

基于局部特征聚合网络的三维语义分割

刘经纬^{1†}, 周彦²

(湘潭大学 自动化与电子信息学院, 湖南 湘潭 411105)

摘要: 激光雷达采集的自动驾驶场景点云数据规模庞大且包含丰富的空间结构信息, 一些方法将点云变换到体素化网格等稠密表示形式进行处理, 但却忽略了点云变换引起的信息丢失问题, 导致分割性能降低。为此, 提出了一种基于局部特征聚合网络的三维语义分割方法。其中的局部特征融合模块, 聚合中心点的 K 个最近点的特征, 并通过强大的注意力机制, 得到增强的点特征, 从而弥补丢失的信息, 提高网络的分割精度。此外, 为了提高小物体的分类精度, 提出了 3D 注意力特征融合块, 通过摒弃常规的特征图拼接, 使用注意力机制来决定不同层次语义特征的权重, 得到更加丰富的语义特征, 提高网络的性能。在 SemanticKITTI 和 nuScenes 数据集上的大量实验表明了该方法的优越性。

关键词: 语义分割; 三维语义分割; 局部特征聚合; 自动驾驶; 激光雷达

中图分类号: TP391

文献标识码: A

Local Feature Aggregation Networks for 3D Semantic Segmentation

LIU Jingwei^{1†}, ZHOU Yan²

(School of Automation and Electronics Information, Xiangtan University, Xiangtan, Hunan 411105, China)

Abstract: The point cloud data of autonomous driving scenes collected by LiDAR is large in scale and contains rich spatial structural detail information, and some methods transform the point cloud to dense representations such as voxelization grids for processing, but ignore the information loss and occlusion problems caused by the point cloud transformations, which leads to degradation of segmentation performance. For this reason, this paper proposes a local feature aggregation networks for 3D semantic segmentation. The local feature aggregation module therein aggregates the features of the K nearest points of the center point and obtains enhanced point features through a powerful attention mechanism, thus compensating for the lost information and improving the segmentation accuracy of the network. In addition, in order to improve the classification accuracy of small objects, this paper proposes a 3D attention feature fusion block, which obtains richer semantic features and improves the performance of the network by discarding the conventional feature map splicing and using the attention mechanism to decide the weights of different levels of semantic features. Extensive experiments on SemanticKITTI and nuScenes datasets demonstrate the superiority of the method.

Key words: semantic segmentation; 3D semantic segmentation; local feature aggregation; autonomous driving; LiDAR

随着三维激光雷达传感器的广泛应用及激光扫描技术的快速发展, 人类能够快速地获取场景以及物体的点云数据。与传统的相机相比, 点云作为

三维数据表示形式之一, 能够较好地表示复杂场景及物体的几何形状, 并且在物体空间关系及拓扑关系的表示方面具有独特的优势。同时, 点云模型的

分割也成了三维特征提取和识别中的一个十分重要的步骤^[1]。

近些年来,三维语义分割在自动驾驶领域的方法激增。例如,SqueezeSeg^[2]使用球面投影获取点云信息,使用 SqueezeNet 网络对投影得到的前视图进行特征提取,然后将条件随机场作为递归层对分割结果进一步优化。RangeNet++^[3]利用球面投影方式处理输入点云,并尝试使用 KNN 作为后处理。虽然这些方法与常规多视图方法相比,球面投影的方式保留了更多的信息。然而,这些方法由于严重的 2D 投影信息损失而具有较低的精度。

针对上述问题,一种替代方法是用基于体素的方法^[4,5],将无序的点云分割成一系列占用一定空间的体素,然后送入 3D CNN 逐步进行体素级别的特征学习,最后为每一格体素内的所有点匹配与该体素相同的语义标签。然而,由于室外激光雷达点云具有无序性以及密度不一致性,导致在体素化过程中损失几何信息导致分割精度不佳等问题。

为了改善这些问题,在基于体素的方法上融合了基于点的方法,提出了局部特征聚合模块,直接应用于输入的点云数据,不用改变三维拓扑结构,没有信息损失。该模块还采用了注意力机制,来获得增强的点特征。通过常规的点一体素融合方法将点特征与体素特征融合,得到更加丰富的语义特征。同时,还提出了一种 3D 注意力特征融合块,通过注意力机制自动选择不同层次的特征,来得到信息量更丰富的特征,从而改善网络性能,提高分割精度。

所提出的方法在两个大型户外数据集 SemanticKITTI^[6]和 nuScene^[7]进行了实验。

本文的贡献概述如下:

提出了一种局部特征聚合模块,能够有效解决体素化过程中改变三维拓扑结构所造成的问题,提高网络分割精度。

为了得到更加丰富的语义特征,进行准确的分类,提出了 3D 注意力特征融合块。

提出的方法在 SemanticKITTI 和 nuScenes 两个数据集的激光雷达分割任务上优于所比较的方法。

1 相关工作

三维语义分割任务根据给定点云的语义信息,

将同一类点归为一个子集,与二维语义分割相比,三维语义分割能够更细致地区分空间物体。以前的方法试图以三种方式解决这个问题。

1.1 基于点的方法

基于点的方法在原始点云上直接进行操作。PointNet^[8]在每个点上应用共享的多层感知器(MLP),通过最大池化将所有点的信息聚合得到全局特征,用于进一步的分割任务。然而,PointNet^[8]在复杂场景中表现较差,缺乏局部上下文提取。接下来的工作^[9,10]提出了球查询和 χ -Conv 来模拟二维卷积,并在室内场景中取得了很好的效果。然而,由于计算和存储成本,它们不适用于自动驾驶场景。为了减少计算成本,RandLA-Net^[11]采用了随机采样,但由于随机采样,精度较低。

1.2 基于体素的方法

基于体素的方法^[12,13]是一种结构化的表示方法,其将无序的点云分割成一系列占用一定空间的体素。该方法有效地促使无序点云规则化,但该方法采用固定网格,容易导致数据错误划分而使得预测结果不准确。Minkowski CNN^[14]是第一个高效的稀疏体素框架,它在精度和速度上都超过了大部分基于点的方法。一个可能的原因是稀疏体素是结构化的,这便于卷积运算。SPVNAS^[15]在文献^[14]中引入了神经结构搜索(NAS),并以较低的计算成本获得了更好的结果。Cylinder3D^[16]设计了圆柱形分割,并证明了其效率。

1.3 基于 2D 投影的方法

2D 投影方法通过将 3D 点投影到 2D 网格上来利用 2D 全卷积网络,然后融合这些特征以预测结果。RangeNet++^[3]提出了一种用于后处理的加速 KNN。SqueezeSegV3^[17]证明了空间自适应卷积的优越性。SalsaNext^[18]基于 SalsaNet^[19]设计了一种新的神经网络架构,实时执行完整的三维激光雷达点云的不确定性感知语义分割。PMF^[20]提出了一种全新的多传感器感知融合方案,提出了一种全新的 perception-aware loss,可以促进网络捕捉不同模态的感知信息。

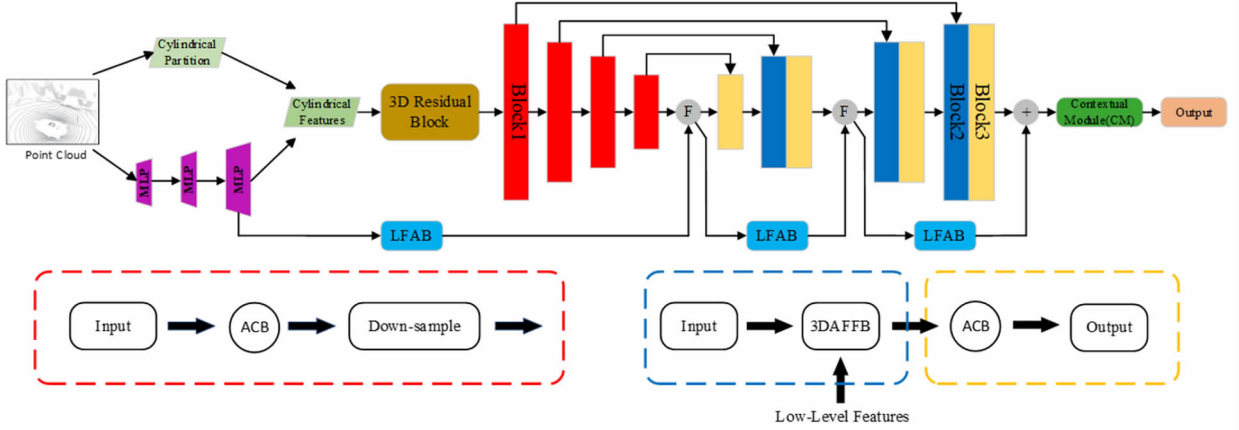
2 本文方案

2.1 总体框架

如图 1 上半部分所示,将数据集中的点云输入圆柱划分网络中,与通过多层感知机得到的点特征

相融合得到体素特征。圆柱划分网络采用的是圆柱分割^[14],并对其进行了微小的修改。与此同时,将数据集中的点云输入到局部特征模块中,得到更详细的点特征。将体素特征输入到非对称三维卷积网络,非对称三维卷积网络由三维残差块、四个

下采样块和三个上采样块组成,每个上采样块中都采用了3D注意力特征融合块以及非对称卷积块。通过这些组件组合同时与点特征相融合得到更加丰富的特征。然后将特征输入到上下文模块,最后通过 loss 函数得到分割结果。



(1) 上半部分是总体框架,⊕表示点-融合模块,LFAB表示局部特征聚合模块,MLP代表多层感知机,3D Residual Block代表3D残差块,CM代表上下文模块。(2) 下半部分详细展示了总体框架中的三个组件,包括Block1、Block2和Block3。Block1代表下采样块,包含一个非对称卷积块(ACB)和下采样操作。Block2和Block3组成了上采样块,包含一个3D注意力特征融合模块(3DAFFB)、一个非对称卷积块和上采样操作。

图1 总体框架图

2.2 局部特征聚合模块

由于室外点云的密度不一致性特性,导致在体素化过程中不可避免地损失几何信息,为此本文采用了局部特征聚合模块,该模块直接应用于原始无序3D点云,没有任何信息损失,通过逐步增加每个点的接收场来自动保留复杂的局部结构,得到增强的点特征。点特征与体素特征相融合,能够得到语义信息更加丰富的特征,从而提高网络的性能。局部特征聚合模块由局部空间编码块和2D注意力特征聚合块组成。

在给定输入的点云数据情况下,局部特征聚合模块能够聚合中心点的 K 个最近点的特征,并最终生成信息量丰富的特征。该模块由局部空间编码块和2D注意力特征聚合块组成,如图2所示。

局部空间编码块(Local Spatial Encoding):输入点云坐标以及每个点的特征,此局部空间编码块通过简单的 K 近邻(KNN)算法收集其相邻点 P_i^k ,明确嵌入所有相邻点 P_i^k 的坐标,从而使相应的点特征始终知道其相对点空间位置。这使局部空间编码块可以显式观察局部几何图案,从而使整个网络受益,有效地学习复杂的局部结构。对于中心点 P_i 的每个最近的 K 个点 $\{P_i^1, \dots, P_i^K\}$,其编码相对点位置为:

$$r_i^k = MLP [P_i; P_i^k; (P_i - P_i^k); \|P_i - P_i^k\|] \quad (1)$$

式中 MLP 为多层感知机, P_i 是中心点坐标, P_i^k 是相邻点坐标, $[\cdot]$ 表示拼接操作, $\|\cdot\|$ 表示欧几里得距离。将编码的相对位置 r_i^k 与它的对应点特征 f_i^k 进行拼接操作,从而获得增强的点特征 \tilde{f}_i^k , \tilde{f}_i^k 表示如下:

$$\tilde{f}_i^k = [r_i^k; f_i^k] \quad (2)$$

该模块最终的输出是一组新的局部特征 $F_i = \{\tilde{f}_i^1, \dots, \tilde{f}_i^K\}$ 。

2D注意力特征聚合块(2D Attentive Feature Aggregation Block):采用强大的注意力机制来自动学习重要的局部特征,输入的局部特征 F_i 通过一个MLP和Softmax操作,为每个空间位置计算权重 W_i 。 W_i 表示为:

$$W_i = Softmax(MLP(F_i)) \quad (3)$$

将权重与输入的局部特征逐元素相乘,然后在通道维度上进行求和,最后通过一个MLP,获得最终的输出特征。2D注意力特征聚合块表示为:

$$\tilde{F}_i = MLP(F_i * W_i) \quad (4)$$

式中 $*$ 表示逐元素相乘, \tilde{F}_i 表示输出的特征。

2.3 3D 注意力特征融合模块

3D 注意力特征融合块的核心思想是使用注意力机制来决定高层特征和低层特征的权重,然后将权重与输入的特征加权求和以生成最终的输出特征如图 3 所示,高层语义特征首先通过上采样操作来保持和底层语义特征相同的通道数。注意力机制的计算是通过 Attention 层来实现的,这个层包含了卷积、归一化、ReLU 和 Softmax 操作,用于计

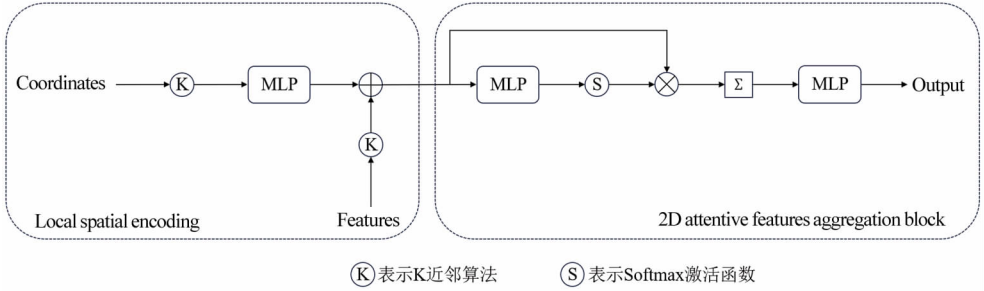


图 2 局部特征聚合模块(LFAB)

设 f_{high} 为输入的高层次特征, f_{low} 为输入的低层次特征,3D 注意力特征融合块表示为:

$$a = Att[ReLU(BN(Conv_{3 \times 3 \times 3}(Up(f_{high}))))); ReLU(BN(Conv_{3 \times 3 \times 3}(f_{low})))] \quad (6)$$

$$F = a \cdot ReLU(BN(Conv_{3 \times 3 \times 3}(f_{low}))) + (1 - a) \cdot ReLU(BN(Conv_{3 \times 3 \times 3}(Up(f_{high})))) \quad (7)$$

算注意力权重。最终输出特征是根据这些权重对输入特征进行加权求和得到的。这种方式可以有效地融合不同层次的语义信息,以改善模型性能,提高分割精度。Attention 层定义为:

$$Att(x) = Soft[Conv_{3 \times 3 \times 3}(ReLU(BN(Conv_{3 \times 3 \times 3}(x))))] \quad (5)$$

式中 a 表示注意力权重, Up 表示上采样操作, F 为输出的特征。

2.4 损失函数

总损失函数由交叉熵损失函数和 Lovasz-Softmax 损失函数^[21]组成。

$$L = L_{CE} + L_{lov} \quad (8)$$

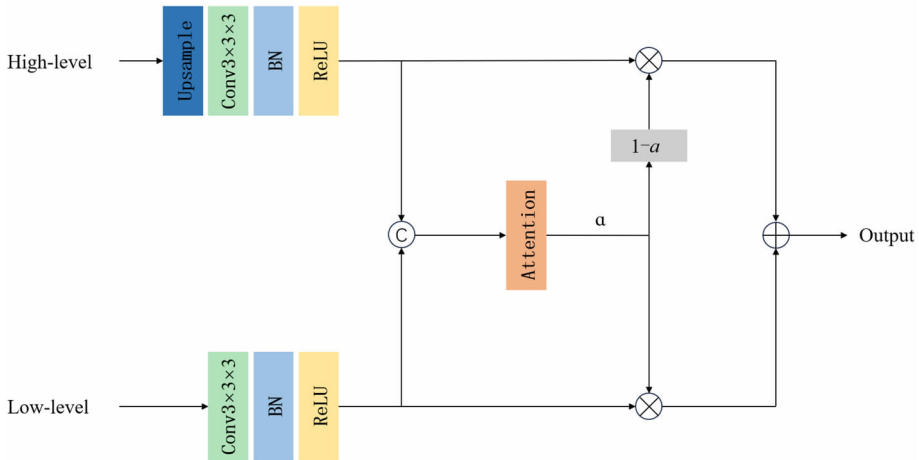


图 3 3D 注意力特征融合块(3DAFFB)

3 实验与分析

在 SemanticKITTI^[4] 和 nuScenes^[5] 数据集上评估了所提出的方法,以证明其有效性。

为了评估所提出的方法,采用了最流行的度

量——平均交并比(mIoU)。mIoU 通过以下方式计算:

$$mIoU = \frac{1}{K} \sum_{k=1}^K \frac{TP_k}{FN_k + FP_k + TP_k} \quad (9)$$

式中, TP 表示真正例,即模型预测为正例,实际是正例。 FN 表示假反例,即模型预测为反例,实

实际是正例。FP表示假正例，即模型预测为正例，实际是反例。K表示类别数。

3.1 在 SemanticKITTI 数据集的结果

如表1所示，将本文提出的方法的结果与 SemanticKITTI 验证集上先进的 LiDAR 分割方法进行了比较。本文提出的方法在所有的的方法中达到最好的性能。以输入方式的不同，将这些方法分为两大类：仅有激光雷达输入和多传感器融合的方法。本文提出的方法在所有的的方法中取得了最高的精度。与仅具有激光雷达的基于投影的方法 2D 空间相比，包括 RandLANet^[11]、RangeNet++^[3]、

SequeezeSegV3^[17] 和 SalsaNext^[18]，由于三维几何信息的建模，点云和体素表示能够更准确地反映物体的三维结构和几何特征，本文的方法在 mIoU 方面实现了 9%~18% 的性能提高。此外，与基于体素的方法相比，包括 MinkowskiNet^[14]，SPV-NAS^[15] 和 Cylinder3D^[16]，由于本文的方法中融入了点特征，分割精度也要优于这些方法。最后，与基于多传感器融合的方法相比，本文提出的方法也优于这些方法。SemanticKITTI 单次扫描的可视化结果如图4所示。

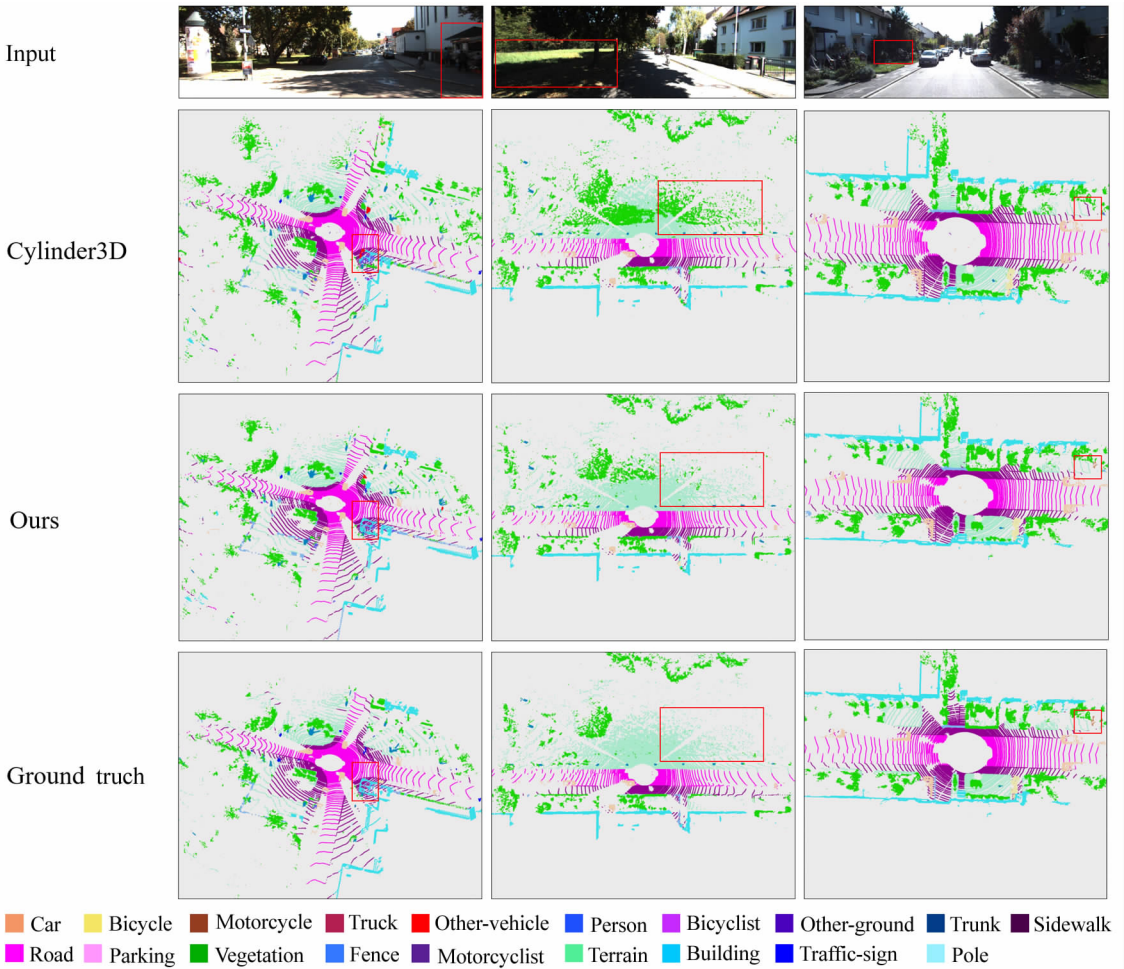


图4 SemanticKITTI 验证集上的可视化结果

3.2 在 nuScenes 数据集上的结果

对比本文提出的方法和最先进的方法在 nuScenes 验证集上的结果，如表2所示。本文方法在这些方法中取得了优异的性能。具体而言，相比于基于投影的方法和基于体素的方法，本文所提出的方法取得约 4%~15% 的性能提高。特别是对于一些需要更加详细语义特征的对象，如行人和

交通锥，本文提出的方法在这些对象分类的精度上取得了大幅提升。此外，本文提出的方法取得了比先进的多传感器融合方法更好的分割结果，即 2dpas^[22]，与之相比 mIoU 提高了 0.6%。这也表明了本文所提出的方法的有效性，可以以解决自动驾驶场景的难点。

3.3 消融实验

本文使用相同的实验设置在 SemanticKITTI 验证集进行消融实验,以确定所提出组件的有效性。如表 3 所示,比较表 3 中的第一行和第二行,

局部特征聚合模块实现了 1.2% 的 mIoU 提升。比较第二行和第三行,3D 注意力特征融合块使网络的 mIoU 提升了 1%。

表 1 在 SemanticKITTI 验证集上的结果

Method	Input	car	bicyclist	motorcyclist	truck	othervehicle	person	bicyclist	motorcyclist	road	parking	sidewalk	otherground	building	fence	vegetation	trunk	terrain	pole	traffic sign	mIoU
RangeNet++ ^[3]	L	89.4	26.5	48.4	33.9	26.7	52.3	46.0	0.0	92.9	37.0	69.9	0.0	83.4	51.0	83.3	54.0	68.1	49.8	34.0	51.2
SSV3 ^[17]	L	87.1	34.3	48.6	47.5	47.1	58.1	53.8	0.0	95.3	43.1	78.2	0.3	78.9	53.2	82.3	55.5	70.4	46.3	33.2	53.3
SalsaNext ^[18]	L	90.5	44.6	49.6	86.3	54.6	74.0	81.4	0.0	93.4	40.6	69.1	0.0	84.6	53.0	83.6	64.3	64.2	54.4	39.8	59.4
MN ^[14]	L	95.0	23.9	50.4	55.3	45.9	65.6	82.2	0.0	94.3	43.7	76.4	0.0	87.9	57.6	87.4	67.7	71.5	63.5	43.6	58.5
SPVNAS ^[15]	L	96.5	44.8	63.1	59.9	64.3	72.0	86.0	0.0	93.9	42.4	75.9	0.0	88.8	59.1	88.0	67.5	73.0	63.5	44.3	62.3
Cylinder3D* ^[16]	L	96.4	61.5	78.2	66.3	69.8	80.8	93.3	0.0	94.9	41.5	78.0	1.4	87.5	50.0	86.7	72.2	68.8	63.0	42.1	64.9
RGBAL ^[23]	L+C	87.3	36.1	26.4	64.6	54.6	58.1	72.7	0.0	95.1	45.6	77.5	0.8	78.9	53.4	84.3	61.7	72.9	56.1	41.5	56.2
PMF ^[20]	L+C	95.4	47.8	62.9	68.4	75.2	78.9	71.6	0.0	96.4	43.5	80.5	0.1	88.7	60.1	88.6	72.7	75.3	65.5	43.0	63.9
2dpass* ^[22]	L+C	96.8	55.0	75.2	93.5	68.9	78.2	91.8	0.1	93.9	50.8	81.3	2.8	92.1	67.7	89.5	72.9	77.0	63.4	52.1	68.6
ours	L	97.3	63.2	78.6	75.8	70.8	81.6	94.2	0	96.1	49.7	82.7	2.4	90.6	66.7	88.4	75.5	76.8	67.2	51.6	68.9

注:L表示仅用LiDAR方法。L+C表示基于LiDAR和图像融合的方法。*表示基于复现的结果。SSv3表示SqueezeSegV3,MN表示MinkowskiNet。

表 2 在 nuScenes 验证集上的结果

Method	barrier	bicyclist	bus	car	construction	motorcyclist	pedestrian	trafficcon	trailer	trunk	driv	otherflat	sidewalk	terrain	manmade	vegetation	mIoU
RangeNet++ ^[3]	66.0	21.3	77.2	80.9	30.2	66.8	69.6	52.1	54.2	72.3	94.1	66.6	63.5	70.1	83.1	79.8	65.5
PolarNet ^[24]	74.7	28.2	85.3	90.9	35.1	77.5	71.3	58.8	57.4	76.1	96.5	71.1	74.7	74.0	87.3	85.7	71.0
SalsaNext ^[18]	74.8	34.1	85.9	88.4	42.2	72.4	72.2	63.1	61.3	76.5	96.0	70.8	71.2	71.5	86.7	84.4	72.2
Cylinder3D ^[16]	76.4	40.3	91.3	93.8	51.3	78.0	78.9	64.9	62.1	84.4	96.8	71.6	76.4	75.4	90.5	87.4	76.1
PMF ^[20]	74.1	46.6	89.8	92.1	57.0	77.7	80.9	70.9	64.6	82.9	95.5	73.3	73.6	74.8	89.4	87.7	76.9
2dpass ^[22]	77.6	50.1	96.1	94.9	59.8	87.9	82.5	66.2	74.6	89.1	96.8	75.1	75.9	75.5	89.4	86.9	79.9
ours	78.6	47.8	95.4	95.6	59.2	87.2	84.8	68.9	76.1	88.7	97.2	75	77.4	77.1	91.4	87.6	80.5

注:加粗数字表示最佳结果。

表 3 SemanticKITTI 验证集进行消融实验

Baseline	LFAB	3DAFFB	mIoU
✓			66.7
✓	✓		67.9
✓	✓	✓	68.9

4 结论

提出了一种基于局部特征聚合网络的三维语义分割方法,其中 3D 注意力特征融合块,使用注

注意力机制来决定高层特征和低层特征的权重,将权重与不同层次的语义特征加权求和得到更加丰富的语义特征度。此外,局部特征聚合模块直接采用输入的原始点云数据,不损失任何信息。同时采用简单的K近邻算法,在保证效率的情况下,得到最近的K个相邻点坐标,聚合其特征,通过注意力机制,最终生成信息更丰富的点特征。最后,本文提出的方法在SemanticKITTI和nuScenes激光雷达分割任务上优于所比较的方法。在未来的工作中,我们将在保证分割精度的情况下,降低网络的计算成本,提高网络的运行速率。

参考文献

- [1] 盛仲飙, 韩慧妍. 散乱点云分割技术研究及实现[J]. 计算技术与自动化, 2016, 35(1): 104–106.
- [2] 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, 2018: 1887–1893.
- [3] 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, 2019: 4213–4220.
- [4] LE T, DUAN Y. Pointgrid: a deep network for 3d shape understanding[C]. Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 2018: 9204–9214.
- [5] GRAHAM B, ENGELCKE M, MAATEN L V D. 3d semantic segmentation with submanifold sparse convolutional networks[C]. Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 2018: 9224–9232.
- [6] 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, 2019: 9297–9307.
- [7] CAESAR H, BANKITI V, LANG A H, et al. nuscenes: A multimodal dataset for autonomous driving[C]. Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition, 2020: 11621–11631.
- [8] 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.
- [9] 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.
- [10] LI Y Y, BU R, SUN M C, et al. Pointnet: convolution on x-transformed points[J]. Advances in neural information processing systems, 2018, 31.
- [11] 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/CVF conference on Computer Vision and Pattern Recognition, 2020: 11108–11117.
- [12] MATURANA D, SCHERER S. Voxnet: a 3d convolutional neural network for real-time object recognition[C]. 2015 IEEE/RSJ international conference on Intelligent Robots and Systems, 2015: 922–928.
- [13] ZHOU Y, TUZEL O. Voxnet: end-to-end learning for point cloud based 3d object detection[C]. Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 2018: 4490–4499.
- [14] CHOY C, GWAK J Y, SAVARESE S. 4d spatio-temporal convnets: Minkowski convolutional neural networks[C]. Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition, 2019: 3075–3084.
- [15] TANG H T, LIU Z J, ZHAO S Y, et al. Searching efficient 3d architectures with sparse point-voxel convolution[C]. European Conference on Computer Vision. Cham: Springer International Publishing, 2020: 685–702.
- [16] ZHU X G, ZHOU H, WANG T, et al. Cylindrical and asymmetrical 3d convolution networks for lidar segmentation[C]. Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition, 2021: 9939–9948.
- [17] 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. Springer, 2020: 1–19.
- [18] CORTINHAL T, TZELEPIS G, ERDAL A E. Salsanext: fast, uncertainty-aware semantic segmentation of lidar point clouds[C]. International Symposium on Visual Computing. Springer, 2020: 207–222.
- [19] AKSOY E E, BACI S, CAVDAR S. Salsanet: fast road and vehicle segmentation in lidar point clouds for autonomous driving[C]. 2020 IEEE Intelligent Vehicles Symposium, 2020: 926–932.
- [20] ZHUANG Z W, LI R, JIA K, et al. Perception-aware multi-sensor fusion for 3d lidar semantic segmentation[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 16280–16290.
- [21] BERMAN M, TRIKI A R, BLASCHKO M B. The lovasz-softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks[C]. Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 2018: 4413–4421.
- [22] YAN X, GAO J T, ZHENG C D, et al. 2dpass: 2d priors assisted semantic segmentation on lidar point clouds[C]. European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 677–695.
- [23] MADAWI K E, RASHED H, SALLAB A E, et al. Rgb and lidar fusion based 3d semantic segmentation for autonomous driving[C]. 2019 IEEE Intelligent Transportation Systems Conference (ITSC). IEEE, 2019: 7–12.
- [24] 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, 2020: 9601–9610.