文章编号:1001-5078(2024)12-1841-09　　·激光应用技术·

面向铁路场景的大规模点云语义分割方法研究

孟维傑[1,2],吴嘉诚[3],孙淑杰[4],刘俊博[4],郭剑勇[5,6],田　媚[1,2],黄雅平[1,7]
(1. 交通数据分析与挖掘北京市重点实验室,北京交通大学,北京 100044;2. 北京交通大学 计算机与信息技术学院,北京 100044;
3. 联通数字科技有限公司,北京 100032;4. 中国铁道科学研究院基础设施检测所,北京 100080;
5. 中国铁路设计集团有限公司城市轨道交通事业部,天津 300142;
6. 城市轨道交通数字化建设与测评技术国家工程研究中心,天津 300308;7. 北京交通大学 唐山研究院,河北 唐山 063000)

摘　要:随着高速铁路和城市轨道交通系统的快速发展,对交通安全技术的研究日益迫切。应用激光扫描技术生成的铁路线路环境三维点云,能够对运行环境实现准确的感知。本文以铁路场景的三维点云数据为研究对象,首次构建了铁路场景下的大规模点云语义分割数据集。针对现有点云语义分割模型主要适用于于小尺度场景,但是铁路线路环境的三维点云数据规模很大,为此,本文提出了一种面向铁路场景的大规模点云语义分割方法,在编码阶段提出了一种基于自注意力的自适应局部特征融合模块,可以更好地聚合不同尺度的局部特征,解决类别不均衡的问题;在解码阶段提出了一种高维语义信息引导下的上采样方法,弥补了在编码阶段较大尺度的下采样造成的信息损失。所提方法在铁路场景数据集及公共室内数据集上都取得了优异的分割性能。

关键词:激光点云;点云分割;深度学习;铁路场景
中图分类号:TP391;TN249　　**文献标识码:**A　　**DOI:**10.3969/j.issn.1001-5078.2024.12.006

Large-scale point cloud semantic segmentation method for railway scene

MENG Wei-jie[1,2], WU Jia-cheng[3], SUN Shu-jie[4], LIU Jun-bo[4], GUO Jian-yong[5,6],
TIAN Mei[1,2], HUANG Ya-ping[1,7]
(1. Beijing Key Laboratory of Traffic Data Analysis and Mining, Beijing Jiaotong University, Beijing 100044, China;
2. School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China;
3. Liantong System Integration Co., Ltd., Beijing 100032, China;
4. Institute of Infrastructure Testing, China Academy of Railway Sciences, Beijing 100080, China;
5. Urban Rail Transit Division, China Railway Design Group, Tianjin 300142, China;
6. National Engineering Research Center for Digital Construction and Evaluation Technology of Urban Rail Transit, Tianjin 300308, China; 7. Tangshan Research Institute of Beijing Jiaotong University, Tangshan 063000, China)

Abstract: With the rapid development of high-speed railway and urban rail transit systems, research on traffic safety technology is becoming increasingly urgent. The 3D point cloud of railway line environment generated by applying laser scanning technology can achieve accurate perception and monitoring of operating environments. In this paper, the three-dimensional point cloud data of railway scenes is taken as the research object, and a large-scale point cloud semantic

基金项目:河北省自然科学基金项目(No. F2022105018)资助。
作者简介:孟维傑(1999 -),男,硕士研究生,主要研究方向为点云分割。E-mail:22120408@bjtu.edu.cn
通讯作者:黄雅平(1974 -),女,教授,博士生导师,主要研究方向为机器学习与认知计算。E-mail:yphuang@bjtu.edu.cn
收稿日期:2024-04-01

segmentation dataset for railway scenes is constructed for the first time. The existing point cloud semantic segmentation models are mainly applicable to small-scale scenes, and large scenic point clouds need to be segmented first. However, three-dimensional point cloud data of railway line environments have the characteristics of high data acquisition frequency and large data scale. Therefore, a large-scale point cloud semantic segmentation method for semantic perception of railway scenes is proposed in this paper. During the coding stage, an adaptive local feature fusion module based on self-attention is proposed in the encoding stage, which can better aggregate local features of different scales and solve the problem of category imbalance. In the decoding stage, an up-sampling method guided by high-dimensional semantic information is proposed to compensate for the information loss caused by large-scale down-sampling in the coding stage. The proposed method achieves excellent segmentation performance on both railway scene datasets and public indoor datasets.

Keywords: laser point cloud; point cloud segmentation; deep learning; railway scene

1 引 言

近年来,我国轨道交通领域实现了许多新的突破和进展。在高速铁路快速发展过程中,,如何保障交通运输的安全尤为关键。高速铁路的运行环境中出现的异物侵限等时间直接影响行车安全,甚至可能危及乘客的生命安全。近年来,在铁路环境安全检测领域中,激光扫描技术能够直接提供物体的几何信息,得到了广泛应用。目前,通过在综合巡检车上安装的车载激光扫描传感器,可以获取铁路线路环境的大量三维点云数据。基于这些数据对线路环境进行精确的识别和语义分割,能够有效地实现环境状态的监控和实时预警,不仅具有重要的理论研究意义,也具有很好的实际应用价值。

目前点云语义分割方法可分为两大类:基于体素的网络和基于点的网络。基于体素的网络首先将3D 空间细分为规则均匀的体素网格,然后进行卷积。虽然这种方法产生了有效的结果,但由于体素化导致丢失了详细的点云位置信息。基于点的网络可以直接处理点作为输入,避免了体素化预处理的需要,从而保证了点云位置信息的准确性。例如 PointNet[1] 通过最大池化的方法聚合特征;PointCNN[2] 采用递归神经网络获得局部信息;RandLANet[3] 通过随机采样及局部特征融合的方法提升性能与效率;KPConv[4] 利用核心点捕获局部信息并引入自适应权重;PointTransformer[5] 通过自注意力机制构建点之间的权重,但缺乏长期上下文信息影响鲁棒性。以上数据集虽然对三维点云数据的识别和语义分割进行了深入探索,但是针对铁路场景下三维点云语义分割的研究还比较少,仍面临着以下挑战:

(1)目前应用于铁路场景的公开点云数据集较少,而基于有监督学习的点云分割模型均严重依赖于大量的标注数据。

(2)现有的三维点云的识别和分割模型均在点云数据较少的数据集进行模型的设计和验证。由于高速铁路运行速度快,采集到的点云数据量大,目前还没有相关研究工作在如此大规模的密集室外数据集上进行验证,因此需要研究面向大规模点云数据的识别与分割算法。而在处理大规模点云数据时为提高效率通常会采用较大尺度的点云下采样,这将损失较多细节信息,影响分割效果。

(3)铁路线路场景中包含大尺度场景(隧道、站台等)和小尺度目标(轨道、道砟/轨道板、接触线、接触网杆柱、护栏、声/风屏障等),面临着类别不均衡的难题。目前针对该问题常用的处理方式主要有在预处理阶段对小样本数据进行增强[6];在模型编码层采用最远点采样[7]、反密度重要性采样[8]等有利于提高小尺度目标被采样概率的下采样方法,以及给小尺度样本赋值更大的权重。尽管这些方法有着不错的表现,但是仍无法达到实际应用需求。

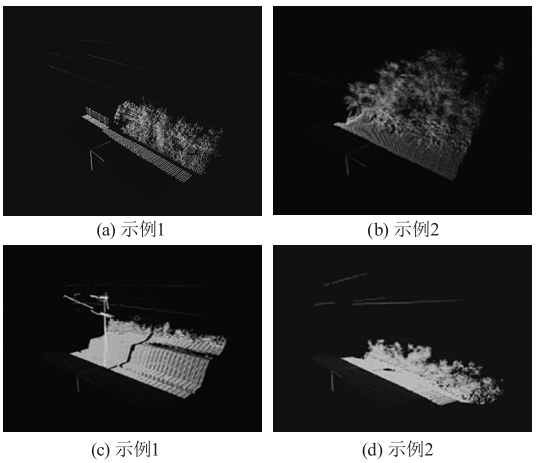
为了解决上述难题,本文首先构建了基于铁路场景的大规模点云分割数据集(Large-scale Point Cloud Dataset for Rail Scene, LSPC4RS),为后续的算法研究和实际应用提供了验证平台;其次,提出了一种面向铁路场景的大规模三维点云语义分割模型 RSFormer。该模型在编码阶段提出了一种基于自注意力的自适应局部特征融合模块,可以更好地聚合不同尺度的局部特征来自适应地获取大尺度场景特征和小尺度目标特征,解决类别不均衡的问题;在解码阶段提出了一种高维语义信息引导下的上采样方

法,该方法有利于将融合的高维特征信息通过上采样方法映射至每一个临近点,弥补了在编码阶段较大尺度的下采样造成的信息损失。RSFormer 在铁路场景数据集 LSPC4RS 及公共室内数据集 S3DIS 上都取得了良好的结果。

2 铁路场景点云语义分割数据集

有监督的深度学习模型依赖于高质量的数据标注,一个好的数据集能够帮助深度学习模型更准确地进行预测和分类,改善模型的泛化能力,提高模型的性能。因此,本文构建了面向铁路场景的大规模点云语义分割数据集 LSPC4RS,该数据集是由安装在综合巡检车上的车载三维激光扫描传感器采集得到的线路点云数据。经比对分析后筛选了 10 条典型线路数据,然后经人工标注获取精确的点云标注。该数据集包含了超 5 亿的人工标注点,近 50,000 个样本,每个样本包含 10,000 个点,涉及铁路场景中的 11 个语义类别,分别为树、杆、线、匝道、钢轨、边坡、护栏、隧道、站台、建筑物、以及其他,其中大尺度数据的种类有树、匝道、边坡、站台、隧道、建筑物;小尺度的种类包括杆、线、钢轨、护栏四种。

为方便高效地进行数据标注,采用了开源的标注软件 Semantic Segmentation Editor 进行数据标注,其是由日立汽车工业实验室(Hitachi Automotive And Industry Lab)开源的基于 Web 的语义对象标注编辑器(Semantic Segmentation Editor),该工具专门用于创建机器学习语义分割的训练数据,不仅支持普通相机拍摄的 2D 图像,还支持 LIDAR 生成的 3D 点云中的目标标注。标注样例如图 1 所示。



(a) 示例1 (b) 示例2

(c) 示例1 (d) 示例2

图 1 标注数据集展示图

Fig. 1 The visualization of labeling datasets

针对标注的 LSPC4RS 数据集,按照类别进行了统计。按照类别标注统计的结果如图 2 所示。

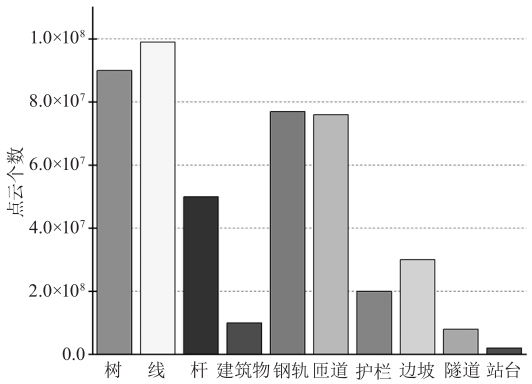


图 2 标注数据集样本统计

Fig. 2 Annotated dataset sample statistics

3 点云语义分割模型

在构建的点云语义分割数据集基础上,提出了一种面向铁路场景的大规模三维点云语义分割模型 RSFormer。针对铁路线路环境中的点云数据量大、以及场景内目标尺度差异大的特点,提出了新颖的 RSFormer 模型,通过编码层和解码层的优化设计,提升了铁路场景中点云语义分割的性能。

目前在处理大规模点云数据时,常用的语义分割模型在应对数据场景中小尺度目标分割任务上均表现较差,这是由于小尺度目标点数少,相对其他目标占比低,在进行下采样时被采样的概率小,而模型在编码阶段往往需要进行多次下采样,尽管可以通过最远点采样(FPS)和反密度采样(IDIS)等特殊设计的采样算法提高小尺度目标被采样的概率,但多次采样仍会使该目标信息熵骤降,从而导致模型效果表现不佳。此外目前大多数模型在局部特征编码阶段采用固定的领域尺度进行局部特征融合,而不同目标间存在尺度偏差,固定的编码尺度无法很好地适应这种偏差,特别是每次采样前后样本空间的同一目标也存在这种尺度差异,其中对小尺度目标的性能影响更为敏感。

为了缓解上述问题,本文提出了一种新的点云分割模型 RSFormer。RSFormer 是一种基于 Transformer 的端到端点云语义分割方法,采用 Encoder-Decoder 架构,如图 3 所示,主要包括一种基于自注意力的局部特征融合模块,该模块通过自适应获取大尺度场景特征以及小尺度目标特征,提高小尺度目标分割准确率;以及一种高维语义引导下的特征

上采样方法,该方法有利于将融合的高维特征信息通过上采样方法映射至每一个近邻点,弥补大尺度下采样造成的信息损失,提高模型的整体分割效果。

具体而言,在编码阶段,RSFormer 设计了一种自适应的多尺度局部特征融合模块。该模块通过对每个目标计算多个尺度的局部特征,之后采用自注意力机制对这多个尺度特征进行自适应融合,从而学习到不同采样空间下不同尺度物体的自适应特征,提高模型表达能力。此外对于分割任务,需要在解码阶段通过多次上采样将高维特征映射回原始点云空间样本上,目前采用的上采样算法采用广播复制或者线性插值的思想进行特征共享,这种简单的上采样方式将产生模糊的目标边界,影响模型效果。因此 RSFormer 设计了一种高维语义信息引导下的上采样方法,通过计算采样前后点云的特征相似度来指导高维特征共享,更准确地完成特征映射以获得更优表现。下面具体介绍模型整体结构及所提出的两种模块结构。

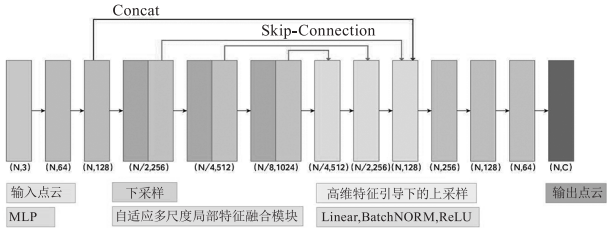


图 3 点云分割模型 RSFormer 架构

Fig. 3 The architecture of RSFormer

3.1 模型整体架构

本文提出的 RSFormer 模型如图 3 所示,采用 Encoder-Decoder 架构,在 Encoder 模块,主要使用全连接层,3.2 节中提出的自适应多尺度局部特征融合模块以及随机下采样层完成编码。Decoder 模块主要使用了 3.3 节提出的高维信息引导的上采样算法,采用 skip-connection 进行特征聚合,最终输入全连接层后使用 softmax 进行每个点的类别预测。

RSFormer 模型为基于 Transformer 的端到端点云语义分割网络模型,是一种直接输入点云即可用于点云分类以及点云分割任务的模型,原始点云通过输入一个共同的 Encoder 模块对点云数据进行多层多尺度局部特征编码,获取点云的全局特征,局部特征以及点特征融合成高维语义信息,该信息不仅可以进行广义的点云场景分类任务,同时在细粒度的点云语义分割任务中达到卓越的效果,这是因为在对每个点特征进行特征编码时,考虑了该点不同尺度的局部特征,并对不同尺度特征进行基于自注意力机制的权重融合,即让模型自适应地为场景内不同尺寸的目标进行相应尺度的局部特征融合,以获取该点精确的局部特征并与该点特征进行融合获得高维特征。

此外,为了获取不同层次特征以及扩充感受野,本章设计的 RSFormer 模型通过进行多次下采样,并在每次下采样后进行多尺度局部特征融合。即在第一次针对输入点云进行多尺度局部特征融合后,每个点融合了该点局部特征即获取原始点云的第一层高维特征,再对该点云进行下采样后获取采样点云,再对该采样点云进行多尺度局部特征融合,获得采样后点云的高维特征即为第二层高维特征,该特征是以第一层特征为基础生成的,因此是在第一层感受野基础上进行局部特征融合,进一步扩充了感受野。通过多次采样及局部特征编码后,完成了对原始点云的编码,再将该编码特征输入后续的分类或分割任务中进行预测。针对分类任务,则直接使用 MLP 将高维特征进行多次降维至类别数即可,而在点云分割任务中,由于需要输出原始点云中每一个点的类别,因此需要对编码模块输出的降采样后的高维特征点云进行逐步上采样恢复至原始点云大小,为有效恢复该特征,因此 RSFormer 模型采用了一种基于高维特征引导下的上采样算法逐步完成特征解码,最终输出各点类别,完成分割。

3.1.1 *编码层*

RSFormer 模型的输入为大尺度点云数据 H,大小为 $N \times d_{\text{in}}$,其中 N 为点数,d_{in} 为每个输入点的特征维度,首先经过两个 MLP 层将点云特征维度映射成 128,然后经过 3 个特征编码层以及 3 个下采样层,通过两两串联构成该模型的 Encoder 模块,其中特征编码为 3.2 节提出的自适应多尺度局部特征聚合模块,下采样层采用的是简单的随机采样算法,通过特征编码层进行特征维度扩展,通过下采样层降低点云尺寸。

3.1.2 *解码层*

通过 Encoder 层获取的高维特征,通过一个 MLP 层后输入 Decoder,该模型中的 Decoder 包括 3 个上采样层,该上采样为 3.3 节提出的高维语义引导的上采样算法。该算法结合 Encoder 中下采样后

的高维特征，通过注意力机制将高维特征映射至下采样前的点云数据，获得采样前点云的高维特征，即完成特征上采样，再通过 skip-connection 与 Encoder 中对应特征图结合，完成特征融合。通过逐步上采样，将所有层特征映射至原始输入点云中，最终通过全连接层以及 softmax 层完成点云类别预测。

3.2　自适应多尺度局部特征融合模块

该模块主要通过并行计算得到每个输入点的多个不同尺度范围的局部特征，然后对不同尺度范围的局部特征利用自注意力机制获取自注意力权重后进行特征加权融合。该模块主要包括三个部分：1）局部特征编码，2）注意力池化，3）自注意力特征融合。该模块设计图如图 4 所示，通过计算多次局部特征编码以及注意力池化获取多个不同尺度范围的局部特征后利用自注意力特征融合模块进行自适应多尺度局部特征融合。

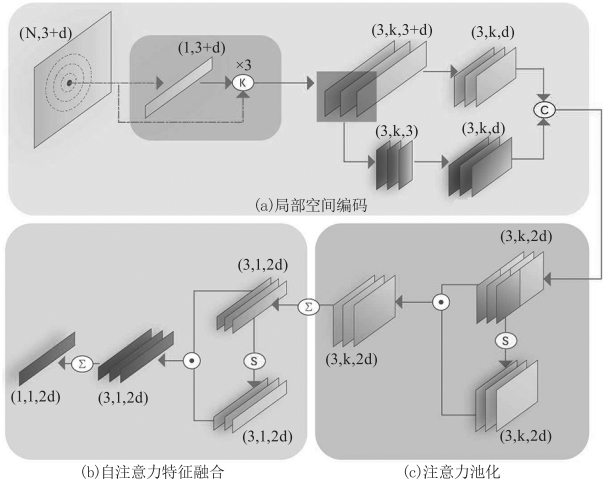


图 4　自适应多尺度局部特征融合模块

Fig. 4 The module of adaptive multi-scale local feature aggregation

3.2.1　局部空间编码

针对局部特征编码，给定一个输入点云 $P \in R^{N\times d}$，其中 N 为个点云个数，每个点的特征维度为 d。对于第 i 个点，利用基于欧式距离的 K 近邻算法获取每个中心点 p_i 的 K 个近邻点 $\{p^1{}_i \cdots p^k{}_i \cdots p^K{}_i\}$，则其相对位置编码如下所示：

$$r^k{}_i = MLP(p_i \oplus p^k{}_i \oplus (p_i - p^k{}_i) \oplus \|p_i - p^k{}_i\|) \tag{1}$$

其中，p_i 表示中心点 (x,y,z) 位置坐标；表示以 p_i 为中心点的近邻点 (x,y,z) 位置坐标；$\oplus$ 表示连接操作，$\|\cdot\|$ 表示计算中心点与近邻点间的欧式距离。

对于每一个近邻点 $p^k{}_i$，结合其相对位置编码 $r^k{}_i$ 与其对应的点特征 $f^k{}_i$ 获取特征增强向量 $\hat{f}^k{}_i$，则局部空间编码模块的输出为新的近邻点集特征：

$$\hat{F}_i = \{\hat{f}^1{}_i \cdots \hat{f}^k{}_i \cdots \hat{f}^K{}_i\}。$$

3.2.2　注意力池化

为聚合近邻点信息，已有 PointNet[1] 考虑到点云数据的无序性，因此首先使用最大池化层等对称函数进行信息聚合，这在一定程度损失了关键信息，因此 RSFormer 中采用了注意力池化策略，具体步骤如下：

给定一个局部特征集 $\hat{F}_i$，通过由一个共享的全连接层以及一个 softmax 层构成的函数 $g(\,)$ 来学习每一个特征的注意力分数。注意力分数 $s^k{}_i$ 定义如下：

$$s^k{}_i = g(\hat{f}^k{}_i, W) \tag{2}$$

针对学习到的特征分数，可以看作每个尺度特征的权重，表示该特征在特征集中的重要性程度，其计算公式定义如下：

$$\tilde{f} = \sum_{k=1}^{K} (\hat{f}^k{}_i \cdot s^k{}_i) \tag{3}$$

在经过局部空间特征编码以及注意力尺度，针对输入点云 P，计算得到其高维信息特征 $\tilde{f}_i$。

3.2.3　自注意力特征融合

由于对大规模点云进行编码需要进行大尺度的点云下采样，但是目前常用的下采样方法，例如最远点采样（FPS）、反密度重要性采样（IDIS）等，仍然可能存在关键信息损失等问题。因此通过融合多次局部空间特征及注意力池化可增加每个输入点的感受野，每次特征信息编码均基于上一次编码特征，从而减少下采样带来的信息损失。而常用的信息融合方式一般使用特征连接方式，这种方式虽然能够融合更多信息，但是根据经验启发，对于某些小尺度物体的检测，一个合适的刚好覆盖该小尺度目标特征融合范围往往可以更好地进行预测，降低更大范围非该物体的特征信息的干扰，因此该设计可以更好地针对不同尺度物体的检测。自注意力特征融合的具体描述如下：

首先对于给定的输入点云 $P \in R^{N\times d}$，前三维为点 (x,y,z) 点位置坐标，后 $d-3$ 维表示该点特征 $f^k{}_i$，就是经过局部空间编码以及注意力池化层后获取该点聚合局部特征后的高维特征 $\tilde{f}_i^{(1)}$。接下来，

将第一次计算得到的高维特征$\hat{f}^{(1)}{}_i$作为第i个点的特征描述子,继续经过局部空间编码以及注意力池化层后获取特征$\hat{f}^{(2)}{}_i$,继续重复上述过程计算获取$\hat{f}^{(3)}{}_i$。每次计算获取的高维特征均在上一次计算的基础上获得,每次计算的感受野均比上一次计算范围广,因此针对每个点获取了不同尺度范围的高维特征。为了能够自适应不同尺度的物体,RSFormer 设计了一种新颖的自适应特征融合方式。该方式利用自注意力机制,首先计算每个尺度范围的注意力分数,将该分数看作每个尺度特征的权重,进行加权求和获取不同尺度融合后的高维特征。其详细步骤如下:

针对每个输入点 i 计算得到的三个不同尺度的特征 $\{\hat{f}^{(1)}{}_i,\hat{f}^{(2)}{}_i,\hat{f}^{(3)}{}_i\}$,通过由一个共享的全连接层以及一个 softmax 层构成的函数 $h(\,)$ 来学习每一个特征的注意力分数,注意力分数定义如下:

$$S_i = g(\hat{f}^{(b)}{}_i, M), b = \{1,2,3\} \tag{4}$$

针对学习到的特征分数,可以看作每个特征的权重,表示该特征在特征集中的重要性程度。其计算公式定义如下:

$$\widetilde{F} = \sum_{k=1}^{K} (\hat{f}^{(b)}{}_i \cdot S_i), b = \{1,2,3\} \tag{5}$$

3.3 高维特征引导下的上采样算法

在 Decoder 阶段中,为了减少上采样的信息损失,RSFormer 引入了一种新的高维信息引导下的上采样算法,用于对 Encoder 中逐步下采样后获取的高维语义特征通过逐步上采样将高维信息映射至原始输入点进行预测。常用的上采样方法,如最近邻上采样、线性插值等,无法很好地弥补大尺度下采样方法带来的信息损失,同时可能导致在不同物体边界区域的误识别,因此提出了一种高维信息引导下的上采样方法,如图 5 所示。

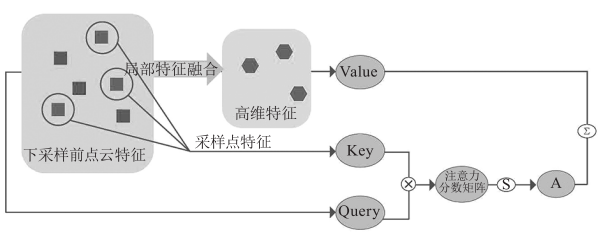


图 5 高维特征引导下的上采样算法

Fig. 5 Upsampling algorithm guided by high dimensional features

该方法将下采样前的点云特征集看作查询(Query)向量,将采样点特征看作键(Key)向量,将下采样点经局部特征聚合模块获取的高维特征看作值(Value)向量。通过计算查询向量与键向量的注意力分数后,将高维特征的值向量通过注意力分数矩阵映射到下采样前的点云内,即完成了高维语义信息引导下的上采样算法。其详细步骤如下。

针对局部特征编码,首先对于给定的输入点云 $P \in R^{N\times d}$,设其采样后点云为 $P_s \in RN^{s\times d}$,而该点云经过编码阶段的自适应多尺度局部特征融合模块后,获得的高维特征点云记为 $\widetilde{F}_s \in R^{N_s\times f}$,则将输入点云 P 当作查询矩阵 Q,采样后点云 P_s 看作键矩阵 K,则通过查询矩阵 P 与键矩阵 P_s 构建的注意力分数矩阵 S 表示如下:

$$S = P \times P_s^T (P \in R^{N\times d}, P_s \in R^{N_s\times d}, S \in R^{N\times N_s}) \tag{6}$$

为获取不同尺度场景的局部特征权重,需要将注意力分数转换成概率值,因此,对注意力分数矩阵进行归一化,然后输入 softmax 函数,表示如下:

$$A = softmax(\frac{S}{\sqrt{d}}) (A \in R^{N_s\times N}) \tag{7}$$

将 $\widetilde{F}$ 看作值矩阵 V,通过 A 对值矩阵 $\widetilde{F}$ 进行加权求和,即完成值向量的映射。其公式如下:

$$Z = A \times \widetilde{F} (A \in R^{N\times N_s}, \widetilde{F} \in R^{N_s\times f}, Z \in R^{N\times f}) \tag{8}$$

4 实验结果与分析

4.1 实验环境与评价指标

本文算法实验环境为 64 位 Linux 操作系统,GPU 为 NVIDIA RTX A4000,编程语言为 Python3.8,深度模型库为 Pytorch。评价指标使用平均交并比(mIoU)、总体类精度(OA)、平均分类精度(mAcc)进行性能的比较。计算公式分别如下:

$$\mathrm{mIoU} = \frac{1}{k}\sum_{i=0}^{k} \frac{S^{pt}{}_i}{S^{p}{}_i + S^{t}{}_i - S^{pt}{}_i} \tag{9}$$

$$\mathrm{OA} = \sum_{i=0}^{k} \frac{S^{pt}{}_i}{S^{p}{}_i} \tag{10}$$

$$\mathrm{mAcc} = \frac{1}{k}\sum_{i=0}^{k} \frac{S^{pt}{}_i}{S^{p}{}_i} \tag{11}$$

式中,k 表示 k 个类别;$S^{pt}{}_i$ 表示实际为第 i 类且预测为第 i 类的点云数量;$S^{p}{}_i$ 表示预测为第 i 类的点云数量;$S^{t}{}_i$ 表示实际为第 i 类的点云数量。

4.2 铁路数据集 LSPC4RS 的语义分割结果

首先对比了所提出的 RSFormer 与已有的先进方法 DGCNN[9] 以及 RandLA-Net[3] 模型的性能，结果如表 1 所示。

表 1 不同模型在铁路数据集 LSPC4RS 上的分割结果对比

Tab. 1 Comparison of segmentation results of different models on railway dataset LSPC4RS

模型	树	线	钢轨	匝道	护栏	边坡	mIoU	OA	mAcc
DGCNN[9]	55.1	99.5	90.8	**95.8**	68.9	87.6	82.9	94.4	75.1
RandLA-Net[3]	**71.6**	**99.5**	85.7	95.0	60.8	87.2	83.3	94.1	78.3
PointTransformer[5]	70.9	99.1	88.1	94.1	67.9	87.0	84.5	94.4	79.6
Ours	70.6	99.0	**91.5**	95.7	**69.2**	**87.8**	**85.6**	**94.6**	**80.9**

表 1 展示了各个不同类别的 IoU 及整体的 mIoU、OA 和 mAcc。本文提出的模型在 mIoU、OA、mAcc 上具有很大优势，这说明所提出的模型设计非常适合处理大规模的铁路场景，这得益于本文提出的特征聚合模块以及高维信息引导下的上采样模块。

本文在真实的铁路线路环境中进行了可视化处理，分割的结果如图 6 所示，可以看出，本文提出的点云语义方法能够获得理想的分割效果。

4.3 消融实验

为了进一步验证提出的自适应多尺度局部特征聚合模块以及高维语义信息引导下的上采样算法的有效性，本节给出相关的消融实验结果，如表 2 和表 3 所示。

表 2 中对比了使用本文的全部模型（Ours），使用自适应多尺度局部特征聚合模块的最后一层输出直接输入下一层的模型（last LFA），以及仅使用一层特征聚合的模型（one LFA）三个模型效果对比，从 mIoU 的结果上可以看出，完整模型依旧表现最优。其中仅使用一层特征聚合的模型（one LFA）效果最差，这也说明通过迭代融合多次局部特征是有效的，这是由于每次迭代融合都基于新的采样后样本空间特征，这不仅可以扩大对目标的感受野，同时可以提高模型的泛化能力，适应不同密度和尺度的输入数据。

(a)示例1 (b)示例2 (c)示例3 (d)示例4

图 6 分割结果展示

Fig. 6 The visualization of segmentation results

表 2 自适应的多尺度局部特征聚合模块有效性验证

Tab. 2 Validation of local feature aggregation module

方法	树	线	钢轨	匝道	护栏	边坡	mIoU
Ours(one LFA)	60.4	98.9	84.5	95.6	61.8	87.1	81.4
Ours(last LFA)	61.0	**99.4**	90.6	**96.0**	**70.8**	**88.0**	84.3
Ours	**70.6**	99.0	**91.5**	95.7	69.2	87.8	**85.6**

表 3 中对比了使用本文的上采样算法（Ours），与采用最近邻上采样算法和 PU-Net 上采样算法的性能对比，从实验结果可以看出，采用本章提出的上采样算法可以获得不错的效率提升。

4.4 公共室内数据集 S3DIS 的语义分割结果

为了进一步证明本文所提方法 RSFormer 的有效性，以大场景室内 S3DIS 数据集为研究对象，分析比较 RSFormer 与已有的先进模型的性能，结果如表 4 所示。

表 3 不同模型在铁路数据集 LSPC4RS 上的分割结果对比

Tab. 3 Validation of upsampling algorithm guided by high-dimensional semantic information

方法	树	线	钢轨	匝道	护栏	边坡	mIoU
最近邻上采样	69.5	95.8	91.2	93.9	67.3	87.1	84.1
PU-Net[10]	69.3	98.7	91.1	92.0	**69.5**	87.7	84.7
Ours	**70.6**	**99.0**	**91.5**	**95.7**	69.2	**87.8**	**85.6**

表 4 不同模型在公共室内数据集 S3DIS Area5 上的分割结果对比

Tab. 4 Comparison of segmentation results of different models on public indoor dataset S3DIS Area5

模型	OA	mIoU	mAcc
PointNet[1]	–	41.1	49.0
SegCloud[11]	–	48.9	57.4
TangentConv[12]	–	52.6	62.2
PointCNN[2]	85.9	57.3	63.9
PointWeb[13]	87.0	60.3	66.6
HPEIN[14]	87.2	61.9	68.3
GACNet[15]	87.8	62.9	–
PAT[16]	–	60.1	70.8
ParamConv[17]	–	58.3	67.0
SPGraph[18]	86.4	58.0	66.5
SegGCN[19]	88.2	63.6	70.4
PAConv[20]	–	66.6	–
KPConv[4]	–	67.1	72.8
PointTransformer[5]	**90.8**	70.4	76.5
Ours	90.7	**70.8**	**79.2**

表 4 对比了本文模型与现有最先进模型的性能，从结果可以看出，本文模型取得了与最先进模型 PointTransformer[5] 相当性能的效果，特别是在 mAcc 这一指标上超过了 PointTransformer[5] 2.7 % 的性能。

5 结 论

本文主要针对铁路场景点云数据提出了一种快速的语义分割模型 RSFormer。首先构建了一个铁路场景的点云语义分割数据集 LSPC4RS，针对铁路场景点云数据各类别尺度差异较大、密度不一致导致小尺度类别识别困难的问题，提出了一种自适应的多尺度局部特征聚合模块，此外，为更好地将编码的高维语义信息逐层映射到原始点云空间，设计了一种高维语义信息引导下的上采样算法，实验结果表明本文提出的基于铁路场景的语义分割算法在真实铁路点云场景中的可行性及有效性。

参考文献：

[1] Qi, Charles R, 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.

[2] Li, Yangyan, et al. Pointcnn: convolution on x-transformed points[C]//Advances in Neural Information Processing Systems 31, 2018.

[3] Hu, Qingyong, 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.

[4] Thomas, Hugues, et al. Kpconv: Flexible and deformable convolution for point clouds[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019.

[5] Zhao, Hengshuang, et al. Point transformer[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021.

[6] Chen Yunlu, et al. Pointmixup: augmentation for point clouds[C]//Computer Vision-ECCV 2020: 16th European Conference, Glasgow, UK, August 23 – 28, 2020, Proceedings, Part III 16. Springer International Publishing, 2020.

[7] Qi, Charles Ruizhongtai, et al. Pointnet++: deep hierarchical feature learning on point sets in a metric space[C]//Advances in Neural Information Processing Systems, 2017.

[8] Shu, Jun, Jie Zhang. CFSA-Net: efficient large-scale point cloud semantic segmentation based on cross-fusion self-attention[J]. Computers, Materials & Continua, 2023, 77(3): 2677.

[9] Wang, Yue, et al. Dynamic graph cnn for learning on point clouds[J]. ACM Transactions on Graphics (tog), 2019, 38(5): 1 – 12.

[10] Yu, Lequan, et al. Pu-net: point cloud upsampling network

[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2018.

[11] Tchapmi,Lyne,et al. Segcloud:semantic segmentation of 3D point clouds[C]//2017 International Conference on 3D Vision(3DV),IEEE,2017.

[12] Tatarchenko,Maxim,et al. Tangent convolutions for dense prediction in 3D[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2018.

[13] Han,Bing,Xinyun Zhang,et al. PU-GACNet:graph attention convolution network for point cloud upsampling[J]. Image and Vision Computing. 2022,118:104371.

[14] Zhao,Hengshuang,et al. Pointweb:Enhancing local neighborhood features for point cloud processing[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,2019.

[15] Jiang,Li,et al. Hierarchical point-edge interaction network for point cloud semantic segmentation[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision,2019.

[16] Yang,Jiancheng,et al. Modeling point clouds with self-attention and gumbel subset sampling[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,2019.

[17] Wang,Shenlong,et al. Deep parametric continuous convolutional neural networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2018.

[18] Landrieu,Loic,Martin Simonovsky. Large-scale point cloud semantic segmentation with superpoint graphs[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2018.

[19] Lei,Huan,Naveed Akhtar,et al. Seggcn:efficient 3D point cloud segmentation with fuzzy spherical kernel[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,2020.

[20] Xu,Mutian,et al. Paconv:position adaptive convolution with dynamic kernel assembling on point clouds[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,2021.