ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 38,No. 2,Mar. 2023,pp. 336－349
DOI:10. 16337/j. 1004-9037. 2023. 02. 010

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

学习几何结构特征的真实点云场景语义分割

李嘉祥[1]，宣士斌[1,2]，刘丽霞[1]，王　款[1]

(1. 广西民族大学人工智能学院,南宁 530006;2. 广西混杂计算与集成电路设计分析重点实验室,南宁 530006)

摘　要：有效获取点云数据在空间上的结构性特征是点云语义分割的关键。针对以往方法没有很好综合利用全局和局部特征问题,提出一种新的空间结构特征——点的盒子特征用于语义分割,设计一种编码-解码结构的网络框架,下采样过程中使用几何结构特征模块学习点云的全局空间特征和局部邻域特征,上采样过程中按分辨率逐级恢复成完整尺寸特征图进行语义分割。其中,几何结构特征模块包含两个子模块,一个是全局特征模块,该模块学习点的“盒子(box)”特征以表现点云在采样空间内概括的粗糙几何特征;另一个是局部特征模块,该模块使用特征提取——注意力机制结构表现点云在局部邻域内精确的细粒度几何特征。在公开数据集S3DIS、Semantic3D上进行了实验并与其他方法比较,实验结果表明mIoU均领先目前大部分主流的方法,部分细则类IoU取得最高。

关键词：深度学习;点云;语义分割;注意力机制;人工智能

中图分类号：TP391.41　　**文献标志码**:A

Semantic Segmentation for Real Point Cloud Scenes via Geometric Features

LI Jiaxiang[1], XUAN Shibin[1,2], LIU Lixia[1], WANG Kuan[1]

(1. College of Artificial Intelligence, Guangxi University for Nationalities, Nanning 530006, China;2. Key Laboratory of Hybrid Computing and Integrated Circuit Design and Analysis, Nanning 530006, China)

Abstract: Effective acquisition of spatial structural features of point cloud data is the key to semantic segmentation of point clouds. To solve the problem that the previous methods do not make good use of global and local features, a new spatial structure feature, point box feature, is proposed for semantic segmentation. A network framework of encoding-decoding structure is designed. The global spatial and local neighborhood features of point clouds are learned by using the geometric structure feature module during the downsampling process, and the full size feature map is restored step by step in the upper sampling process for semantic segmentation. The geometric structure features module contains two sub-modules, one is the global features module, which learns the “box” features of points to represent the rough geometric features of point clouds in the sampling space. Another is the local features module, which uses feature extraction, the attention mechanism structure, to represent precise, fine-grained geometric characteristics of point clouds within local neighborhoods. Experiments are performed on the public dataset S3DIS and Semantic3D and compared with other methods. The results show that mIoU is ahead of most of the current mainstream methods, and some of the detail class IoU is the highest.

Key words: deep learning; point cloud; semantic segmentation; attention mechanism; artificial intelligence

---

基金项目:国家自然科学基金(6186603)。
收稿日期:2022-04-19;**修订日期**:2022-06-28

引　言

近年来深度学习在计算机图像处理领域得到广泛应用,包含三维点云语义分割。无论是无人驾驶还是增强现实、虚拟现实等领域,都要求大规模真实场景的点云语义分割技术支持,为了促进这些领域的发展,对大规模真实场景的点云语义分割研究必不可少。如图1、2所示,对比于传统的二维图像语义分割,三维点云语义分割存在着离散化、无结构性和无序性等问题,这些问题使得应用于一般的二维图像的分割方法难以直接应用于点云数据,例如常用于二维图像的卷积操作,如图3所示,图中左侧代表二维图像的卷积,右侧代表某种点云卷积,深色点代表卷积中心,浅色部分代表卷积的区域和卷积的点(像素),可以看出二维图像卷积有结构,而点云的卷积则无结构特征。但点云数据具有二维图像所没有的丰富的空间位置信息,因此根据点云的空间位置信息学习空间几何结构特征才能有效进行点云语义分割,如何学习到有效的空间几何结构特征则是关键。

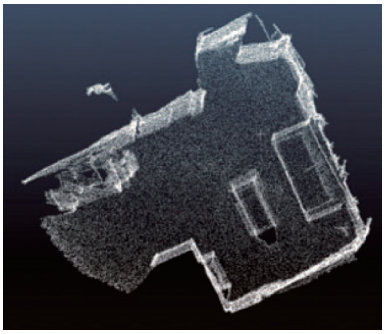

图1　未标注的点云数据俯视图

Fig.1　Unlabeled point cloud data top view

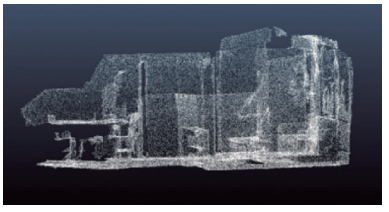

图2　未标注的点云数据侧视图

Fig.2　Unlabeled side view of point cloud data

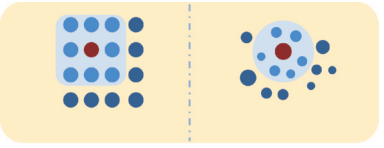

图3　2D图像卷积与3D点云卷积对比

Fig.3　Comparing 2D image convolution with 3D point cloud convolution

为了解决这一问题,Charles等[1]提出了PointNet,用k最近邻法(k-nearest neighbor,KNN)[2]对点云数据进行分组,并对每一组进行卷积,首次实现了直接对点进行卷积的网络。在PointNet的基础之上,许多研究者的工作主要在3个方面。第一,针对点云卷积核的设计,如PointConv[3]、KPConv[4]、PAConv[5]等。这类方法设计卷积核使用可适应性或者可学习的权重构造卷积核的权重参数,成功把二维图像的卷积操作运用到了三维点云中,并且在多个点云分类数据集的测试上也取得较好的结果,但是在点云的语义分割中的表现相对较差,主要是点云数据在本质上和二维图像的不同之处在于点云数据具有空间几何信息和空间结构信息。如果没有充分利用到这些信息即使使用二维图像的卷积操作也难以在点云语义分割上面取得较好的结果。另外,在大规模真实点云场景的语义分割中,需要处理的点云数据量十分巨大,学习权重库多个权重矩阵是会带来巨大的时间花费。第二,针对点云采样方法设计,如PointNet[1]、RandLA-Net[6]等。由于传统PointNet[1]的最远点采样时间花费大,不适合大规模真实场景的点云语义分割。为了把更多的计算机时间空间应用于对点云直接的语义分割,本文采用RandLA-Net[6]提出的随机采样法。第三,针对点云卷积的特征设计,如PointNet++[7]、DGCNN[8]、PointWeb[9]、SCF-Net[10]等,点的空间坐标、邻域内坐标、点与中心点的距离等是这类方法常用特征,这类方法没有区分全局背景信息和局部背景信息。SCF-Net[10]在局部上利用了中心点与邻域点的角度信息,在全局上仅利用了局部邻域与全局采样空间体积比信息,没有充分利用全局背景信息可能导致的问题是相邻点特征权重过大无法学习到整体几何结构特征。例如在同一采样空间里的两个“桌子(Table)”一个能够正确分类,而另一个却被误判为“书架(book)”。为解决以往方法未充分利用全局背景信息,本文根据点邻域内点的坐标极值构造一个长方体,称之为“盒子(box)”,并通过盒子的顶点、全局角度、体积比等特征,用于

获取全局空间背景信息,与此法最为相近的是RandLA-Net[6],二者都使用随机采样方法进行大规模真实场景点云语义分割,但是本文方法设计了一个全局特征提取层利用点的盒子特征。

提出的网络模型使用编码-解码结构作为主框架,在下采样中提出一种新几何结构特征模块,它由两个子模块组成:(1)局部特征提取模块,对输入的点云邻域进行处理然后输出点云邻域内的多种特征,用于获取局部邻域背景信息;(2)全局特征提取模块,用于获取全局空间背景信息。在上采样过程中,通过一个残差块连接相对应的同等分辨率大小的特征图和前面一层通过最邻近插值法生成下采样特征图,以此类推逐级恢复到完整尺寸的特征图,最后传入一个全连接层进行语义分割。

本文的主要贡献如下:

(1)提出了一个全局特征提取模块,根据点的盒子特征来获得全局背景信息,扩大点在全局三维空间上的特征差异。

(2)提出了一个几何结构特征模块,把点云的特征分为全局特征和局部特征来处理,能够兼顾全局和局部的几何结构特征表现。

1 相关工作

根据对三维点云数据处理方式的不同,主流的点云语义分割方法大致划分为:基于投射的网络、基于体素的网络、基于点的网络,而使用自注意力的方法在点云语义分割也越来越普遍。

1.1 基于投射、体素的网络

为了将卷积神经网络应用于三维点云语义数据,人们设计一种基于投射的网络,将点云数据投射到不同视角下的平面转换为二维图像形式,再按照二维图像的卷积方法进行处理,处理后的数据聚合成点云数据,这种方法能够像处理二维图像一样处理三维点云数据。如MV3D模型[11]通过前视图和俯视图处理点云数据,Li等[12]把输入点云通过投影函数投影为二维图再使用全卷积神经网络进行处理。基于体素的网络则是通过将点云转换为一种体素形式的数据[13-14],再将这种体素数据通过一种三维卷积神经网络进行处理,然而这些方法在进行数据转换时不可避免地带来巨大的开销,并且在数据转换之后,会丢失点云原有的一些细节信息。

1.2 基于点的网络

为了避免点云数据在转换时造成的信息丢失和额外花费,一种基于点的网络被提出用于直接对点云数据进行处理,PointNet[1]采用一种适用于直接对点云数据进行卷积的方法,随后PointNet++[7]等方法进一步对PointNet进行改进。基于点的网络能方便快捷处理点云数据,相对于投射、体素的方法有一定优越性,但是这种方法所学习的特征不完整,并且1×1的卷积核也显得不足以模拟二维的卷积过程,为此对于卷积核的设计也是一个改进方向,此外它所使用的最远点采样方法需要一定的代价[1,7]。

1.2.1 卷积核设计

为了直接对点云数据进行卷积,需要设计特定的卷积核。PointConv[3]直接预测核权重基于相关的局部信息来设计卷积核。KPConv[4]连接核权重跟固定的核点并用相关函数去调整核的权重。PA-Conv[5]通过使用得分网络去计算权重库里每一个核的得分,再累加起来得到一个动态可适应性核。这些卷积核的设计方法都可以直接对点云数据进行卷积操作,然而无论是哪一种卷积核都不能应用于所有的点云场景,并且这些卷积核无法学习到全局信息,即使是自适应学习到的卷积核卷积处理得到的特征或许还不如直接地手动设计的特征提取。

1.2.2 卷积特征设计

只有有效的提取点云数据在空间上几何结构特征才能解决点云数据的稀疏性、无结构性等问题,

从而提高分割的准确率。为此,PointNet[1]主要学习的特征是点的坐标组成; DGCNN[8]则是学习点与邻域点之间连接边的特征;PointWeb[9]提取邻域内每一点与其他点连接的边作为特征;RandLA-Net[6]通过合并点的坐标、相对坐标、邻域点与中心点之间的距离作为特征;SCF-Net[10]在RanLA-Net基础上进一步利用点之间的角度信息作为特征。这些方法都是通过设计专门三维空间点的特征表示,对点云数据的语义分割针对性较强。但这类方法大多以局部邻域为单位学习卷积特征,缺乏学习全局采样空间下的特征。受此启发,本文提出一种不仅能够有效学习局部细粒度特征信息,还能学习全局采样空间下特征的新方法。

1.2.3 采样方法设计

由于三维点云不能直接通过卷积进行下采样,所以在PointNet[1]提出最远点采样方法之后,一些新的基于点的网络被提出。如:RandLA-Net[6]中的随机采样方法可大幅度减少点云采样过程的时间花费,从而能有效应用于大规模真实点云场景分割邻域。为解决点云数据没有固定拓扑结构问题,球邻域法[15]被提出,该方法根据点的局部密度以半径为r的球为单位进行采样,具有较强的适应性。刘鹏等[16]提出基于自适应动态球半径的k邻域搜索算法,根据周围点云密度调整球半径提高搜索效率缩短搜索时间;欧阳峻岭[17]通过改进球邻域算法中的空间栅格划分和设定搜索极限进一步提高搜索效率、克服搜索过程中可能陷入死循环等问题。这类方法需要事先指定球半径r,对不规则点云数据会导致圆内点数过多或过少且对于边缘数据搜索效率低鲁棒性差等问题。受以上方法启发,本文根据“最小包围盒”[15]概念提出一种“盒子”特征用于点云语义分割,点云数据组成的物体都可以表示为几何体的组合,例如:桌子可以看作是一个正方体桌面和4个长方体桌脚组成。与球邻域法使用的球体邻域相比,本文使用的长方体邻域有更强的表现力,因为长方体具有丰富的边、角特征有助于模型学习空间几何结构特征克服点云数据的无序性、不规则性问题。

1.3 使用自注意力方法

自从Vaswani等[18]提出Transformer模型以来,机器翻译和自然语言处理得到了飞快的发展,许多点云邻域的研究者开始将Transformer引入到点云邻域[6,19-20]。而注意力机制作为Transformer模型的关键一环受到了很多的关注,一些方法使用注意力机制来改进它的性能。Yang等[19]提出了一种在点之间插入注意力模块,Chen等[20]提出了一种利用注意力去学习空间分布权重和获取局部几何结构的方法,而RandLA-Net[6]提出了一种特征提取——注意力机制的结构,在一个特征提取模块提取到特征之后输入到一个注意力模块去学习特征间的注意力权重并更新特征值,在本文提出的模型中也引用这种方式。

2 方 法

本节主要介绍本文提出的模型和方法,该模型使用U-Net结构的网络[21]。算法流程如图4所示,模型总的结构如图5所示,上半部分是下采样阶段,下半部分是上采样阶段,其中虚线代表着残差连接。首先输入的点云数据经过一个全连接层和5个下采样层进行多分辨率的特征提取,如果输入的点的数量为N,特征数量为8,那么每一个下采样层都会进行一次随机采样(Random sampling, RS)随机选取$N/4$个点以节省由采样所带来的时间开销,然后经过几何结构特征模块(Geometric structure feature module, GSFM)学习点云全局背景信息和局部邻域信息以构建空间几何结构提高模型的语义分割能力,每一层输出维度为:16、64、128、256、512,最后经过一个多层感知机(Multilayer perceptron, MLP)和5个上采样层通过残差连接对应维度上采样层的结果逐级恢复到原始尺寸的特征图,保留了网络低层

信息和高层信息又使它们有效融合,最后传入到3个带有批标准化和ReLU函数的全连接层进行分类输出语义分割结果,每个全连接层后都添加了dropout层避免过拟合。其中在下采样阶段使用的几何结构特征模块如图6所示,输入的点云数据主要经过两个模块处理:全局特征提取层和局部特征提取层。左边的全局特征提取层以点云邻域为单位学习它的全局空间几何结构信息,中间的局部特征提取层以邻域内的每一个点为单位学习它在局部邻域内的几何结构信息并且有残差连接在右边,这两个层输出的结果会串联后通过一个LeakyReLU层提取到特征,与一般U-Net结构不同的是主要利用了“盒子”特征,因此称之为“Box-Net”。

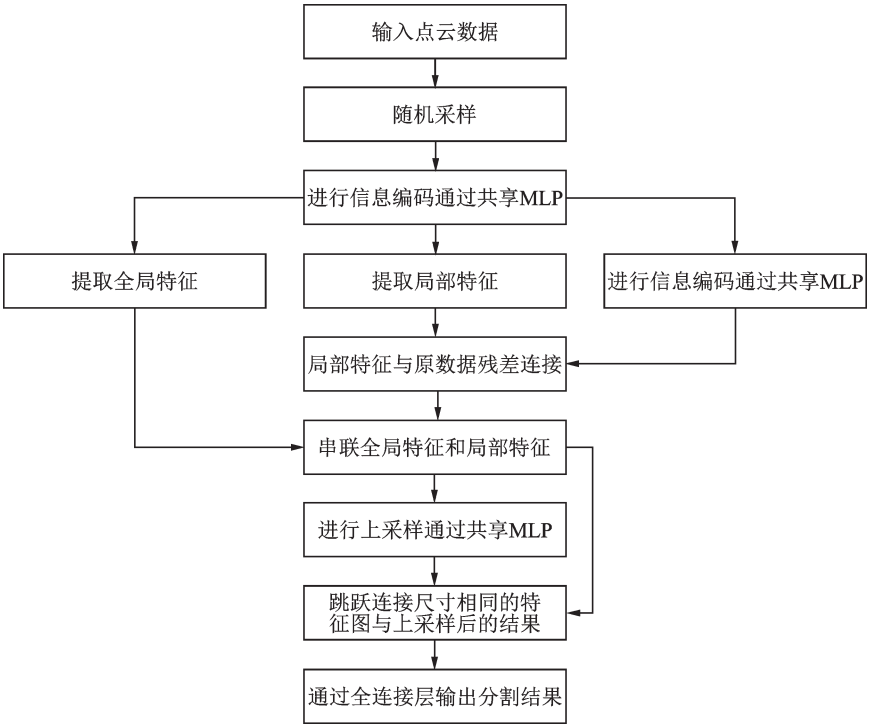

图4 算法流程图

Fig.4 Algorithm flow chart

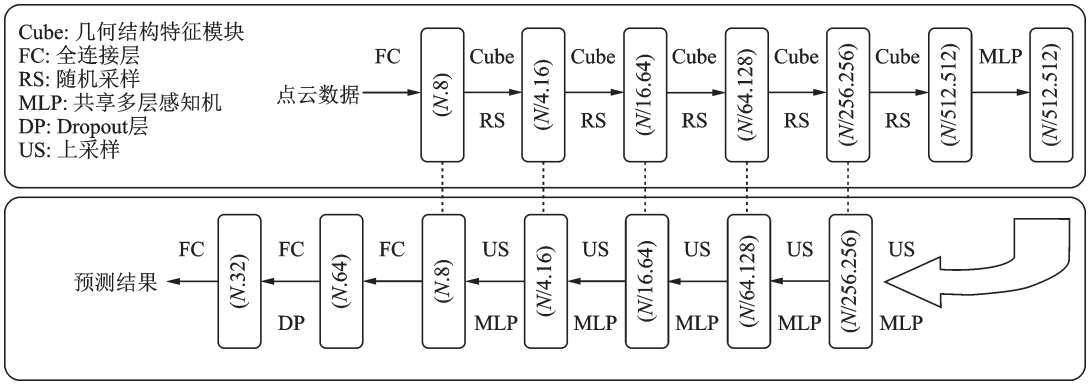

图5 Box-Net结构

Fig.5 Structure of Box-Net

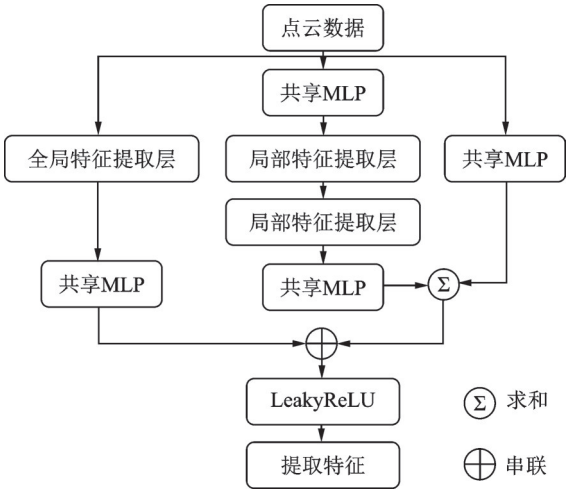

图6 几何结构特征模块

Fig.6 Geometric structure feature module

2.1 全局特征提取层

已有方法很少明确地把点云特征区分为全局、局部特征来分开处理。SCF-Net提出了全局特征提取层[10],但是仅仅提取了邻域体积和全局点云体积比这一特征,过于侧重邻域中心点与邻域内其他点间的局部结构特征,导致网络模型难以学习全局采样空间下完整物体结构,主要表现为同一场景下两个相同物体一个正确分类一个错误分类。受球邻域法启发本文提出全局特征提取层,利用长方体的几何属性提取了点云局部邻域多种区别性的特征,点云数据是物体表面信息,与物体长轴方向一致的长方体更能反映物体表面特征信息。若使用球邻域法的球体则结构相对简单难以提高语义分割准确率,因为球邻域法更多用于点云搜索、配准要求提高时间效率为主。对于输入的点云数据,每一个点都会带有x轴、y轴、z轴坐标和RGB颜色特征等,这些点首先会通过KNN方法进行分组,分组后的每一组称之为点的局部邻域。

每一个以点p_i为中心的点通过KNN方法分组后的局部邻域有k个点$\left\{ p_i^1, p_i^2, \cdots, p_i^k \right\}$,每一个局部邻域都是存在于一个长方体空间里,称之为“盒子”,因此每一片点云的全局特征记为“Box”,使用式(1)将它们进行编码以获得它们的全局特征。

$$\mathrm{Box} = \mathrm{MLP}\left(p_i \oplus s \oplus \alpha \oplus \beta \oplus \rho \oplus V \right) \tag{1}$$

式中:MLP(Multilayer perceptron)表示一个多层感知机,$\oplus$表示串联操作,p_i为第i个点的x、y、z轴坐标,s为一个向量表示盒子的长、宽、高,α为盒子几何中心点与整个点云三维空间中心点的连线在xoy平面上的角度,β为盒子几何中心点与整个点云三维空间中心点的连线与投影线在空间上的角度,ρ为盒子几何中心点到整个点云三维空间中心点的距离,V代表盒子的体积与整个点云三维空间的体积比例。

点的局部邻域所处在的盒子空间即邻域内以点p_i为中心的k个点中x、y、z坐标的最大值减去最小值所形成的边连接而成的一个长方体,记为式(2),其中x_i^k、y_i^k、z_i^k分别为p_i^k的x、y、z轴坐标。

$$s = \left(\left(\max\left(x_i^k \right) - \min\left(x_i^k \right) \right), \left(\max\left(y_i^k \right) - \min\left(y_i^k \right) \right), \left(\max\left(z_i^k \right) - \min\left(z_i^k \right) \right) \right) \tag{2}$$

记盒子的几何中心点为式(3),利用空间几何的中点公式计算出盒子的几何中心点p_{bc}坐标代表着点云邻域在三维空间中的位置信息,式(4)通过计算盒子中心点p_{bc}到全局点云空间中心点的距离可获

得点云邻域在空间中的距离信息，$\|\cdot\|$表示欧氏几何距离。

$$p_{bc}=\left(\frac{\max\left(x_i^k\right)+\min\left(x_i^k\right)}{2},\frac{\max\left(y_i^k\right)+\min\left(y_i^k\right)}{2},\frac{\max\left(z_i^k\right)+\min\left(z_i^k\right)}{2}\right) \tag{3}$$

$$\rho=\left\|p_{bc}\right\| \tag{4}$$

在计算盒子的几何中心点p_{bc}后，根据空间几何知识并按式(5,6)进一步可得出盒子中心点p_{bc}与全局点云空间中心点的偏移角度α和β。

$$\alpha=\arctan\left(\frac{y_{bc}}{x_{bc}}\right) \tag{5}$$

$$\beta=\arctan\left(\frac{z_{bc}}{\sqrt{x_{bc}^2+y_{bc}^2}}\right) \tag{6}$$

式(7)以盒子空间内距离盒子中心点最远点的欧氏距离的三次方作为盒子的体积V_l，而式(8)距离坐标原点最远点的欧氏距离的三次方作为全局点云空间的体积V_g，式(9)即盒子体积占全局空间体积的比重代表着局部邻域在全局空间的体积信息。在采样空间下提取的全局特征如图7所示，图中长方体中心虚线连接原点的点为邻域中心点，以此点作为中心点使用KNN算法搜索最近16个点，以这些点的坐标最值构造最小包围盒即图中长方体，提取长方体几何特征进行学习解决全局采样空间下物体几何结构表达不足问题，对比以往方法学习点、点和邻域点的特征捕抓到点、线(点与邻域点连接而成)特征，盒子特征能够捕抓到点、线、体(点与邻域点组成)特征改进模型学习全局几何结构的能力。

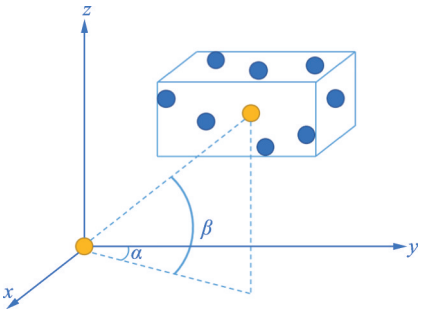

图7 点的全局特征提取

Fig.7 Global feature extraction of points

$$V_l=\left(\max\left\|p_i-p_i^k\right\|\right)^3 \tag{7}$$

$$V_g=\left(\max\left\|p_i\right\|\right)^3 \tag{8}$$

$$V=\frac{V_l}{V_g} \tag{9}$$

2.2 局部特征提取层

以往的方法大多把点云局部邻域看作球体空间，而本文把点云局部邻域看作盒子空间，局部特征提取层在点云盒子空间内对每一个点进行特征提取，提取得到各种点云数据特征串联一起进行注意力池化操作。点云数据从本质上适合于注意力机制，注意力机制提出用于解决自然语言处理中一个句子里面不同的单词对其他单词的影响力权重问题，那么也可以把局部特征提取层提取到的特征看作是一个句子，提取到的邻域内的n个特征可以看作是句子里面的n个单词，就可以像自然语言处理计算句子内不同单词之间的影响权重那样去计算邻域内不同特征之间的影响权重。图8展示了局部特征提取层。

2.2.1 提取局部特征

RandLA-Net及其之前的许多方法把x、y、z轴坐标输入模型进行学习，忽略了真实场景中z轴比x、y轴更具有代表性，例如两张相同的桌子具有相同z轴坐标而x、y轴坐标不同，桌子和电风扇具有不同的z轴坐标而相同的x、y轴坐标，因此本文使用具有z轴旋转置换不变性的局部特征提取层解决此问题，设计式(10)使得点绕z轴旋转任意角度后仍能提取同样的特征。

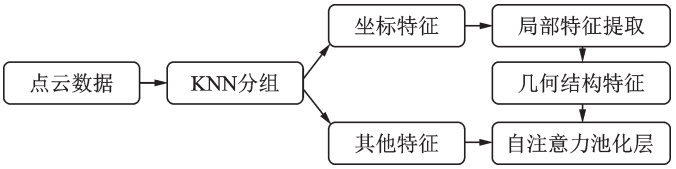

图8 局部特征提取层

Fig.8 Local feature extraction layer

$$F_1 = \mathrm{Concat}(p_i \oplus p_i^k \oplus l_a \oplus l_b \oplus \| p_i - p_i^k \| \oplus f_i^k \oplus f_d) \tag{10}$$

记提取到点云邻域内的特征为"F_1",通过串联多种不同局部特征可以得到F_1,其中p_i是第i个点的x、y、z轴坐标,p_i^k是以第i个点为中心的局部邻域内的第k个点的x、y、z轴坐标,使用球坐标转换公式(11,12)计算邻域内每一个点与中心点的偏移角度l_a和l_b,式(13)中$\| p_i - p_i^k \|$表示局部邻域内每一个点到中心点之间的欧氏几何距离,f_i^k表示邻域内每一个点的原始特征,式(14)中f_d代表着邻域内点的特征与中心点特征的平均差值通过一个负指数函数作为激活函数。

$$l_a = \arctan\left(\frac{y_i^k}{x_i^k}\right) \tag{11}$$

$$l_b = \arctan\left(\frac{z_i^k}{\sqrt{\left(x_i^k\right)^2 + \left(y_i^k\right)^2}}\right) \tag{12}$$

$$\| p_i - p_i^k \| = \sqrt{\left(x_i - x_i^k\right)^2 + \left(y_i - y_i^k\right)^2 + \left(z_i - z_i^k\right)^2} \tag{13}$$

$$f_d = \exp\left[-\mathrm{mean}\left(\left| f_i - f_i^k \right|\right)\right] \tag{14}$$

2.2.2 自注意力池化层

提取到局部特征后关键问题是如何充分利用这些特征去训练网络模型使其更有效学习,设计一个自注意力池化层计算提取到的局部特征,定义n个不同的特征如下形式:$F_l^n = \left\{ f_l^1, f_l^2, \cdots, f_l^n \right\}$,式(15)通过一个全连接层作为MLP接着softmax激活函数得到这些特征的注意力得分。

$$W^n = \mathrm{softmax}\left(\mathrm{MLP}\left(F_l^n\right)\right) \tag{15}$$

式中:W^n为学习到的权重向量,代表着n个特征的注意力得分。式(16)根据注意力得分调整原特征,凸显更为关键的特征。

$$\widetilde{F_l^n} = \mathrm{MLP}\left(\sum_{n=1}^{n}\left(F_l^n \cdot W^n\right)\right) \tag{16}$$

总而言之,对于一个输入的点云P的第i个点p_i,局部特征提取层以点p_i为中心的局部邻域内k个最邻近点为单位学习它的几何样式和特征,最后利用自注意力池化层筛选出关键的特征扩大特征差异提高语义分割的识别度。

2.3 上采样

经过前面多层的几何结构特征模块对输入的点云数据进行下采样后输出5个不同尺寸的特征图,然后使用最邻近插值法方法残差连接这些不同尺寸的特征图,既保留了上采样后的特征细节信息又融合了高层信息和低层信息,按分辨率从小到大重建完整尺寸的预测结果图。

上采样方法是每一个查询点使用KNN算法搜索一个最邻近点，然后使用最邻近插值法进行上采样，最后再传入3个全连接层对点云特征与语义分割结构建转换关系。

3 实　　验

实验数据集选用两个大型的公开数据集：S3DIS（StanfordLarge-scale 3D indoor spaces）数据集[22]、Semantic3D数据集[23]上进行测试。在使用Quadro RTX8000显卡和Python 3.5，Tensorflow 1.11，CUDA 9.0，cuDNN v7在Ubuntu 16.04的服务器上进行所有测试。在S3DIS数据集上与9种不同方法的结果进行对比，在Semantic3D数据集上与10种不同的方法的结果进行对比。

实验中使用Adam优化器并设置初始学习率为10^{-2}，训练的batch为4，训练100个周期，使用k值为16的KNN方法对点云进行聚类分组，以上参数在RandLA-Net[6]中经由实验证明为最佳，且为与RandLA-Net[6]比较显示本文方法的优越性，故不调整以上参数。

3.1 评价指标

实验使用平均交并化（Mean intersection over union，mIoU）、平均类间准确率（Mean class accuracy，mAcc）和总体准确度（Overall accuracy，OA）作为评价指标。其中mIoU为

$$\mathrm{mIoU}=\frac{1}{k}\sum_{i=1}^{k}\frac{p_{ii}}{\sum_{j=1}^{k}p_{ij}+\sum_{j=1}^{k}p_{ji}-p_{ii}} \tag{17}$$

式中：k为实验数据的类别总和，p_{ii}代表输出结果为i类真实类别为i类，p_{ij}代表输出结果为i类真实类别为j类，p_{ji}代表输出结果为j类真实类别为i类。OA计算公式为

$$\mathrm{OA}=\frac{n}{N} \tag{18}$$

式中：n表示正确分类的样本数，N表示所有样本的总数。而mAcc计算公式为

$$\mathrm{mAcc}=\frac{1}{k}\sum_{i=1}^{k}\frac{p_{ii}}{\sum_{j=1}^{k}p_{ij}} \tag{19}$$

mIoU值、OA值、mAcc值越大，分割效果越好。

3.2 S3DIS数据集

S3DIS[22]是一个大型的室内真实场景点云数据集，它包含了6个分区和271个空间。每一篇点云都是一个中等尺寸的空间并且每一个点都被标注为13个类中的一个。使用6折交叉验证法（6-fold cross validation）去测试本文提出的方法。图9～11展示了S3DIS数据集上实验的结果与真实标注对比图。实验结果分别如表1、2所示，本文方法领先于大部分目前主流方法。在表1中，本文提出的方法比原方法RandLA-Net[6]和同样是改进RandLA-Net[6]的方法SCF-Net[10]在9个细类下IoU更高，在4个IoU更低的细类差距不超过0.9%，本文提出的方法在房梁（Beam）、桌子（Table）、沙发（Sofa）、木板（Board）、杂物（Clutter）这几个细类取得领先结果，这些细类的mIoU分别是65.0%、73.5%、70.5%、67.5%、62.2%，分别提高了0.1%、1.9%、1.2%、1.6%、1.3%，这几个细类能够用几个简单的长方体穿插组合而成，从全局采样空间来看具有显著性几何结构特征，而“盒子”特征提取模块能够有效地提取点云数据的几何结构，在墙（Wall）、柱子（Col.）上表现不佳，因为点云数据的采集大多用激光雷达传感器发射激光接触到物体表面随后反射回点云数据，点云中的墙（Wall）仅是二维平面而不是现实生活中的长方体，“盒子”特征针对大部分具有立体几何特征的点云物体必然导致难以捕抓二维特征，柱子（Col.）在S3DIS数据集中仅占1.5%，表中所有方法对柱子（Col.）分割IoU最高仅54.3%，说明目前方法难以识别少样本类别。表2比较了多种方法的mIoU、OA、mAcc，本文的方法比相似的方法RandLA-Net[6]

(2020年)mIoU提高了2.2%、OA提高了0.6%、mAcc提高了1.0%,比同样是大规模真实点云场景分割领域的先进方法SCF-Net[10](2021年)mIoU提高了0.6%、OA提高了0.2%、mAcc提高了0.3%,主要是结合了点云的全局和局部特征,尤其是在全局特征方面进一步学习了其空间结构。

图9 S3DIS数据集Area 6部分数据(未标注)可视化

Fig.9 S3DIS dataset Area 6 partial data (unlabeled) visualization

图10 S3DIS数据集Area 6部分数据(真实标注)可视化

Fig.10 S3DIS dataset Area 6 partial data (true background) visualization

图11 S3DIS数据集Area 6部分数据(Box-Net标注)可视化

Fig.11 S3DIS dataset Area 6 partial data (Box-Net annotated) visualization

表1 S3DIS数据集上细则类的结果

Table 1 Results of detail classes on 3DIS dataset

Method	Ceil.	Floor	Wall	Beam	Col.	Wind.	Door	Table	Chair	Sofa	Book	Board	Clutter
PointNet[1]	88.0	88.7	69.3	42.4	23.1	47.5	51.6	54.1	42.0	9.6	38.2	29.4	35.2
RSNet[24]	92.5	92.8	78.6	32.8	34.4	51.6	68.1	59.7	60.1	16.4	50.2	44.9	52.0
3P-RNN[25]	92.9	93.8	73.1	42.5	25.9	47.6	59.2	60.4	66.7	24.8	57.0	36.7	51.6
SPG[26]	89.9	95.1	76.4	62.8	47.1	55.3	68.4	**73.5**	69.2	63.2	45.9	8.7	52.9
PointCNN[27]	**94.8**	**97.3**	75.8	63.3	51.7	58.4	57.2	71.6	69.1	39.1	61.2	52.2	58.6
PointWeb[9]	93.5	94.2	80.8	52.4	41.3	64.9	68.1	71.4	67.1	50.3	62.7	62.2	58.5
ShellNet[28]	90.2	93.6	79.9	60.4	44.1	64.9	52.9	71.6	**84.7**	53.8	64.6	48.6	59.4
KPConv[4]	93.6	92.4	**83.1**	63.9	**54.3**	**66.1**	**76.6**	57.8	64.0	69.3	**74.9**	61.3	60.3
RandLA-Net[6]	93.1	96.1	80.6	62.4	48.0	64.4	69.4	69.4	76.4	60.0	64.2	65.9	60.1
SCF-Net[10]	93.3	96.4	80.9	64.9	47.4	64.5	70.1	71.4	81.6	67.2	64.4	**67.5**	60.9
Ours	93.7	96.0	81.4	**65.0**	50.2	65.3	69.6	**73.5**	80.7	**70.5**	64.1	**67.5**	**62.2**

表2 S3DIS数据集上6文件交叉测试的各种不同的结果

Table 2 Quantitative results of different approaches on the S3DIS dataset (6-fold cross validation)

Method	mIoU/%	OA/%	mAcc/%
PointNet[1]	47.6	78.6	66.2
RSNet[24]	56.5		66.5
3P-RNN[25]	56.3	86.9	
SPG[26]	62.1	86.4	73.0
PointCNN[27]	65.4	88.1	75.6
PointWeb[9]	66.7	87.3	76.2
ShellNet[28]	66.8	87.1	
KPConv[4]	70.6		79.1
RandLA-Net[6]	70.0	88.0	82.0
SCF-Net[10]	71.6	88.4	82.7
Ours	**72.2**	**88.6**	**83.0**

3.3 Semantic3D数据集

Semantic3D数据集[23]来自对许多乡村和城市景观扫描生成的真实场景三维点云数据,总共包含了划分为8个类别大概400亿个点,每个点都有空间坐标和RGB颜色特征。Semantic3D数据集的真实标注结果没有公开,需要在Semantic3D数据集上进行语义分割实验并且把预测的结果上传到Semantic3D数据集的官方网站上进行评测,在Semantic3D数据集的官方网站上有两种不同的评测方法,一种是使用规模较小省略版的语义标注结果进行评测即是“Reduce-8”,另一种是把所有完整的语义标注结果进行评测即是“Semantic-8”,在表3和表4使用“Semantic-8”的方法进行评测并和大部分其他主流的方法进行对比结果证明本文的方法能够取得最好的效果。实验结果可视化如图12所示。具体每一类的mIoU在表3中展示,可以看出在人造的地形(man-made terrain)、建筑物(buildings)和汽车(cars)这3个方体特征比较明显的细类中,mIoU取得了最好的结果分别是96.1%、96.1%、93.1%,分别提高了0.1%、0.2%、3.8%,而在和其他方法相比较之下,在高植被(high vegetation)、矮植被(low vegetation)和扫描伪影(scanning artefacts)这些不规则物体中表现得更差,原因是这几类的点云数据占比相对较低并

表3 Semantic3D数据集上细则类的结果

Table 3 Results of Semantic3D (semantic-8) detail class

Method	Man-made terrain	Natural terrain	High vegetation	Low vegetation	Buildings	Hard scape	Scanning artefacts	Cars
TMLC-MS[29]	91.1	69.5	32.8	21.6	87.6	25.9	11.3	55.3
EdgeConv-PN[30]	91.2	69.8	51.4	58.5	90.6	33.0	24.9	68.6
PointNet++[7]	81.9	78.1	64.3	51.7	75.9	36.4	43.7	72.6
SnapNet[31]	89.6	79.5	**74.8**	56.1	90.9	36.5	34.3	77.2
PointConv[3]	92.2	79.2	73.1	62.7	92.0	28.7	43.1	82.3
PointGCR[32]	93.8	80.0	64.4	**66.4**	93.2	39.2	34.3	85.3
PointConv-CE[33]	92.4	79.6	72.7	62.0	93.7	40.6	**44.6**	82.5
RandLA-Net[6]	96.0	**88.6**	65.3	62.0	95.9	**49.8**	27.8	89.3
Ours	**96.1**	87.3	64.3	58.4	**96.1**	41.8	37.9	**93.1**

且这几类缺乏显著的几何结构特征,说明相对于RandLA-Net,“盒子”特征可以更好应用于城市多建筑、汽车的场景,在室内场景(S3DIS)数据集上表现优异,在室外场景(Semantic3D)表现效果也不降低。在表4与大部分主流的方法进行了对比,可以看出本文提出的方法mIoU最高,OA虽然次优,但在语义分割方面从目前主流的观点来看,相对于OA,mIoU有更高的可靠性,因此本文提出的方法优于其他点云语义分割方法。

表4 Semantic3D数据集上Semantic-8测试的结果

Table 4 Results of Semantic-8 test on Semantic3D dataset

Method	mIoU/%	OA/%
TMLC-MS[29]	49.4	85.0
EdgeConv-PN[30]	61.0	89.4
PointNet++[7]	63.1	85.7
SnapNet[31]	67.4	91.0
PointConv[3]	69.2	91.8
PointGCR[32]	69.5	92.1
PointConv-CE[33]	71.0	92.3
RandLA-Net[6]	71.8	**94.2**
Ours	**71.9**	93.4

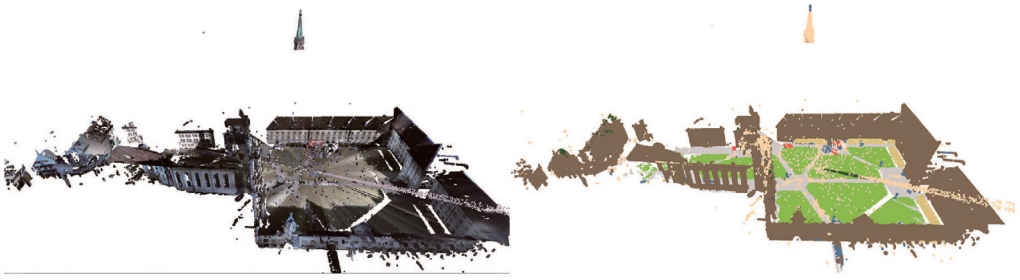

图12 Semantic3D数据集(左)与Box-Net标注结果(右)可视化

Fig.12 Visualization of Semantic3D dataset and Box-Net annotation results

4 结束语

本文提出了一个使用了几何结构特征模块的模型,其中首先提出使用“盒子”特征作为全局特征并能有效综合局部特征,应用于大规模真实场景点云语义分割并在多个数据集上取得了优秀的结果。提出的几何结构特征模块对点云数据进行特征提取,有效地获取点云的区别性特征,解决点云数据的无序性、无结构性等问题,主要是因为此方法根据空间几何学上的数学原理提取几何结构特征,对于解决目前大规模真实场景点云语义分割网络面临的缺乏全局几何结构信息具有重要意义和实际价值,而在局部邻域使用自注意力机制确保网络具有特征权重判断能力。在上采样过程使用了最邻近插值法对下采样得到的特征图进行复原,能够简单快捷地恢复完整尺寸特征图进行最后的语义分割。在S3DIS数据集和Semantic3D数据集上进行了测试,并与其他主流方法进行了对比,证明了本文提出的方法的有效性。未来计划把这种方法推广到其他邻域,例如:点云分类、点云目标识别等。

参考文献：

[1] CHARLES R Q, SU H, KAICHUN M, et al. PointNet: Deep learning on point sets for 3D classification and segmentation [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). [S.l.]: IEEE, 2017: 77-85.

[2] COVER T, HART P. Nearest neighbor pattern classification[J]. IEEE Transactions on Information Theory, 1967, 13(1): 21-27.

[3] WU W, QI Z, FUXIN L, PointConv: Deep convolutional networks on 3D point clouds[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). [S.l]: IEEE, 2019: 9613-9622.

[4] THOMAS H, QI C R, DESCHAUD J E, et al. KPConv: Flexible and deformable convolution for point clouds[C]//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). [S.l]: IEEE, 2019: 6410-6419.

[5] XU M, DING R, ZHAO H, et al. PAConv: Position adaptive convolution with dynamic kernel assembling on point clouds [C]//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). [S.l]: IEEE, 2021: 3172-3181.

[6] HU Q, YANG B, XIE L, et al. RandLA-Net: Efficient semantic segmentation of large-scale point clouds[C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). [S.l.]: IEEE, 2020: 11105-11114.

[7] RUIZHONGTAIQI C, YI L, SU H, et al.PointNet++: Deep hierarchical feature learning onpoint sets in a metric space[C]//Proceedings of Advances in Neural Information Processing Systems (NeurIPS). [S.l.]: NIPS, 2017: 5100-5109.

[8] WANG Yue, SUN Yongbin, LIU Ziwei, et al. Dynamicgraph CNN for learning on point clouds[J]. ACM Transactions on Graphics (TOG), 2019, 146: 1-12.

[9] ZHAO H, JIANG L, FU C W, et al. PointWeb: Enhancing local neighborhood features for point cloud processing[C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). [S.l.]: IEEE, 2019: 5560-5568.

[10] FAN S, DONG Q, ZHU F, et al. SCF-Net: Learning spatial contextual features for large-scale point cloud segmentation[C]//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). [S.l.]: IEEE, 2021: 14499-14508.

[11] CHEN X, MA H, WAN J, et al. Multi-view 3D object detection network for autonomous driving[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). [S.l.]: IEEE, 2017: 6526-6534.

[12] LI Bo, ZHANG Tianlei, XIA Tian. Vehicle detection from 3D lidar using fully convolutional network[C]//Proceedings of Robotics: Science and Systems. [S.l.]: [s.n.], 2016.

[13] GRAHAM B, ENGELCKE M, MAATEN L V D, 3D semantic segmentation with submanifold sparse convolutional networks[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2018: 9224-9232.

[14] LE T, DUAN Y. PointGrid: A deep network for 3D shape understanding[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2018: 9204-9214.

[15] 马娟,方源敏,赵文亮,等. 利用空间微分块与动态球策略的k邻域搜索算法研究[J]. 武汉大学学报, 2011, 36(3): 358-362.
MA Juan, FANG Yuanmin, ZHAO Wenliang, et al. Algorithm for finding k-neighbors based on spatial sub-cubes and dynamic sphere[J].Journal of Wuhan University, 2011, 36(3): 358-362.

[16] 刘鹏,王明阳,王焱. 基于自适应动态球半径的k邻域搜索算法[J]. 机械设计与制造工程, 2016, 45(6): 83-86.
LIU Peng, WANG Mingyang, WANG Yan. k-neighborhood search algorithm based on adaptive dynamic sphere radius[J]. Machine Design and Manufacturing Engineering, 2016, 45(6): 83-86.

[17] 欧阳峻岭. 基于高斯映射的点云分类算法研究[D]. 成都: 成都理工大学, 2018.
OUYANG Junling. Research on point cloud classification algorithm based on gaussian mapping[D]. Chengdu: Chengdu University of Technology, 2018.

[18] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of Advances in Neural Information Processing Systems(NeurIPS). [S.l.]: Curran Associates, Inc, 2017.

[19] YANG Jiancheng, ZHANG Qiang, NI Bingbing, et al.Modeling point clouds with self-attention and gumbel subset sampling

[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2019: 3318-3327.
[20] CHEN L, LI X, FAN D, et al.LSANet: Feature learning on point sets by local spatial attention[EB/OL]. (2019-06-20)[2022-02-24]. https://arxiv.org/abs/1905.05442v3.
[21] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation[C]//Proceedings of Medical Image Computing and Computer-Assisted Intervention (MICCAI). [S.l.]: [s.n.], 2015: 234-241.
[22] ARMENI I, SENER O, ZAMIR A R, et al. 3D semantic parsing of large-scale in door spaces[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2016: 1534-1543.
[23] HACKEL T, SAVINOV N, LADICKY L, et al. NET: A new large-scale point cloud classification benchmark[C]//Proceedings of ISPRS Annals of the Photogrammetry. [S.l.]: Remote Sensing and Spatial Information Sciences, 2017: 91-98.
[24] HUANG Q, WANG W, NEUMANN U. Recurrent slice net-works for 3D segmentation of point clouds[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2018: 2626-2635.
[25] YE X, LI J, HUANG H, et al. 3D recurrent neural networks with context fusion for point cloud semantic segmentation[C]//Proceedings of the European Conference on Computer Vision (ECCV). [S.l.]: [s.n.], 2018: 415-430.
[26] 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.
[27] LI Y, BU R, SUN M, et al. PointCNN: Convolution on x-transformed points[C]//Proceedings of Advances in Neural Information Processing Systems(NeurIPS). [S.l.]: Curran Associates, Inc, 2018: 820-830.
[28] ZHANG Z, HUA B S, YEUNG S K. ShellNet: Efficient point cloud convolutional neural networks using concentric shells statistics[C]//Proceedings of IEEE International Conference on Computer Vision. [S.l.]: IEEE, 2019: 1607-1616.
[29] HACKEL T, WEGNER J D, SCHINDLER K. Fast semantic segmentation of 3D point clouds with strongly varying density [C]//Proceedings of ISPRS Annals of the Photogrammetry, Remotesensing and Spatial Information Sciences. [S.l.]: [s.n.], 2016: 177-184.
[30] CONTRERAS J, DENZLER J. Edge-convolutionpoint net for semantic segmentation of large-scale pointclouds[C]//Proceedings of IGARSS 2019—2019 IEEE International Geo-science and Remote Sensing Symposium. [S.l.]: IEEE, 2019: 5236-5239.
[31] BOULCH A, GUERRY J, SAUX B, et al. Snapnet: 3D point cloud semantic labeling with 2D deep segmentation networks [J]. Computers & Graphics, 2018, 71: 189-198.
[32] MA Yanni, GUO Yulan, LIU Hao, et al. Global context reasoning for semantic segmentation of 3D point clouds[C]//Proceedings of the IEEE Winter Conference on Applications of Computer Vision. [S.l.]: IEEE, 2020: 2931-2940.
[33] LIU H, GUO Y, MA Y, et al. Semantic context encoding for accurate 3D point cloud segmentation[C]//Proceedings of IEEE Transactions on Multimedia. [S.l.]: IEEE, 2020: 1-10.

作者简介:

李嘉祥(1995-),男,硕士研究生,研究方向:计算机视觉、点云处理、语义分割等,E-mail:15902054490@163.com。

宣士斌(1964-),**通信作者**,男,博士,教授,硕士生导师,研究方向:计算机视觉,E-mail:xuanshibin@gxun.edu.cn。

刘丽霞(1997-),女,硕士研究生,研究方向:计算机视觉、医学图像处理。

王款(1995-),男,硕士研究生,研究方向:计算机视觉、姿态估计。

(编辑:夏道家)