



武汉大学学报(信息科学版)

*Geomatics and Information Science of Wuhan University*

ISSN 1671-8860, CN 42-1676/TN

## 《武汉大学学报(信息科学版)》网络首发论文

题目: 使用局部几何特征与空洞邻域的点云语义分割  
作者: 项学泳, 李广云, 王力, 宗文鹏, 吕志鹏, 向奉卓  
DOI: 10.13203/j.whugis20200567  
收稿日期: 2021-03-01  
网络首发日期: 2021-06-06  
引用格式: 项学泳, 李广云, 王力, 宗文鹏, 吕志鹏, 向奉卓. 使用局部几何特征与空洞邻域的点云语义分割. 武汉大学学报(信息科学版).  
<https://doi.org/10.13203/j.whugis20200567>



**网络首发:** 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

**出版确认:** 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

DOI:10.13203/j.whugis20200567

引用格式：

项学泳, 李广云, 王力, 等. 使用局部几何特征与空洞邻域的点云语义分割 [J]. 武汉大学学报·信息科学版, 2021, DOI: 10.13203/j.whugis20200567 (Xiang Xueyong, Li Guangyun, Wang Li, et al. Semantic segmentation of point clouds using local geometric features and dilated neighborhoods[J]. Geomatics and Information Science of Wuhan University, 2021, DOI: 10.13203/j.whugis20200567)

## 使用局部几何特征与空洞邻域的点云语义分割

项学泳<sup>1</sup>, 李广云<sup>1</sup>, 王力<sup>1</sup>, 宗文鹏<sup>2</sup>, 吕志鹏<sup>1</sup>, 向奉卓<sup>3</sup>

1 信息工程大学地理空间信息学院, 河南 郑州, 450000

2 西安测绘研究所, 陕西 西安, 710000

3 61287 部队, 四川 成都, 610036

**摘要:** 点云具有数据量大、无拓扑结构等特点, 现有的深度学习语义分割模型难以充分挖掘大范围邻域内点云中所隐藏的几何特征。为此, 提出一种基于空洞邻域并结合角度等几何特征作为模型输入的点云语义分割模型。首先, 在局部邻域构建过程中, 将图像处理的空洞卷积操作扩展至点云, 建立空洞邻域结构, 以扩大感受野。其次, 在特征提取过程中, 采用中心点与邻域点之间相对坐标、距离、角度等基本几何特征作为模型输入, 最大程度挖掘邻域内的几何特征。最后, 基于本文提出的邻域结构与特征提取算法构建了点云语义分割模型。采用 Semantic3D 数据集进行实验验证, 结果表明: 本文提出的模型分割效果优于对比的点云语义分割算法, 空洞邻域与局部几何输入特征能够有效改善点云语义模型的性能。

**关键词:** 点云; 语义分割; 空洞邻域; 几何特征

**项目资助:** 国家自然科学基金 (42071454), 国家重点研发计划资助 (2017YFF0206000)。

**第一作者:** 项学泳, 博士生, 主要从事三维场景环境感知研究, [ahhsxxy@163.com](mailto:ahhsxxy@163.com)。

**通讯作者:** 李广云, 博士, 教授, 博士生导师, 主要从事精密工程测量研究, [guangyun\\_li\\_chxy@163.com](mailto:guangyun_li_chxy@163.com)。

三维激光扫描测量技术是三维空间信息获取的一种重要方式。原始点云数据由于缺乏拓扑结构, 除可视化与距离测量外, 可以直接进行应用的领域相对有限<sup>[1]</sup>。需要通过数据处理从中提取出感兴趣的信息, 其中获取点云的语义信息是实现三维场景识别的重要手段, 其在基础测绘、智慧城市建设、机器人环境感知等方面有着广泛应用<sup>[2-3]</sup>。

为了解决点云的语义分割问题, 相关学者提出了大量解决方案, 包括基于边缘的分割方法<sup>[4]</sup>、区域生长型算法<sup>[5]</sup>、模型匹配法<sup>[6]</sup>以及以上述方法为基础加以组合构成的复合算法<sup>[7]</sup>。由于点云的非结构化和分布不均匀等特性, 加之现实场景中地物类别繁多、分布杂乱, 因而难以采用某一种特定的算法解决点云语义分割问题, 迫切需要一种以数据作为驱动的方式进行处理。

近年来,深度学习在图像和自然语言处理方面取得了巨大进展,相关研究也逐步扩展至点云数据处理领域。代表性的工作主要包括:文献[8]首先获得图像语义分割效果,再将图像分割结果传递至点云,RangeNet++<sup>[9]</sup>将点云转换为二维图像,在图像上执行语义分割后,再将图像语义分割结果映射至点云空间,但在映射过程中由于遮挡会导致三维空间信息丢失。VoxNet<sup>[10]</sup>直接将三维卷积应用于点云实现点特征提取,然而该处理方式会导致内存消耗严重、计算代价高昂。PointNet<sup>[11]</sup>在考虑置换不变性的基础上实现了直接基于点云完成语义分割任务,具有开创性的意义,但该网络逐点提取点特征,在网络后期仅融合了全局特征,忽略了局部邻域信息。PointNet++<sup>[12]</sup>虽建立了两种邻域结构,但在每个邻域内部,却迭代的使用 PointNet 提取邻域特征,运算量较大。

本文所提出的模型基于多层感知机,该模型直接以点云作为输入,无需进行中间数据转换。然而现有基于多层感知机的语义分割模型在邻域构建与特征提取过程中均存在一定的不足。首先,特征提取阶段获取更大范围邻域信息有利于提升语义分割准确性,但在现有模型<sup>[11,12]</sup>基础上盲目地扩大邻域范围会增加模型参数数量,导致模型训练困难。因此,本文从图像处理领域的空洞卷积<sup>[13]</sup>获得启发,将图像空洞卷积扩展至点云处理中,该处理方式能够在不增加模型参数的前提下获取更大范围的邻域几何特征。其次,在特征提取过程中,文献[11,12,14,15]单纯以点坐标作为网络的输入,这种输入方式难以充分挖掘点云中所隐含的几何信息。为此,本文在特征提取时增加了距离与角度作为特征提取阶段的额外输入特征,这种看似冗余的输入方式却能提升语义分割性能。

## 1 算法介绍

### 1.1 邻域构建

目前深度学习点云处理模型普遍采用 K 最近邻算法 (K-Nearest Neighbor, KNN) 获取邻域点。当点云比较稠密时, KNN 算法所获取的邻域范围比较有限,虽可通过增加邻近点的数量进行扩大,然而邻域点数目的增多却导致特征提取阶段模型参数的增多,最终带来网络模型难以优化、语义分割效率降低等不良影响。

文献[13]在进行图像卷积操作过程中,通过间隔像素建立空洞卷积结构。考虑到点云数据结构的不规则性,本文依据欧氏距离构建了两种空洞邻域结构,分别如图 1 (b) 和图 1 (c) 表示。在图 1 (b) 中,首先通过 KNN 算法搜索  $2 \times k$  个点,并按照欧氏距离对其进行排序,之后并每间隔一点进行抽取,最终获得  $k$  个邻域点,将上述邻域结构命名为空洞 KNN。考虑到近中心的邻域点与中心点的相关性较强,本文构建了如图 1 (c) 所示的第二种邻域结构,先获取距离中心点最近的  $k/2$  个点,之后在外层的  $k$  个点中,每隔一点抽取剩余的  $k/2$  个点,最终组合获得  $k$  个邻域点,将上述邻域结构命名为复合 KNN。

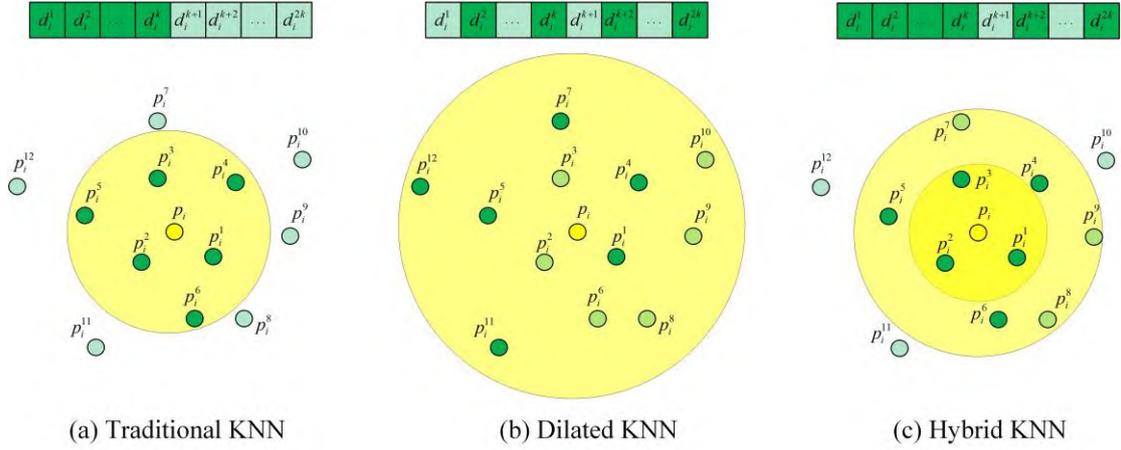


图 1 三种点云邻域结构对比

Fig.1 Comparison of Three Kinds of Point Cloud Neighborhood Structure

相对于文献[14,15]采用经典 KNN 算法直接获取  $k$  个距离中心点最近的点构成邻域, 本文提出的空洞 KNN 和复合 KNN 根据邻域点与中心点之间的欧氏距离构建邻域空间, 可有效扩大邻域范围, 同时保持邻域内点的数量不变, 因此并未增加模型的参数数量。

## 1.2 特征提取

点云  $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^N \in \mathbb{R}^{N \times F}$  一般由点坐标及颜色、回光强度等附加特征组成, 其中  $N$  为点个数,  $F$  为点特征维度数。特征提取的主要目的是使得其中的每一个点均能获取与邻域点的相关特征。对于任意一点  $p$ , 首先利用上文所提出的邻域搜索算法获取邻域点, 然后通过公式 (1) 对其编码。

$$\mathbf{f}_i^m = h([\mathbf{p}_i \quad \mathbf{p}_i^m \quad d_i^m \quad \mathbf{v}_i^m \quad \mathbf{a}_i^m]) \quad (1)$$

上式中,  $\mathbf{f}_i^m$  表示中心点  $i$  与其第  $m$  个邻域点之间的编码特征;  $h$  为特征提取函数, 本文模型采用多层感知机 (Multi layer perceptron, MLP) 实现;  $\mathbf{p}_i$  与  $\mathbf{p}_i^m$  分别表示中心点与其第  $m$  个邻域点的三维坐标;  $d_i^m = \|\mathbf{p}_i^m - \mathbf{p}_i\|$  表示中心点与邻域点之间的欧氏距离;  $\mathbf{v}_i^m = (\mathbf{p}_i^m - \mathbf{p}_i)$  表示以中心点为原点的邻域点相对坐标;  $\mathbf{a}_i^m$  为邻域点在  $x$ 、 $y$ 、 $z$  三个方向上相对于中心点的角度反正切值, 计算公式如下:

$$\mathbf{a}_i^m = (\arctan(\frac{z_i^m - z_i}{x_i^m - x_i}), \arctan(\frac{z_i^m - z_i}{y_i^m - y_i}), \arctan(\frac{y_i^m - y_i}{x_i^m - x_i})) \quad (2)$$

上式中,  $x_i$ 、 $y_i$ 、 $z_i$  表示中心点坐标,  $x_i^m$ 、 $y_i^m$ 、 $z_i^m$  表示第  $m$  个邻域点的坐标。与文献[11, 12, 15]仅采用几何坐标作为模型的输入不同, 式 (1) 在输入信息中融合了距离、角度在内的多种基本几何特征, 为网络挖掘更深的几何特征提供了良好的初始条件。最终, 将获得的中心点与邻域点之间的特征  $\mathbf{f}_i^m$  与颜色、回光强度等其它输入特征进行结合, 若采集到的点云无附加特征, 则直接将本处理单元输出的特征作为特征聚合的输入, 做进一步的处理。特征提取过程如图 2 所示。

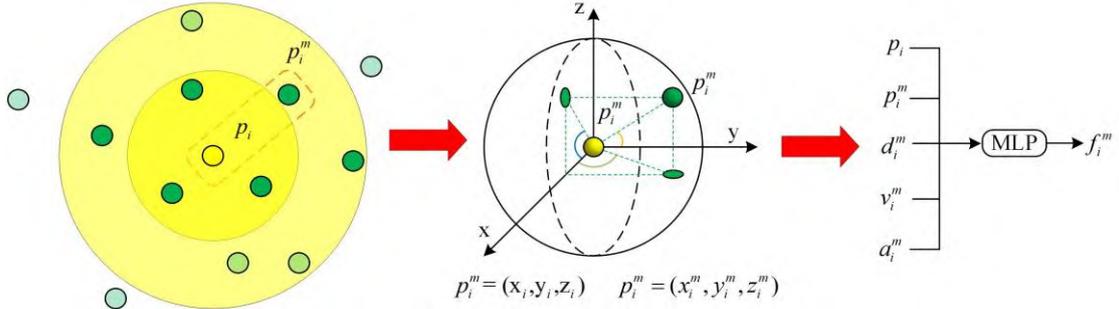


图2 点云特征编码过程

Fig.2 Point Cloud Feature Encoding Process

### 1.3 特征聚合

特征提取阶段获得了中心点与各邻域点的交互特征,本节对交互特征的进行聚合以获得中心点的特征。简单地采用求和平均虽可实现,但在语义分割过程中,更希望相同语义类内部特征差异小,而不同类别间的特征差异较大,从而有效地区分各语义类别。本文与文献[16]的处理方式类似,采用了具有注意力机制的特征聚合操作。特征聚合输出各点特征表示为  $\{f_1, f_2, \dots, f_N\} \subseteq \mathbb{R}^{N \times H}$ ,  $\{f_i^1, f_i^2, \dots, f_i^k\} \subseteq \mathbb{R}^{k \times H}$  表示第  $i$  个中心点与其  $k$  个邻域点之间的特征,聚合过程由式 (3) 所示:

$$f_i = \sum_{m=1}^k \hat{w}_i^m f_i^m \quad (3)$$

上式中,  $\hat{w}_i^m$  表示中心点  $i$  与其第  $m$  个邻域点之间的相关性,可由式 (4)–(5) 计算得到。

$$w_i^m = g(\mathbf{w}, f_i^m) \quad (4)$$

$$\hat{w}_i^m = \text{softmax}(w_i^m) = \frac{w_i^m}{\sum_{m=1}^k w_i^m} \quad (5)$$

式 (4) 中,  $g: \mathbb{R}^H \rightarrow \mathbb{R}^1$ , 可由神经网络实现,  $\mathbf{w}$  为神经网络的权值。对式 (4) 中的  $w_i^m$  进一步使用 softmax 函数做归一化处理,可获得式 (3) 中的  $\hat{w}_i^m$  权重。训练过程中,依据反向传播的回馈,通过式 (4) 自动调节权重  $\hat{w}_i^m$ ,从而由网络自适应地决定中心点与邻域点之间的相关性。

### 1.4 池化处理

在点云特征编码过程中,池化处理与降采样是同步处理的流程。池化操作可降低数据量,提升模型处理效率。本文模型直接以点作为输入,池化还需满足文献[11]所提及的置换不变性的要求,即当点云的输入顺序发生变化时,模型学习获得的点云特征仍然保持不变。 $\{f_1, f_2, \dots, f_N\} \subseteq \mathbb{R}^{N \times H}$  为前一阶段特征聚合所生成的一组特征向量,假设降采样率为  $1/4$ ,则池化后的特征为  $\{\hat{f}_1, \hat{f}_2, \dots, \hat{f}_{N/4}\} \subseteq \mathbb{R}^{N/4 \times H}$ 。为满足置换不变性,本文采用最大池化,池化过程如图 3 所示。

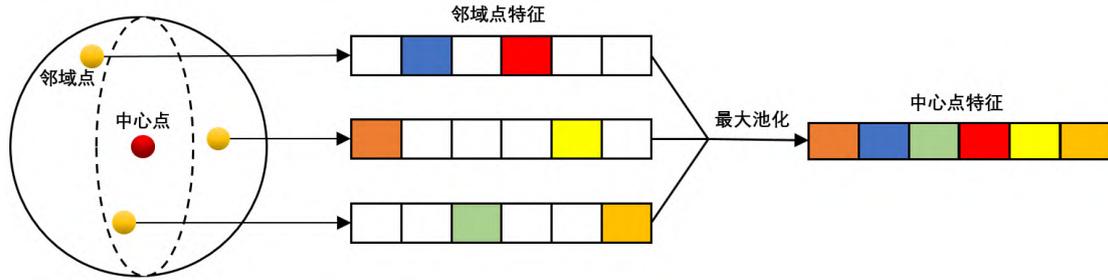


图3 池化处理过程

Fig.3 Pooling Process

图3中，首先使用KNN算法获取若干邻域点，每个邻域点表示为 $n$ 维特征向量，最大池化后中心点的特征向量则为所有邻域点特征向量每一维度的最大值。因此，即使邻域点的顺序发生变化，仍可保证所提取的特征保持不变。

## 2 网络结构

模型的整体结构采用U-net<sup>[17]</sup>，U-net最初是针对图像语义分割而专门开发的网络，该网络结构已被广泛地应用于点云语义分割<sup>[14, 15, 18, 19]</sup>。与U-net类似，本文模型整体上分为编码和解码两个阶段，如图4所示。

模型直接以点云坐标、RGB颜色、回光强度作为输入，输出为每个点的标签预测值，通过交叉熵损失函数计算预测值与标签值之间的损失值，反向传播依据损失值并使用Adam优化器对网络进行监督训练。

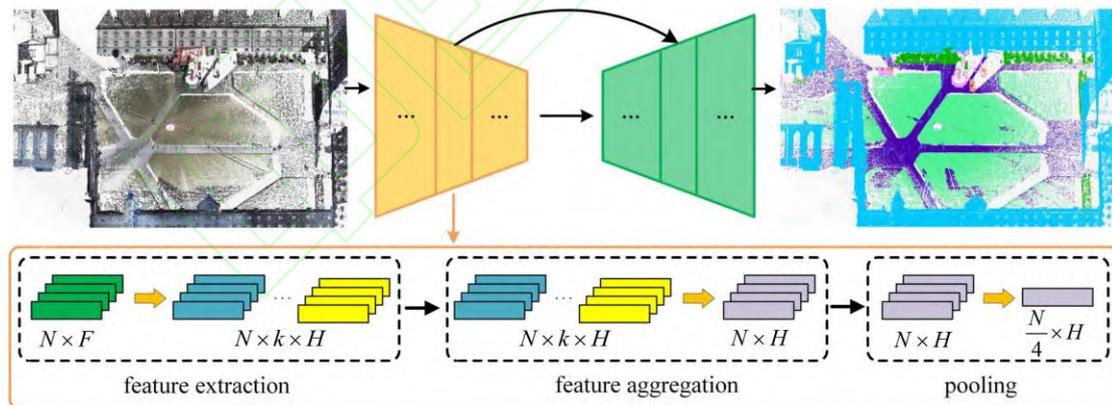


图4 模型整体结构

Fig.4 Overview of the Proposed Architecture

编码阶段设置了4个处理层，上一处理层的输出为下一处理层的输入，每个处理层分为特征提取、特征聚合、池化3个处理单元，特征编码和特征聚合单元逐步地提取点云中的空间几何特征，池化操作逐步地稀疏点云并保留点云中高阶特征，与池化同步的降采样采用反密度采样算法<sup>[37]</sup>，降采样率为1/4，相较于文献[12]中模型所采用的最远点采样，该采样算法在效率上更具优势。特征提取、特征聚合、池化具体过程以及图4中各符号的意义与第一

章一致,在此不再赘述。经过4个处理层,每一层输出的点特征向量维度分别为:64、128、256、512,邻域点的数值k取16。

解码阶段同样设置了4个处理层,每个处理层之间通过最邻近点插值算法逐步恢复点云几何细节,升采样率为4。在每一层,借助于U-net的跨层连接,合并来自编码阶段对应分辨率的特征图,从而增强上下文信息关联<sup>[17]</sup>,经过4层插值与合并操作,最终获得点云中单独每一点的特征,进一步通过全连接网络和softmax分类器获得各点语义标签的预测值。

## 3 实验与讨论

### 3.1 实验准备

为评价本文所提模型的语义分割效果,采用大型室外场景数据集Semantic 3D对本文模型进行测试。

Semantic 3D数据集<sup>[20]</sup>由苏黎世联邦理工学院下属的大地测量与摄影测量学院于2017年通过地面静态扫描方式采集获得,点云数据包含XYZ坐标,RGB颜色与回光强度,总点数超过40亿。采集场景包含了教堂、街道、铁轨、广场、村庄、足球场、城堡等。数据集包含8个语义类别,分别为:人造地形、自然地形、高植被、低植被、建筑物、硬质景观、扫描人工制品和车辆。

采用平均交并比(mean Intersection over Union,mIoU)与总体精度(Overall Accuracy,OA)两类衡量标准评价模型性能,其中mIoU通过计算所有语义类别交并比(Intersection over Union,IoU)的平均值获得,计算方法分别如式(6)和式(7)所示。

$$mIoU = \frac{1}{n} \sum_{i=1}^n \frac{TP_i}{TP_i + FP_i + FN_i} \quad (6)$$

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

式(6)中, $TP_i$ 、 $FP_i$ 、 $FN_i$ 分别表示第i类别的真阳例、假阳例、假阴例中所包含的点的个数,n为场景中总的语义类别数;式(7)中, $TP$ 、 $TN$ 、 $FP$ 、 $FN$ 分别表示场景中所有类别的真阳例、真阴例、假阳例、假阴例中所包含的点的个数。

实验硬件环境为Intel Core i9-9980Xe CPU、GPU为4块的Nvidia RTX 2080Ti,RAM为128GB。软件环境采用ubuntu16.04下的CUDA10.2+cuDNN7.6.5+python3.5+anaconda3.6+pytorch1.5。训练期间批处理块的大小为100,每批次处理的点数为8192个,模型的初始学习率为0.02,学习衰减率为0.9,训练迭代次数为50次。

### 3.2 实验结果与分析

为验证本文所提出的两种空洞邻域结构对点云语义分割的影响,模型编码阶段设置为4层,且保持输入几何特征、实验组邻域点数量等均相同,使用Semantic3D数据集的15个场景建立训练集和验证集,建立方式参照文献[19]。依次将模型1、2、3、4处理层按顺序替

换为本文两种空洞邻域结构和 4 层均设置为经典 KNN 邻域结构，共获得 9 个模型。为测试模型的效率，使用了 semantic3D 下的 semantic-8 测试集，各模型的总体精度、平均交并比与测试阶段耗时如表 1 所示。为进一步测试邻域点的数目对语义分割效果的影响，在模型的第 1、2 层均使用复合 KNN 的模型基础上变化邻域点数目，实验结果如表 2 所示。

表 1 不同邻域结构下模型的总体精度、平均交并比与推理耗时

Tab.1 The OA Scores, mIoU Scores and Inference Time of Models with Different Neighborhood Structure

处理层	传统 KNN			空洞 KNN			复合 KNN		
	OA/ (%)	mIoU/ (%)	推理耗时 (min)	OA/ (%)	mIoU/ (%)	推理耗时 (min)	OA/ (%)	mIoU/ (%)	推理耗时 (min)
1	-	-	-	88.3	62.4	10.90	87.8	62.1	10.93
1-2	-	-	-	87.4	61.8	11.01	88.5	62.8	10.98
1-3	-	-	-	85.7	60.4	11.07	87.1	61.6	11.02
1-4	86.8	61.4	10.90	83.7	59.3	11.13	85.2	60.3	11.13

表 2 不同邻域点数目下模型的语义分割结果

Tab.2 Semantic Segmentation Results of the Model for Different Number of Neighborhood Points

邻域点数	8	12	16	20	24
OA / (%)	87.1	87.9	88.5	88.7	88.8
mIoU / (%)	61.8	62.3	62.8	62.9	63.0

从表 1 可得，相较于将模型整个 4 层均设置为传统 KNN 邻域结构，在模型的浅层设置本文所提出的两种空洞邻域结构，mIoU 和 OA 均有所提升。其中，在第一层设置空洞 KNN 的 OA 和 mIoU 分别高于将 4 层均设置为 KNN 提升了 1.5% 和 1%。在 1-2 层均设置复合 KNN 可获得最好的语义分割效果，OA 和 mIoU 分别高于将 4 层均设置为传统 KNN 提升了 1.7% 和 1.4%。由此可见，邻域点范围的扩大对点云语义分割起到了积极作用。此外，随着处理层的加深，点的数量变得越来越少，最后一层在池化前的点数量仅有模型输入点数量的 1/64，在这种稀疏点云状态下加入空洞 KNN 和复合 KNN 舍弃了部分空间相关性较强的邻近点，OA 和 mIoU 均出现不同程度的下滑。从表 1 还可得，各模型的推断耗时基本在 11 分钟左右，这与预期的一致，可见两种邻域模型虽扩大了邻域范围，但如 1.1 小节所述，邻域结构改善并未增加模型的参数数量。从表 2 可得，模型的 OA 和 mIoU 随着邻域点数量的增加而增加，当邻域点数量达到 16 上下时，虽可进一步增加邻域点的数目以改善语义分割结果，但提升并不明显，为平衡模型效率和表现力，将模型的邻域点数目设定为 16。

为验证几何输入特征对语义分割效果的影响，进一步开展对比实验，各实验组模型的 1-2 层的邻域几何结构均设置为复合 KNN，邻域点数均保持相同，仅对模型的输入类别做出变化，实验结果及模型测试阶段耗时如表 3 所示。

表 3 不同输入几何特征下模型的总体精度、平均交并比与推理耗时

Tab.3 The OA Scores, mIoU Scores and Inference Time of Models under Different Input Geometric Features

model	OA / (%)	mIoU / (%)	推理耗时 / (min)
$p_i + p_i^m + v_i^m$	10.90	10.93	9.47

$p_i + p_i^m + v_i^m + d_i^m$	11.01	10.98	10.02
$p_i + p_i^m + v_i^m + a_i^m$	11.07	11.02	10.05
Ours	11.13	11.13	10.98

由于现有大多数模型均以中心点、邻域点、以及相对坐标作为输入量，因此本文只探讨中心点与邻域点间的相对距离与角度对点云语义分割效果的影响。由表 3 可知：相比于单纯以中心点坐标、邻域点坐标以及相对坐标作为模型输入，以距离作为附加输入，OA 和 mIoU 分别提升了 6.3% 和 4.8%，以角度作为附加输入，OA 和 mIoU 分别提升了 7.5% 和 6.1%。当在以坐标为模型输入的基础上加入角度与距离，OA 和 mIoU 分别提升了 10.7% 和 8.8%。由此可见，在输入端增加不同类基本的几何特征可以有效地改善模型的语义分类结果。在推理耗时上，模型的推理时间随着输入几何特征类别的增加而递增，最终本文模型的推理耗时为 10.98 分，相对于目前流行的单纯以点坐标作为输入模型仅增加了 1.51 分，差距不大。

将最佳邻域设置与输入特征设置组合为本文的最终模型，并以 Semantic3D 训练集的 15 个场景训练本文模型，在测试集的 15 个场景对本文模型进行测试，由于测试集无真值标签，本文仅展示了输入值与预测值，部分场景的分割效果如图 5 所示。

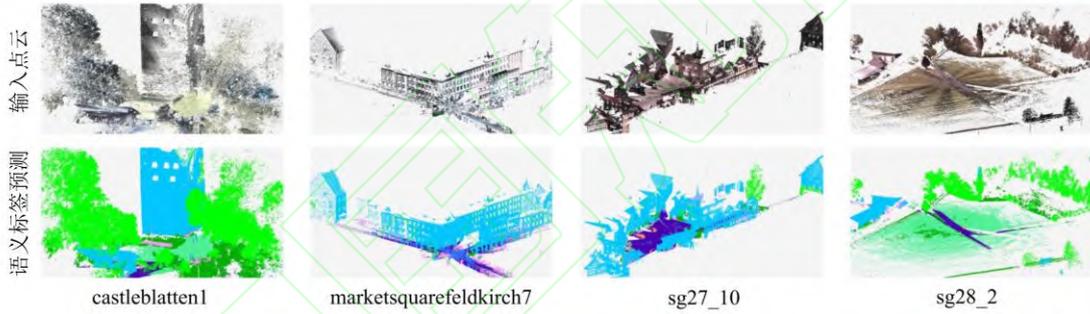


图 5 本文模型在 Semantic3D 分割效果

Fig.5 Qualitative Results of Proposed Approach on Semantic3D

图 5 展示了乡村、城市、市郊三类具有代表性的场景，由图可得，本文所提模型点云分割边界清晰，地物内部杂点较少，整体目视分割效果良好。为进一步定量地评价本文所设计模型的分割效果，将模型在 Semantic3D 的 reduced-8 测试集上的结果，与近几年文献[21-24]所提出的语义分割模型的测试结果进行对比，以各类别的 IoU、所有类别的 mIoU 以及 OA 为评价指标，结果如表 4 所示，表 4 数据来自 Semantic3D 官网。

表 4 不同点云语义分割算法在 Semantic3D 上的结果对比 (%)

Tab.4 Quantitative Results of Different Approaches on Semantic3D (%)

地物类别	OctreeNet_CRF[21]	new_net[22]	RF_MSSF[23]	MSDeepVoxNet[24]	Ours
人工地面	90.7	84.5	87.6	83.0	96.5
自然地面	82.0	70.9	80.3	67.2	86.9
高植被	82.4	76.6	81.8	83.8	81.8
矮植被	39.3	26.1	36.4	36.7	38.2
建筑物	90.0	91.4	92.2	92.4	91.6

硬质景观	10.9	18.6	24.1	31.3	28.7
扫描人工制品	31.2	56.5	42.6	50.0	55.8
车辆	46.0	51.4	56.6	78.2	65.8
mIoU %	59.1	59.5	62.7	65.3	68.2
OA %	89.9	87.9	90.3	88.4	91.2

由表 4 可知, 本文模型分类的整体精度和平均交并比分别为 91.2% 和 68.2%, 达到了较为理想的分类效果。从每一类别看, 自然地面与人工地面分类的交并比分别为 96.5% 和 86.9%, 较大程度优于其它模型; 从植被分类结果看, 本文模型在高植被分类方面效果较好, 但对矮植被的分类效果较差; 从建筑物的分类结果看, 5 种模型分类结果在 90.0%-92.4% 之间, 相差不大; 从硬质景观、扫描人工制品、车辆来看, 本文模型均表现较弱, 其中硬质景观的分类结果仅为 28.7%。从分类结果整体来看, 本文在地表分类结果上表现较为出色, 但在场景中类别分类结果上均一定程度上落后于其他算法。

## 4 结论

本文以空洞邻域与多组几何特征输入为基础建了点云特征提取单元, 并基于特征聚合、最大池化构建了点云语义分割模型, 模型构建简洁, 通过大型室外场景数据集 Semantic3D 的测试, 验证了本文模型具有良好的点云语义分割能力。空洞邻域能够在不增加模型参数的前提下, 有效地提取更大范围内的点云信息, 多类别基本几何特征输入能够最大程度挖掘邻域内所隐藏的几何相关性。从分类结果来看, 本文分类的准确性倾向于包含地面在内的大的地物类别, 传统算法通常可从实验结果溯源分析, 而深度学习对于网络设计者来说在一定程度上也是一个“黑匣子”, 因此关于网络不同处理模块设计对于实验结果的具体影响仍需做进一步分析。

## 参考文献

- [1] Hackel T, Wegner J D, Schindler K. Fast Semantic Segmentation of 3D Point Clouds With Strongly Varying Density[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2016, 3(3): 177-184
- [2] Yang Bisheng, Dong Zhen. Point Cloud Intelligent Processing[M]. Beijing: Science Press, 2020(杨必胜, 董震. 点云智能处理[M]. 北京: 科学出版社, 2020)
- [3] YANG Yuqi, CHEN Chi, YANG Bisheng, et al. 3D Change Detection of Buildings Based on Multi-level Segmentation of Dense Matching Point Clouds from UAV Images[J]. *Geomatics and Information Science of Wuhan University*, 2021, 46(4): 489-496 (杨钰琪, 陈驰, 杨必胜, 等. 基于 UAV 影像密集匹配点云多层次分割的建筑物层高变化检测[J]. 武汉大学学报·信息科学版 2021, 46(4): 489-496)
- [4] Castillo E, Liang J, Zhao H. Point Cloud Segmentation and Denoising Via Constrained Nonlinear Least Squares Normal Estimates[M]. *Innovations for Shape Analysis*. Springer, Berlin, Heidelberg, 2013
- [5] Vo A V, Truong-Hong L, Laefer D F, et al. Octree-Based Region Growing for Point Cloud Segmentation[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2015, 104(7): 88-100
- [6] Zhang Liangpei, Zhang Yun, Chen Zhenzhong, et al. Splitting and Merging Based Multi-model Fitting for Point Cloud Segmentation[J] *Acta Geodaetica et Cartographica Sinica*, 2018, 47(6): 833-843(张良培, 张云,

- 陈震中, 等. 基于分裂合并的多模型拟合方法在点云分割中的应用[J]. *测绘学报*, 2018, 47(6): 833-843
- [7] Yan Li, Xie Hong, Hu Xiaobin, et al. A New Hybrid Plane Segmentation Approach of Point Cloud [J] *Geomatics and Information Science of Wuhan University*, 2013, 35(5): 517-521(闫利, 谢洪, 胡晓斌, 等. 一种新的点云平面混合分割方法[J]. *武汉大学学报·信息科学版*, 2013, 35(5): 517-521)
- [8] Xiong Hanjiang, Zheng Xianwei, Ding You, et al. Semantic Segmentation of Indoor 3D Point Cloud Model Based on 2D-3D Semantic Transfer [J] *Geomatics and Information Science of Wuhan University*, 2018, 12(12): 2303-2309(熊汉江, 郑先伟, 丁友, 等. 基于 2D-3D 语义传递的室内三维点云模型语义分割 [J]. *武汉大学学报·信息科学版*, 2018, 7(12): 2303-2309)
- [9] Andres M, Ignacio V, Jens B, et al. RangeNet++: Fast and Accurate Lidar Semantic Segmentation[C]. *IEEE International Conference on Intelligent Robots and Systems*, Macau, China, 2019
- [10] Maturana, Daniel, Sebastian Scherer. Voxnet: A 3d Convolutional Neural Network for Real-time Object Recognition[C]. *IEEE International Conference on Intelligent Robots and Systems*, Hamburg, Germany, 2015
- [11] Qi C, 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*, Honolulu, HI, USA, 2017
- [12] Qi C, Yi L, Su H, et al. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space[C]. *Advances in Neural Information Processing Systems*, Long Beach, CA, USA, 2017
- [13] Yu, Fisher, and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions[C]. *International Conference on Learning Representations*, San Juan, Puerto Rico, 2016
- [14] Jiang M, Wu Y, Zhao T, et al. Pointsift: a Sift-like Network Module for 3D Point Cloud Semantic Segmentation[J]. *Arxiv Preprint*, ArXiv: 1807.00652, 2018
- [15] Li Y, Bu R, Sun M, et al. Pointcnn: Convolution on X-Transformed Points[C]. *Advances in Neural Information Processing Systems*, Montréal, Canada, 2018
- [16] Liang Z, Yang M, Li H, et al. 3D Instance Embedding Learning With a Structure-Aware Loss Function for Point Cloud Segmentation[J]. *IEEE Robotics and Automation Letters*, 2020, 5(3): 4915-4922
- [17] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation[C]. *International Conference on Medical image computing and computer-assisted intervention*, Munich, Germany, 2015
- [18] Zhao H, Jiang L, Fu C. PointWeb: Enhancing local neighborhood features for point cloud processing[C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, USA, 2019
- [19] Thomas H, Qi C, Deschaud J, et al. Kpconv: Flexible and Deformable Convolution for Point Clouds[C]. *Proceedings of the IEEE International Conference on Computer Vision*, Seoul, South Korea, 2019
- [20] Hackel T, Savinov N, Ladicky L, et al. Semantic3D Net: a New Large-scale Point Cloud Classification Benchmark[C]. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2017, (IV-1)W1: 91-98
- [21] Wang F, Zhuang Y, Gu H, et al. OctreeNet. A Novel Sparse 3-D Convolutional Neural Network for Real-Time 3-D Outdoor Scene Analysis[J]. *IEEE Transactions on Automation Science and Engineering*, 2019, 17(2): 735-747
- [22] Contreras J, Joachim D. Edge-Convolution Point Net for Semantic Segmentation of Large-Scale Point Clouds[C]. *IEEE International Geoscience and Remote Sensing Symposium*, Yokohama, Japan, 2019
- [23] Thomas H, Goulette F, Deschaud J, et al. Semantic Classification of 3D Point Clouds with Multiscale Spherical Neighborhoods[C]. *International Conference on 3D Vision*, Verona, Italy, 2018
- [24] Roynard X, Deschaud J, Goulett F. Classification of Point Cloud Scenes with Multiscale Voxel Deep Network[J]. *ArXiv Preprint*, ArXiv: 1804.03583, 2018

# Semantic segmentation of point clouds using local geometric features and dilated neighborhoods

Xiang Xueyong<sup>1</sup>, Li Guangyun<sup>1</sup>, Wang Li<sup>1</sup>, Zong Wenpeng<sup>2</sup>, Lü Zhipeng<sup>1</sup>, Xiang Fengzhuo<sup>3</sup>

<sup>1</sup> Institute of Geospatial Information, Information Engineering University, Zhengzhou, 450000, China;

<sup>2</sup> Xi'an Institute of Surveying and Mapping, Xi'an, 710000, China;

<sup>3</sup> 61287 Troops, Chengdu, 610036, China

**Abstract: Objectives** Point cloud has no topological structure, current deep learning semantic segmentation algorithm is difficult to capture geometric features implied in irregular points. In addition, the point cloud is in three-dimensional space with a large amount of data size. If we blindly expand the receptive field size during extract neighborhood information, it will increase the number of model parameters, which will make model training difficult. **Methods** To this end, we propose a point cloud semantic segmentation model based on the dilated convolution and combining elementary geometric features such as angle as the model input. First of all, during feature extraction, basic geometric features such as the relative coordinates, distance and angle between the centroid and the neighboring points are used as the model input to mine the geometric information. Secondly, in the process of building local neighborhoods, we expand the image dilated convolution operator to point cloud processing, the point cloud dilated operator can expand the receptive field size with no increasing the number of parameters of the model. Then the dilated convolution operator, multi-geometric features encoding modules and U-Net architecture are combined to form a complete point cloud semantic segmentation model. **Results** Semantic3D is applied to verify the proposed algorithms. The results show that compared with the traditional neighborhood structure, the OA of dilated neighborhood structure is increased by 1.4%. Compared with the model that only uses coordinates as input, multi-geometric features encoding module is increased by 10.7%. The final model based on the two proposed algorithms get mIoU and OA are 91.2% and 68.2%, respectively. **Conclusions** The dilated neighborhood structure can effectively extract point cloud information in a larger range without increasing the number of model parameters. multi-geometric features encoding module can maximize the capture of shape information in the neighborhood.

**Key words:** point cloud; semantic segmentation; dilated neighborhood; geometric features

**First author:** XIANG Xueyong, PhD candidate, specialize in 3D scene understanding, E-mail: ahhsxy@163.com

**Corresponding author:** LI Guangyun, professor, Doctoral supervisor, specialize in surveying and mapping engineering, E-mail: guangyun\_li\_chxy@163.com

**Foundation support:** The National Natural Science Foundation of China (42071454); The National Key Research and Development Program of China (2017YFF0206000)

收稿日期: 2021-03-01

DOI:10.13203/j.whugis.20200567