3	38.8	4.8	58.6	47.9	55.9
LatticeNet[21]	52.9	90.0	**74.1**	**59.4**	22.0	88.2	92.9	26.6	16.6	22.2	21.4	81.7	63.6	63.1	35.6	43.0	46.0	58.8	51.9	48.4
PolarNet[20]	54.3	90.8	74.4	61.7	21.7	**90.0**	93.8	22.9	**40.2**	30.1	28.5	**84.0**	**65.5**	**67.8**	43.2	40.2	5.6	**61.3**	**51.8**	**57.5**
gRandLA-Net (ours)	**54.5**	90.8	74.4	59.9	20.0	86.3	93.8	39.4	28.0	31.2	38.3	80.2	63.0	66.1	**50.0**	**51.2**	**11.4**	56.8	50.1	45.6

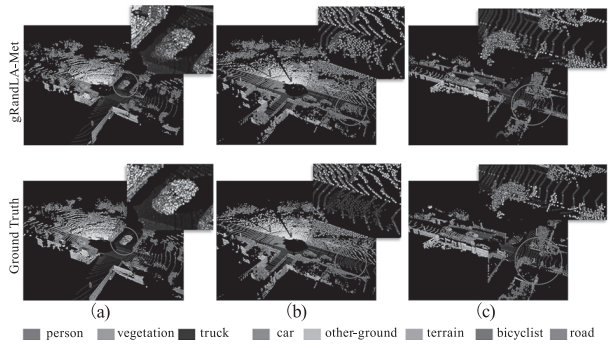

图 6 gRandLA-Net 的分割结果图

Fig. 6 Qualitative results of gRandLA-Net

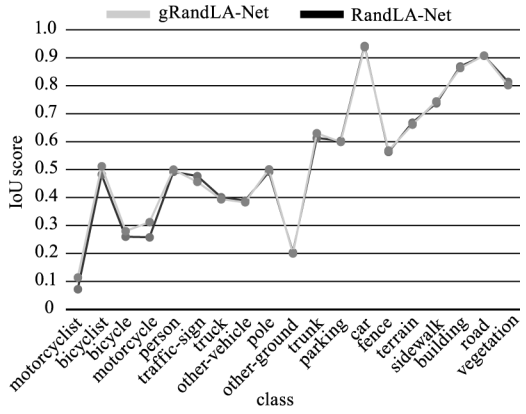

图 7 改进前后网络在各类别上的 IoU 分数对比

Fig. 7 The comparison of IoU score of gRandLA-Net and RandLA-Net

改进后网络对小目标分割性能有明显提升,同时,网络平均性能 mIoU 提升 0.6 %。

3.4 消融实验

为了验证分析 gRandLA-Net 模型的可行性和必要性,本文设置了消融实验。该部分通过对比多种算法来测试 gRandLA-Net 的效果,并进一步验证了平均池化单元、注意力门控单元在模型性能中发挥的具体作用。

消融实验均基于 SemanticKITTI[12] 数据集,评估指标为网络收敛速度(epoch),均交并比,为了保证本文方法可行、可信,本文取五次实验结果的平均数作为稳定的模型表现。

3.4.1 验证 gRandLA-Net 和各个模块的性能

网络各个模块消融实验结果如表 3 所示。

表 3 不同消融网络的均交并比和收敛速度

Tab. 3 The training epoch and meanIoU score of ablated networks

Network	Average Pooling Unit	Attentive Gating Unit	Epoch	mIoU/%
RandLA-Net	/	/	58	53.5
RandLA-Net + Average Pooling Unit	√	/	37	53.6
RandLA-Net + Attentive Gating Unit	/	√	50	54.1
gRandLA-Net(ours)	√	√	**30**	**54.5**

RandLA-Net[10] 收敛需要 58 轮,而基于 Average Pooling Unit 的 RandLA-Net 收敛需 37 轮;gRandLA-Net 使用了 Average Pooling Unit 后收敛轮数由 50 降至 30,因此平均池化单元使得网络收敛速度提高超过 40 %。

对比第一组和第三组网络性能,注意力门控单元使 mIoU 提高了 0.6 %;对比第二组和第四组网络性能,注意力门控单元使 mIoU 提高 0.9 %,因此注意力门控单元能有效提升网络性能。

3.4.2 进一步验证注意力门控单元的有效性和作用

为进一步验证注意力门控单元的有效性和作用,本文做了两组对比实验,量化结果如表 4 所示。

RandLA-Net[10]加上注意力门控单元后，mIoU 提升 0.6 %，motorcyclist 的 IoU 提升了 2 %，且在其他小目标（如 motorcycle，bicycle，bicyclist，person，等）上 IoU 均有小幅提升。gRandLA-Net 加上注意力门控单元后，mIoU 提升 0.9 %，在 motorcyclist 的 IoU 由 6.2 % 到 11.4 %，提升了 5.2 %，且在其他小目标（如 bicycle，bicyclist，person，traffic-sign 等）上 IoU 均有大幅提升。

表 4 针对注意力门控单元的消融网络上部分小目标的交并比和所有 19 类目标的平均交并比的量化结果

Tab. 4 IoU score ofsome small objects and the mean IoU score of all 19 classes in ablated study for attentive gating unit

Network	IoU of Motorcyclist / %	IoU of Bicyclist / %	IoU of Bicycle / %	IoU of Person / %	mIoU of Allclasses / %
RandLA-NetwithoutGating	6.1	43	24.7	47.1	53.5
RandLA-NetwithGating	**8.1**	**45.8**	**24.8**	**47.4**	**54.1**
gRandLA-NetwithoutGating	6.2	47.7	24	49.5	53.6
gRandLA-NetwithGating(ours)	**11.4**	**51.2**	**28**	**50**	**54.5**

因此证得，注意力门控单元利用几何上下文增强语义上下文，并融合多尺度感受野的局部聚合点特征，使得网络在稀疏的室外大场景点云中，对目标的几何模式感知能力更强，能够更加有效地感知相似模式的小目标点云之间的差异，对小目标的分割更加准确。

3.4.3 改进前后分割效果可视化对比

改进前后，模型在 SemanticKITTI[13]序列 08 上的分割效果如图 8 所示。RandLA-Net[10]在（a）场景中未能正确分割出 person，在（b）中未能正确分割 bicyclist，而本文方法 gRandLA-Net 分割更加准确。

图 8 改进前后模型在 SemanticKITTI[13]验证集上的分割效果图

Fig. 8 Qualitative results of RandLA-Net and gRandLA-Net on the validation set of SemanticKITTI

4 结 语

面对室外大场景点云中小目标点云难以识别的问题，本文提出注意力机制和多尺度上下文融合的方法，将点云不同局部感受野的几何模式结合起来，充分利用点云的局部几何信息，显著地提高了小目标的精度，同时还优化了网络训练速度。本文证明了融合多尺度的注意力上下文信息能够使得网络更加有效地感知具有相似模式的小目标点云之间的差异，在针对被大目标包围的小目标识别研究中具有明显的效用。

该方法虽然实现了更准确地分割，但是容易模糊各类目标点云的边界点，对边界点容易产生歧义。因此，下一步我们将研究增强网络对不同目标边界点的特征提取能力，以进一步优化对小目标的分割效果。

参考文献：

[1] Chen X，Ma H，Wan J，et al. Multi-view 3D object detection network for autonomous driving[C]//Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. 2017：1907 - 1915.

[2] 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）. IEEE，2019：4213 - 4220.

[3] Meng H Y，Gao L，Lai Y K，et al. Vv-net：voxel vae net with group convolutions for point cloud segmentation[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision，2019：8500 - 8508.

[4] Riegler G，Osman Ulusoy A，Geiger A. Octnet：learning deep 3D representations at high resolutions[C]//Proceedings of the IEEE Conference on Computer Vision and

Pattern Recognition,2017:3577 – 3586.

[5] QI C R,SU H,MO K,et al. Pointnet:deep learning on point sets for 3D classification and segmentation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2017:652 – 660.

[6] Liu Youqun,Ao Jianfeng. 3D point cloud semantic segmentation based on multi-informaion deep learnin[J]. Laser & Infrared,2021,51(5):675 – 680. (in Chinese)
刘友群,熬剑锋. 基于多信息深度学习的 3D 点云语义分割[J]. 激光与红外,2021,51(5):675 – 680.

[7] Wang L,Huang Y,Hou Y,et al. Graph attention convolution for point cloud semantic segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,2019:10296 – 10305.

[8] Luo C,Li X,Cheng N,et al. MVP-net:multiple view pointwise semantic segmentation of large-scale point clouds[J]. arXiv preprint arXiv:2201.12769,2022.

[9] Hu Q,Yang B,Xie L,et al. Randla-net:efficient semantic segmentation of large-scale point clouds[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020:11108 – 11117.

[10] Geng X,Ji S,Lu M,et al. Multi-scale attentive aggregation for LiDAR point cloud segmentation[J]. Remote Sensing,2021,13(4):691.

[11] Choe J,Park C,Rameau F,et al. PointMixer:MLP-mixer for point cloud understanding[J]. arXiv Preprint arXiv:2111.11187,2021.

[12] Behiey 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,2019:9297 – 9307.

[14] 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,2018:4558 – 4567.

[15] Su H,Jampani V,Sun D,et al. Splatnet:sparse lattice networks for point cloud processing[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2018:2530 – 2539.

[16] QI C R,YI L,SU H,et al. Pointnet ++ :deep hierarchical feature learning on point sets in a metric space[J]. Advances in Neural Information Processing Systems,2017,30.

[17] Tatarchenko M,Park J,Koltun V,et al. Tangent convolutions for dense prediction in 3D[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2018:3887 – 3896.

[18] Liu K,Gao Z,Lin F,et al. FG-conv:large-scale LiDAR point clouds understanding leveraging feature correlation mining and geometric-aware modeling[C]//2021 IEEE International Conference on Robotics and Automation (ICRA),IEEE,2021:12896 – 12902.

[19] Wu B,Zhou X,Zhao S,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),IEEE,2019:4376 – 4382.

[20] Zhang Y,Zhou Z,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,2020:9601 – 9610.

[21] Rosu R A,SchÜTT P,Quenzel J,et al. Lattice net:fast spatio-temporal point cloud segmentation using permutohedral lattices[J]. Autonomous Robots,2022,46(1):45 – 60.

[22] Qi C R,Su H,Mo K,et al. Pointnet:deep learning on point sets for 3D classification and segmentation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2017:652 – 660.