基于多特征融合的城市场景三维点云语义分割

刘贤梅,刘鹏飞,贾　迪,赵　娅
(东北石油大学 计算机与信息技术学院,黑龙江 大庆 163318)

摘　要:城市场景三维点云语义分割存在点云覆盖范围广、数据规模大、局部点云量稀疏、城市建筑风格各异等问题,仅依靠形状特征与颜色特征的分割方法无法对城市点云进行准确分割。该文提出了一种基于多特征融合的城市场景三维点云语义分割方法 MFFN(Multi-Future Fusion Network)。在预处理阶段对三维点云进行网格采样,降低了点云数据量,但同时最大程度保留点云的几何形状特征;引入每个采样点的法向量特征,有效弥补几何形状特征与颜色特征的不足;设计多特征局部聚合模块,将点云法向量特征和几何形状特征、颜色特征进行融合,增强网络对城市场景中表面凹凸程度相差较大的物体类别的学习能力。在 SensatUrban 城市数据集上的结果显示,该方法的平均交并比为 55.90%,总体精度为 91.90%,相比 RandLA-Net 网络分别提高了 3.21 百分点和 2.12 百分点,并且在多个城市类别上的分割精度均有较大提升。

关键词:城市规模三维点云;语义分割;法向量;特征融合;多特征局部聚合
中图分类号:TP391　　**文献标识码**:A　　**文章编号**:1673-629X(2023)11-0078-08
doi:10.3969/j.issn.1673-629X.2023.11.012

3D Point Cloud Semantic Segmentation of Urban Scene Based on Multi-feature Fusion

LIU Xian-mei,LIU Peng-fei,JIA Di,ZHAO Ya
(School of Computer and Information Technology,Northeast Petroleum University,Daqing 163318,China)

Abstract:The semantic segmentation of urban scene 3D point cloud has the problems of wide point cloud coverage,large data scale,sparse local point cloud volume,and different styles of urban buildings,etc. The segmentation method relying only on shape features and color features cannot accurately segment the urban point cloud. We propose a semantic segmentation method MFFN (Multi-Future Fusion Network) based on multi-feature fusion for 3D point cloud of urban scene. 3D point cloud is meshed during the preprocessing phase,which reduces the amount of data in the point cloud,while preserving the geometric features of the point cloud to the maximum. The normal vector feature of each sample point is introduced to effectively compensate for the shortcomings of geometric and color features. A multi-feature local aggregation module is designed,which combines the normal vector feature,geometric shape features and color features of point cloud to enhance the learning ability of the network for objects with different surface concaveness and convexity in urban scene. The results on the SensatUrban urban dataset show that the mean intersection over union of the proposed method is 55.90%,and the overall accuracy of it is 91.90%,which is 3.21 percentage points and 2.12 percentage points higher than that of the RandLA-Net network,respectively,and the segmentation accuracy is greatly improved in several urban categories.

Key words:urban-scale 3D point cloud;semantic segmentation;normal vector;feature fusion;multi-feature local aggregation

0　引　言

三维点云是在同一空间参考系下用于表达物体表面特征和空间分布的海量点的集合,相比于二维图像,点云可以提供丰富的几何形状信息,并且不易受光照变化和其它物体遮挡的影响[1]。三维点云语义分割的目的是为每个三维点分配语义标签,是三维场景理解和环境智能感知的关键问题之一,广泛应用于自动驾驶、高精地图、智慧城市等领域[2],大规模城市场景的三维点云覆盖范围广、数据规模大、局部点云量稀疏、城市建筑风格各异,使得城市场景的三维点云语义分

收稿日期:2022-12-07　　**修回日期**:2023-04-12
基金项目:黑龙江省教育科学“十四五”规划重点课题(GJB1421114);黑龙江省自然科学基金项目(LH2020F003);黑龙江省高等教育教学改革重点委托项目(SJGZ20200037)
作者简介:刘贤梅(1968-),女,硕士,CCF 高级会员(07945S),教授,硕导,研究方向为虚拟现实与媒体信息处理;通信作者:贾　迪(1995-),男,硕士,助教,研究方向为虚拟现实技术。

割面临严峻的挑战。

当前基于深度学习的点云语义分割方法可分为基于投影的方法和基于点的方法。

基于投影的方法:为了将成熟的二维图像语义分割方法应用于三维点云,文献[3]首次提出多视图投影方法,将输入的三维点云投影为多组二维图像,再利用图像语义分割网络对每个视图的图像进行联合分数预测。SqueezeSeg[4]利用球面投影将三维点云转换为二维图像,然后使用 SqueezeNet[5] 网络进行特征提取与分割,并应用条件随机场(Conditional Random Field,CRF)优化分割结果。文献[6]在 SqueezeSeg 的基础上设计了上下文聚合模块(Context Aggregation Module,CAM)以进一步提高分割精度。

基于点的方法:文献[7]提出了首个能够直接处理非规则点云的 PointNet 方法,该方法采用多层感知器和对称函数来学习和聚合点云特征,但其捕捉局部特征的能力很弱。

为解决该问题,PointNet++[8] 通过划分邻域的方法提取局部特征,该模型在将点云划分为多个有重叠的局部邻域后利用 PointNet 捕获局部邻域特征。

PointConv[9] 根据近邻点的距离赋予其不同的权重,再通过加权卷积聚合局部特征。

文献[10]提出可变形核点卷积分割框架 KPConv,该方法中的卷积权重是由每个邻域内定义的核点与其余非核点之间的欧几里得距离计算得出,核点的选择可根据不同情况进行修改,相比于 PointConv 更灵活。

文献[11]提出动态图卷积分割网络 DGCNN,用构建的动态图中每个节点代表点的特征,每条边代表邻域内点间的特征关系,且边会根据计算的邻域特征矩阵动态变化,使网络更容易聚类邻域内的相似特征。

文献[12]基于谱图理论设计了 RGCNN 模型,在构建动态图保存点云特征的基础上,利用图拉普拉斯矩阵自适应地捕获每一层动态图结构。

文献[13]将动态图的思想融入 PointNet++,设计了 DGPoint 动态图卷积网络,通过 K 近邻算法确定新的局部区域以达到动态图更新的目的。

文献[14]借助图的思想构建超点图(SuperPoint Graph,SPG),使网络捕获点云的上下文结构变得更精准。

GACNet[15] 通过注意力机制计算邻域中心点与每一个邻接点的边缘权重,从而使得网络能在分割的边缘部分取得更好的效果。

DALNet[16] 提出了一种基于双注意力机制的语义分割网络,结合空间注意力以及双线性插值法实现在解码阶段空间信息的高效恢复,在处理城市道路场景时有不错的效果。

此外,最近的 RandLA-Net[17] 设计了一个局部特征聚合模块,通过增加感受野的方式聚合局部点云的几何形状特征与颜色特征,极大程度地减少了信息损失,并采用随机采样的方法提高了网络可以同时处理的点云量。

尽管基于点的方法在三维点云语义分割上取得了不错的效果,但几乎都只适用于小规模室内场景或道路场景,无法扩展到大规模城市场景,这主要是由于城市场景点云数据规模更大,覆盖面积更广,对网络训练时的处理速度与内存开销要求极大。RandLA-Net 虽然通过随机采样的方法降低了网络训练过程中处理的点云量,但却牺牲了网络提取点云特征的准确性,而城市点云数据的局部区域点云量本身就稀疏,采样后几何形状信息更加难以提取,同时由于城市建筑风格的差异,颜色特征的描述能力也极大下降,因此网络已无法仅依靠几何特征和颜色特征来分割城市点云。为解决上述问题,该文提出了一种基于多特征融合的三维点云语义分割方法 MFFN,该方法的贡献如下:

为解决几何形状与颜色特征对城市物体描述能力减弱的问题,引入了点云的法向量特征,点云法向量在表面凹凸程度与光滑度相差较大的城市物体间有明显的差异,利用这种特性可有效弥补几何形状与颜色特征的不足,并基于 RandLA-Net 特征聚合思想设计了多特征局部聚合模块 MFLA(Multi-Feature Local Aggregation),将点云的法向量特征、颜色特征与几何特征进行融合,进一步提高了网络对城市场景三维点云的分割精度。此外,为解决城市点云数据规模大,局部点云量稀疏的问题,在数据预处理阶段与网络训练阶段分别采用网格采样与随机采样进行点云降采样。预处理过程中的网格采样保证了经过一次预处理之后输入到网络中的点云可以最大程度保留原始点云的几何形状特征,既保证了后续网络的训练速度,又缓解了局部点云量稀疏导致形状特征提取不准确的问题;网络训练过程中的多次逐层随机采样凭借其采样速度快的优势,大幅降低每层需要训练的数据量,进一步加快训练速度并降低内存开销。

1　基于多特征融合的城市场景三维点云语义分割网络 MFFN

1.1　MFFN 整体结构

MFFN 采用了带有跳跃连接的编码-解码结构,整体网络结构如图 1 所示。首先将预处理后的 N 个携带 D 维特征的采样点输入网络,利用四组编码解码层学

习每个点的局部特征,各层的特征维度为(8,32,128,256,512)。在每个编码层中利用一个多特征局部聚合模块(MFLA)融合局部邻域内的点云法向量、颜色和几何形状信息,并通过逐层随机采样方法降低训练点云量;之后在每个解码层中插入一组多层感知机(MLP)和近邻插值上采样(US),使采样点大小与每个采样点携带的特征维度逐步恢复到原始大小;其间利用跳跃连接将编码解码过程中提取的相同维度的特征信息进行融合;最后利用三个全连接层与一个 dropout 层对其进行输出,输出结果为 $N\times$Class,Class 为点云中的类别个数。

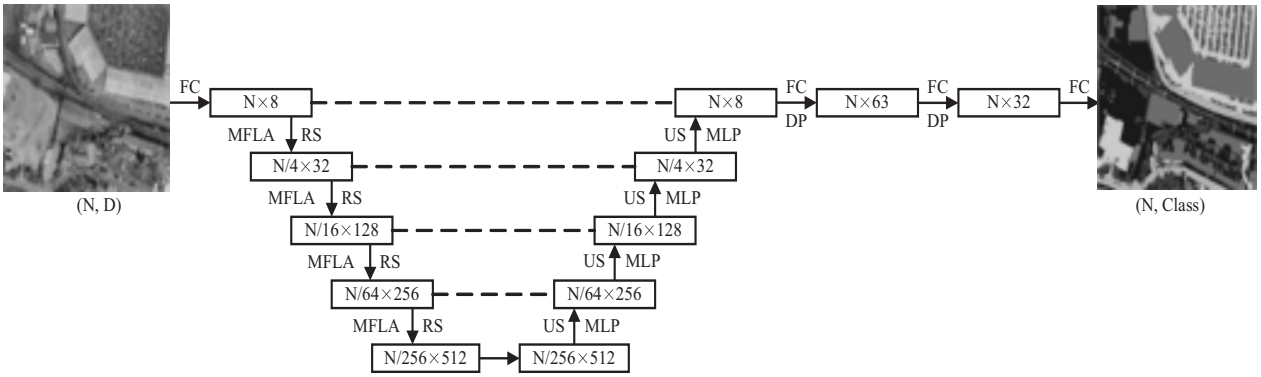


图 1 MFFN 架构

1.2 不同阶段的点云降采样

1.2.1 数据预处理阶段的网格采样

为了解决城市点云数据规模大、网络训练困难的问题,该文利用网格采样在降低点云数量的同时,能最大程度地保留点云几何结构的特点,在网络训练前先利用网格采样对点云进行预处理。

首先,通过遍历查找分别找出点云数据在 X、Y、Z 轴上的最大、最小坐标值,为输入点云建立一个能包围全部三维点的最小立方体。然后把该立方体划分为多个大小一样的小体素;然后,确定每个三维点所在的体素网格;最后,计算每个网格内三维点的重心,并利用该重心点代替网格内的所有三维点,即可得到网格采样后的点云数据。

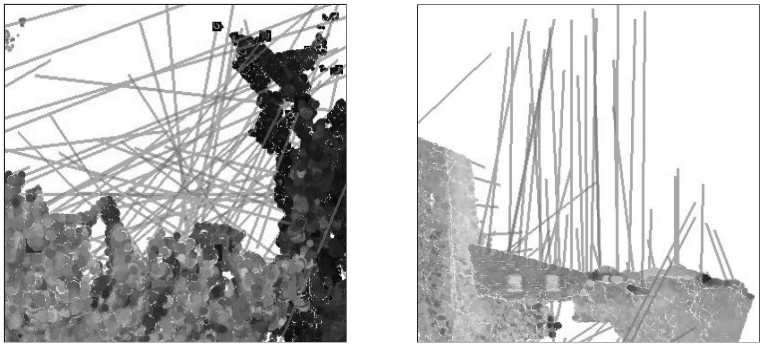
1.2.2 网络训练阶段的随机采样

随机采样会根据指定输出的采样点个数从输入点云中进行随机点选取。与网格采样、最远点采样[18]、反密度采样[19]等方法相比,随机采样的采样速度与输入点数无关,且在采样过程中没有中间运算步骤,计算效率极高。因此,在网络编码过程中利用随机采样逐层降低三维点数,以大幅提高网络的训练速度。

1.3 点云法向量特征分析与计算

1.3.1 法向量特征分析

为便于分析点云法向量的特性,图 2(a)(b)分别展示了植被和建筑物两个类别的法向量特征局部放大图,图中线条为相应三维点的法向量。



(a)植被局部放大图 (b)建筑物局部放大图

图 2 部分类别法向量特征放大图

从图 2 中可以看出,植被等不规则物体的法向量朝向参差不齐,但建筑物等光滑物体的法向量朝向基本一致,可见点云法向量在表面凹凸程度与光滑度相差较大的城市物体间有明显的差异,这种特征在多数的城市物体上均有所体现。因此,利用法向量的这种特性辅助语义分割网络,可有效加强网络对城市场景中这些类别的学习能力。

1.3.2 法向量计算

法向量的计算方法有三种[20]:基于 Delaunay 三角分割的方法会受到离群点和噪声的影响,因此不适用

于现场采集的数据;基于统计学原理的方法计算复杂度很高,且要求点云在尖锐特征处的采样密度足够稠密,因此不适用于局部区域稀疏的城市点云;基于局部表面拟合的方法计算原理简单清晰、速度快、使用范围广,该文采用该方法,具体计算过程如下:

首先进行局部区域表面拟合,对于每个采样点 p,利用 k 近邻搜索算法搜索到与其最近的 k 个近邻点,然后根据最小二乘法对 k 个点进行曲面拟合,形成曲面的表达形式,如公式(1)所示。

$$p(\vec{n},d)=\operatorname{argmin}\sum_{i=1}^{k}(\vec{n}p_i-d)^2 \tag{1}$$

式中,d 为平面 P 与坐标原点间的相对距离,$\vec{n}$ 为拟合平面 P 的法向量,即该采样点的法向量。为了求解法向量,需先计算出该邻域内点云的质心 p_c,如公式(2)所示。

$$p_c=\frac{1}{k}\left(\sum_{i=0}^{k}x_i,\sum_{i=0}^{k}y_i,\sum_{i=0}^{k}z_i\right) \tag{2}$$

随后,对式(3)中的协方差矩阵 $\boldsymbol{M}$ 进行特征值分解,求得 $\boldsymbol{M}$ 的所有特征值,其中最小的特征值所对应的特征向量即为所求的法向量。

$$\boldsymbol{M}=\sum_{i=1}^{k}(p_i-p_c)(p_i-p_c)^{\mathrm{T}} \tag{3}$$

上述求得的法向量只确定了其所在的直线而未确定其方向,因此为了消除二义性需为法向量定向。假设 $\vec{n}_i$、$\vec{n}_j$ 为相邻两点 p_i、p_j 的法向量,则 $\vec{n}_i$、$\vec{n}_j$ 应近乎平行,即其内积为 1;反之,则说明其中一点的法向量方向错误,需要反转。基于上述方法,只需设定好初始点的法向量朝向,再遍历其余点即可重定向所有点的法向量。

1.4　多特征局部聚合模块 MFLA

在网络的随机采样过程中,不可避免地会丢失一些携带重要信息的三维点,为了同时解决信息丢失和几何形状与颜色特征对城市场景表达不充分的问题,该文引入 RandLA-Net 局部特征聚合的思想,设计了多特征局部聚合模块,整体框架如图 3 所示。该模块主要由多特征局部编码模块、注意力池化模块两部分堆叠而成,并将堆叠后的输出特征与输入点云经过多层感知机处理后的特征相加,获得最终的聚合特征。通过多特征编码的方式,缓解重要三维点丢失带来的精度下降问题;同时,通过聚合法向量特征,网络可以更好地学习一些特定城市类别的特征信息,进一步提高模型精度。

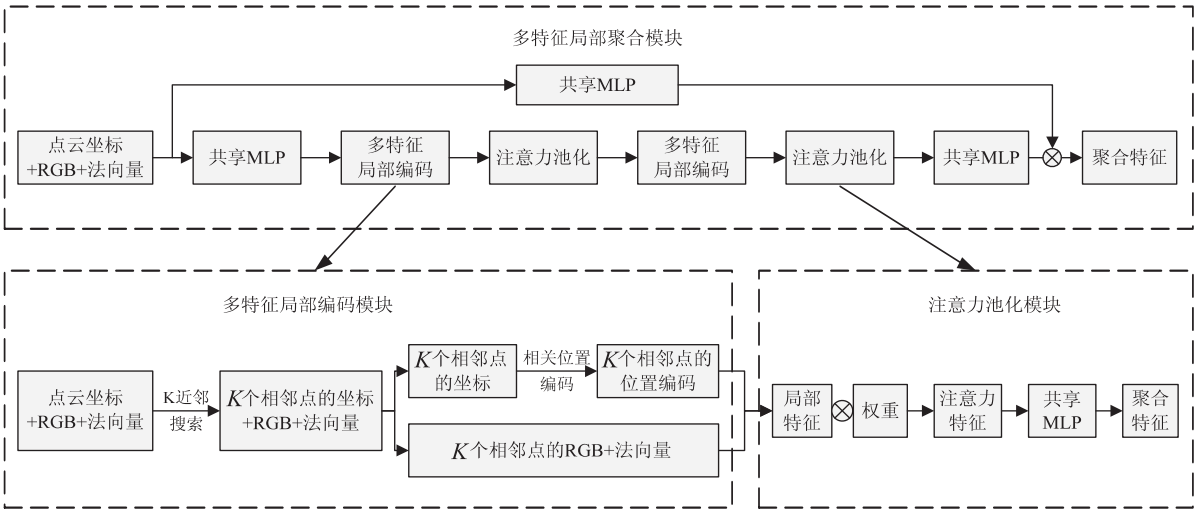


图 3　多特征局部聚合模块

1.4.1　多特征局部编码模块

多特征局部编码模块通过编码的方式将每个采样点与其近邻点之间建立联系,使每个采样点除了携带自身的多特征信息之外,还会携带与其它邻域点之间的特征关系,这样即使在随机采样过程中一些重要的三维点丢失,其部分特征信息仍能保留在其它邻域点的特征编码中,使网络后续模块可以更好地聚合局部特征;同时,考虑到点云的法向量特征在一些特定的城市类别上的差异性很大(铁轨、植被、建筑等),通过将相对位置编码与颜色信息、法向量信息进行级联的方式,加强网络对这些类别的学习能力,进而提高整体的分割精度。

该模块首先采用 k 近邻搜索算法得到每个输入的采样点 p 的 k 个临近点的三维坐标、颜色以及法向量,然后对点 p 及其 k 个近邻点的位置信息进行编码,使其间建立联系。编码的相对位置信息包括 p 点的三维坐标 p_i、k 个临近点的位置坐标 p_i^k、p 点与 k 个邻近点的相对位置关系 $p_i-p_i^k$ 以及它们的欧氏距离 $\|p_i-p_i^k\|$,$\oplus$ 表示级联运算,最后输出点 p 的相关位置编码 r_i^k,整体编码方式如公式(4)所示。

$$r_i^k = \mathrm{MLP}(p_i \oplus p_i^k \oplus (p_i - p_i^k) \oplus \| p_i - p_i^k \|) \tag{4}$$

得到每个采样点与其局部邻域内近邻点的相对位置编码后,将编码结果 r_i^k 与点云的颜色特征信息 fc_i^k 、法向量特征信息 fn_i^k 级联就得到了每个采样点的增强特征 $\hat{f}_i^k$,如公式(5)所示。

$$\hat{f}_i^k = \mathrm{fc}_i^k \oplus \mathrm{fn}_i^k \oplus r_i^k \tag{5}$$

1.4.2 注意力池化模块

注意力池化模块用于聚合点的局部特征。首先,将多特征局部编码模块获得的邻域点多特征信息送入可学习的全连接层;然后,使用 Softmax 激活函数获得该点多特征信息的权重;最后,将得到的所有邻近点特征权重加权求和获得局部邻域聚合特征。在处理大规模城市场景的点云时,该方法相比于最大池化或平均池化的优势在于注意力池化可以自动选择重要的局部特征,进一步降低随机采样丢失关键点信息的影响,提高分割网络的精度。

2 实 验

2.1 实验数据集

实验使用的数据集是牛津大学的胡等人在 2021 年公开的 SensatUrban 数据集[21],该数据集是城市规模摄影测量点云数据集,其中包含三个英国城市(伯明翰,剑桥以及约克)7.6 平方公里中的近 30 亿具有详细语义标注的点,同时包含每个点的位置信息和颜色信息,共分为地面、植被、建筑物、墙面、桥梁、停车场、铁轨、交通路、街道设施、汽车、人行道、单车和水 13 个语义类别。

其中,伯明翰城市数据中类别具体占比如表 1 所示,其它城市中的类别占比与伯明翰类似。

表 1 伯明翰数据集中各类别占比(误差在 0.001 ~0.01 之间)

类别	地面	植被	建筑物	墙面	桥梁	停车场	铁轨	道路	设施	汽车	人行道	单车	水
占比	0.273	0.148	0.297	0.025	0.002	0.068	0.001	0.116	0.021	0.027	0.006	0.000 01	0.002

2.2 实验环境

实验环境为 Linux Ubuntu 18.04 操作系统、Intel(R) Xeon(R) Silver 4210 处理器、RTX 3090 显卡,使用 CUDA11.2 加速 GPU 计算,深度学习框架为基于 python3.8 的 tensorFlow2.6.0。

2.3 实验结果与分析

2.3.1 预处理阶段采用不同采样方法对网络分割结果影响的对比分析

为验证网格采样法在应用于大规模城市场景点云数据预处理时的优越性,分别采用随机采样与网格采样对数据进行降采样,检测其对模型训练的影响,评价指标包括总体精度(OA)、平均交并比(mIoU)和各个类别的交并比(IoU),实验结果如表 2 所示。由于目前无法获得测试集标签,该文展示的相关实验结果均是将训练好的模型上传至 SensatUrban 数据集发布者提供的官方网站后获得的。从表 2 中可以看出,采用网格采样的方法处理后的数据训练出来的模型,其各项指标均优于随机采样处理后的训练数据,尤其是在一些分割精度本就较低的类别上差距尤为明显,如停车场、铁轨、道路等。这主要是由于网格采样与随机采样相比,保证了相对稀疏的位置也会有适量的三维点得以保留,使网络可以更好地学习点云局部特征。

表 2 随机采样与网格采样对训练模型的影响 %

采样方法	OA	mIoU	地面	植被	建筑物	墙面	桥梁	停车场	铁轨	道路	设施	汽车	人行道	单车	水
随机采样	91.20	54.40	84.40	98.20	94.40	56.50	38.80	47.10	12.90	51.40	37.90	77.40	37.60	0.00	69.00
网格采样	91.90	55.90	84.60	98.20	95.90	57.60	40.50	53.90	13.60	55.90	39.60	78.88	37.80	0.50	69.70

2.3.2 融合不同点云特征的实验结果对比分析

表 3 对比了在几何特征中依次融入颜色特征与法向量特征的分割结果。可以看出,当点云数据中存在多种类型时,该文提出的多特征融合算法对地面、建筑物、墙面、铁轨、植被、停车场、人行横道等类别的分割精度均有较大提升,其中对铁轨分割精度的提升最为明显,由 0% 提升至 13.60% 。这主要是由于这些类别有着自己独特的法向量特征,极大程度地降低了这些类别之间相互错分的概率,如地面与停车场、道路与铁轨、植被与城市设施等,说明融入法向量特征提高了网络对城市场景点云中这些类别的分辨能力。

表 3 融合单几何特征、几何特征+颜色特征、几何特征+颜色特征+法向量特征三种情况的分割结果对比 %

	OA	mIoU	地面	植被	建筑物	墙面	桥梁	停车场	铁轨	道路	设施	汽车	人行道	单车	水
XYZ	88.60	47.10	77.40	87.50	90.20	43.80	39.90	41.00	0.00	47.30	30.00	77.30	15.30	0.00	62.30

续表 3

	OA	mIoU	地面	植被	建筑物	墙面	桥梁	停车场	铁轨	道路	设施	汽车	人行道	单车	水
XYZ+RGB	90.10	53.00	80.80	97.20	91.80	52.90	40.80	43.60	6.70	56.10	35.90	79.14	33.40	0.00	69.30
XYZ+RGB+$\vec{n}$	91.90	55.90	84.60	98.20	95.90	57.60	40.50	53.90	13.60	55.90	39.60	78.88	37.80	0.50	69.70

2.3.3　MFFN 与其它分割网络的实验结果对比分析

为进一步验证文中算法在进行城市场景三维点云语义分割中的有效性和优越性,将 MFFN 与其它基于深度学习的分割方法进行了比较。表 4 列出了这些算法在 SensatUrban 数据集上的分割性能,其相应结果是由该数据集发布者在文献[21]中提供,与 MFFN 的评估方法完全一致,可以对比分析。文中算法得到的平均交并比为 55.90%,总体精度为 91.90%,表明该算法在所有类别上的评价指标上均好于其它的分割方法,证明其能够有效地提高大规模城市场景三维点云语义分割精度。

从表 4 中可以看出,文中算法对铁轨、单车两个类别的分割效果仍较差,这主要是因为铁轨和单车的训练数据太少(数据集中类别占比如表 1 所示),这使得网络无法很好地学习二者的特征,进而很难将它们精准分割;但表 1 显示在数据集中墙面和桥梁的训练数据也较少,而它们在表 4 的结果中却比铁轨和单车效果好很多,这主要是因为墙面和桥梁表现出与其它类别完全不同的法向量特征,从而获得了较高的分类性能,进一步说明了融合法向量特征对整个网络的分割精度有极大提升,由此表明融合法向量特征可有效提高大规模城市场景三维点云语义分割模型的性能。

表 4　文中方法与其它先进分割方法的实验结果对比　%

算法	OA	mIoU	地面	植被	建筑物	墙面	桥梁	停车场	铁轨	道路	设施	汽车	人行道	单车	水
PointNet[7]	80.78	23.71	67.97	89.52	80.05	0.00	0.00	3.95	0.00	31.55	0.00	35.14	0.00	0.00	0.00
PointNet++[8]	84.30	32.92	72.36	94.24	84.77	2.72	2.09	25.79	0.00	31.54	11.42	38.84	7.12	0.00	56.93
SPG[13]	85.27	37.29	69.93	94.55	88.87	32.83	12.58	15.77	15.48	30.63	22.96	56.42	0.54	0.00	44.24
SparseConv[22]	88.66	42.66	74.10	97.90	94.20	63.30	7.50	24.20	0.00	30.10	34.00	74.40	0.00	0.00	54.80
RandLA-Net[17]	89.78	52.69	80.11	98.07	91.58	48.88	40.75	51.62	0.00	56.67	33.23	80.14	32.63	0.00	71.31
MFFN (文中方法)	91.90	55.90	84.60	98.20	95.90	57.60	40.50	53.90	13.60	55.90	39.60	78.88	37.80	0.50	69.70

2.3.4　MFFN 与其它分割网络的模型参数和训练时间的对比分析

为了验证文中网络模型在内存开销和训练速度上的优越性,分别从模型参数和每轮的训练时长两方面与其它网络进行了对比,结果如表 5 所示。为了保证对比的公平性,其它方法也在训练之前进行了和文中方法相同的网格采样预处理,使训练的数据量保持一致。从表中可以看出,虽然 SPG 的模型训练参数最少,但由于其依赖于昂贵的超点图构造,反而训练时间最长;PointNet++由于在网络训练过程中采用的是最远点采样法,其训练速度远低于采用随机采样的文中方法;且文中方法在网络的分割性能明显提升的前提下,两项数据均与 RandLA-Net 几乎持平,且训练速度大幅领先于其它网络,证明了该方法十分适用于数据量庞大的城市点云。

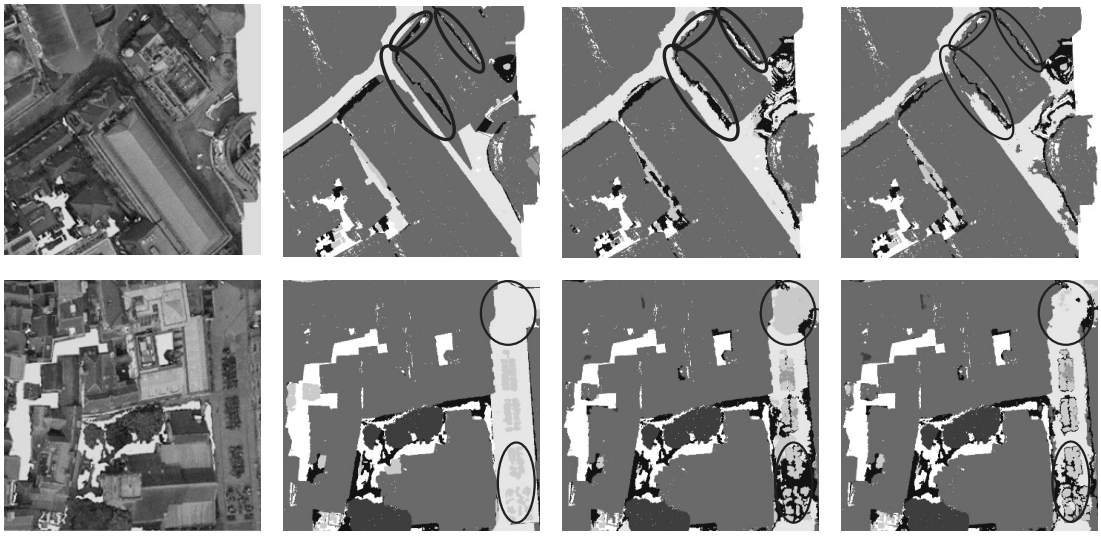
表 5　MFFN 与其它分割网络的模型参数和训练时间的对比

算法	模型参数/百万	训练时长/(秒/epoch)
PointNet++[8]	0.97	1 020.2
SPG[13]	0.25	2 520.2
SparseConv[22]	14.3	1 800.4
RandLA-Net[17]	1.24	420.1
MFFN(文中方法)	1.26	446.6

2.3.5 MFFN 与 RandLA-Net 分割结果的对比分析

图 4 为 MFFN 与 RandLA-Net 在 SensatUrban 数据集中的分割结果,由于无法获得测试集标签,该文从原数据集的训练数据中选取出合适的点云数据用于测试出图,选取的数据仅用于最终展示分割结果,并不参与模型训练,因此不会对模型性能评估产生影响。从图 4 第一行可以看出,文中方法相比于 RandLA-Net,极大程度地降低了地面、道路、人行横道三个相似度较高的类别的分割误差;从图 4 第二行可以看出,RandLA-Net 将一处道路错分为停车场,可能是该处道路与相连的其它道路颜色略有不同的原因导致出现错分,而文中方法引入的法向量特征对其进行了矫正,使错分面积明显降低。

对比其它方法,MFFN 方法具有更好的分割结果,这主要是受益于引入的法向量特征,通过多特征融合模块将点云几何特征、颜色特征与法向量进行融合,使三者特征相辅相成,降低了单一特征带来的分割误差,且对局部区域内占比较少的小型物体更加友好,有效地提高了大规模城市场景点云的分割精度。



(a)输入点云数据 (b)分割真值图 (c)RandLA-Net 分割结果 (d) MFFN 分割结果

图 4 RandLA-Net、MFFN 对 SensatUrban 数据集的分割结果

3 结束语

该文引入了点云的法向量特征,有效地弥补了几何形状与颜色特征的不足,并基于 RandLA-Net 的特征聚合思想设计了多特征局部聚合模块,将点云的法向量特征、颜色特征与几何特征进行融合,大幅提高了城市场景三维点云的分割精度。并且,在数据预处理阶段与网络训练阶段分别采用网格采样法与随机采样法进行点云降采样,保证了大规模城市点云的训练速度。在 SensatUrban 城市语义数据集上的结果显示,该算法的平均交并比为 55.90%、总体精度为 91.90%,相比其它分割网络在绝大多数类别上的分割精度均有大幅提升。但由于城市场景点云数据中物体类别分类不均衡,部分类别的占比过低,导致这些物体难以被分割,如铁轨、单车等,引入法向量后虽有所提升,但并未达到预期效果,如何解决该问题是下一步研究重点。

参考文献:

[1] GUO Y,WANG H,HU Q,et al. Deep learning for 3d point clouds:a survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2020,43(12):4338-4364.

[2] 许安宁. 基于深度学习的三维点云语义分割方法综述[J]. 长江信息通信,2021,34(1):59-62.

[3] LAWIN F J,DANELLJAN M,TOSTEBERG P,et al. Deep projective 3D semantic segmentation[C]//Computer analysis of images and patterns. [s. l.]:Springer,2017:95-107.

[4] WU B,WAN A,YUE X,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 (ICRA). Brisbane:IEEE,2018:1887-1893.

[5] IANDOLA F N, HAN S, MOSKEWICZ M W, et al. SqueezeNet:AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size[J]. arXiv:1602.07360,2016.

[6] 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). Montreal:IEEE,2019:4376-4382.

[7] QI C R,SU H,MO K,et al. Pointnet:deep learning on point sets for 3d classification and segmentation[C]//Proceedings of the 2017 IEEE conference on computer vision and pattern recognition. Honolulu:IEEE,2017:77-85.

[8] QI C R,YI L,SU H,et al. Pointnet++:deep hierarchical fea-

ture learning on point sets in a metric space[C]//Proceedings of the annual conference on neural information processing systems. Long Beach:Curran Associates,2017:5105-5114.

[9] WU W,QI Z,FUXIN L. Pointconv:deep convolutional networks on 3D point clouds[C]//2019 IEEE CVF conference on computer vision and pattern recognition (CVPR). Long Beach:IEEE,2019:9621-9630.

[10] THOMAS H,QI C R,DESCHAUD J E,et al. Kpconv:flexible and deformable convolution for point clouds[C]//Proceedings of the IEEE/CVF international conference on computer vision. Seoul:IEEE,2019:6411-6420.

[11] WANG Y,SUN Y,LIU Z,et al. Dynamic graph CNN for learning on point clouds[J]. ACM Transactions on Graphics,2019,38(5):1-12.

[12] TE G,HU W,ZHENG A,et al. Rgcnn:regularized graph CNN for point cloud segmentation[C]//Proceedings of the 26th ACM international conference on multimedia. Seoul: ACM,2018:746-754.

[13] 刘友群,敖建锋,潘仲泰. DGPoint:用于三维点云语义分割的动态图卷积网络[J]. 激光与光电子学进展,2022,59(16):209-216.

[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. Salt Lake City:IEEE,2018:4558-4567.

[15] 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. Long Beach:IEEE,2019:10296-10305.

[16] 石　敏,沈佳林,易清明,等. 快速超轻量城市交通场景语义分割[J]. 计算机科学与探索,2022,16(10):2377-2386.

[17] 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. Seattle:IEEE,2020:11108-11117.

[18] LIN Y,CHEN L,HUANG H,et al. Beyond farthest point sampling in point-wise analysis[J]. arXiv:2107.04291,2021.

[19] GROH F,WIESCHOLLEK P,LENSCH H P A. Flex-convolution:million-scale point-cloud learning beyond grid-worlds[C]//Computer vision - ACCV 2018:14th Asian conference on computer vision. Perth:Springer,2019:105-122.

[20] 李　宝,程志全,党　岗,等. 三维点云法向量估计综述[J]. 计算机工程与应用,2010,46(23):1-7.

[21] HU Q,YANG B,KHALID S,et al. Towards semantic segmentation of urban-scale 3D point clouds:a dataset,benchmarks and challenges[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Nashville:IEEE,2021:4977-4987.

[22] GRAHAM B,ENGELCKE M,VAN DER MAATEN L. 3d semantic segmentation with submanifold sparse convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Salt Lake City:IEEE,2018:9224-9232.