

# 控制与决策

Control and Decision

## 基于深度学习的三维点云分割综述

卢健, 贾旭瑞, 周健, 刘薇, 张凯兵, 庞菲菲

引用本文:

卢健, 贾旭瑞, 周健, 刘薇, 张凯兵, 庞菲菲. 基于深度学习的三维点云分割综述[J]. *控制与决策*, 2023, 38(3): 595–611.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2021.1648>

---

## 您可能感兴趣的其他文章

### Articles you may be interested in

#### [基于机器视觉的喷涂机器人轨迹规划与涂装质量检测研究综述](#)

Trajectory planning for spray-painting robot and quality detection of paint film based on machine vision: A review

*控制与决策*. 2023, 38(1): 1–21 <https://doi.org/10.13195/j.kzyjc.2022.1438>

#### [基于紧凑混合网络的视网膜血管自动分割](#)

Automatic segmentation of retinal vessel via compact mixed network

*控制与决策*. 2022, 37(2): 353–360 <https://doi.org/10.13195/j.kzyjc.2020.0883>

#### [基于DST融合多视图模糊推理赋值的三维目标检测](#)

3D object detection based on DST fusion multi-view fuzzy reasoning assignment

*控制与决策*. 2021, 36(4): 867–875 <https://doi.org/10.13195/j.kzyjc.2019.0434>

#### [融合稀疏编码与深度学习的草图特征表示](#)

A feature representation of sketch based on fusion of sparse coding and deep learning

*控制与决策*. 2021, 36(3): 699–704 <https://doi.org/10.13195/j.kzyjc.2019.0941>

#### [基于广义罚函数可行性准则的DE算法对不确定数据的处理](#)

Application of improved DE algorithm based on generalized penalty function feasibility criteria in uncertain data processing

*控制与决策*. 2021, 36(2): 498–504 <https://doi.org/10.13195/j.kzyjc.2019.0728>

# 基于深度学习的三维点云分割综述

卢健<sup>†</sup>, 贾旭瑞, 周健, 刘薇, 张凯兵, 庞菲菲

(西安工程大学 电子信息学院, 西安 710600)

**摘要:** 作为三维场景理解的重要技术之一, 三维点云分割受到广泛的关注, 具有重要的研究价值和广阔的应用前景. 基于此, 梳理基于深度学习的三维点云分割技术的最新研究进展; 在介绍三维点云分割常用的 8 个室内和室外数据集的基础上, 重点阐述和分析现有主要基于深度学习的语义分割、实例分割和部件分割方法, 并基于量化数据进行部分方法的效能比较; 最后从 10 个方面总结现有方法的不足, 并针对性地提出工作展望.

**关键词:** 深度学习; 三维点云; 数据集; 语义分割; 实例分割; 部件分割

中图分类号: TP391

文献标志码: A

DOI: 10.13195/j.kzyjc.2021.1648

引用格式: 卢健, 贾旭瑞, 周健, 等. 基于深度学习的三维点云分割综述[J]. 控制与决策, 2023, 38(3): 595-611.

## A review of deep learning based on 3D point cloud segmentation

LU Jian<sup>†</sup>, JIA Xu-rui, ZHOU Jian, LIU Wei, ZHANG Kai-bing, PANG Fei-fei

(School of Electronics and Information, Xi'an Polytechnic University, Xi'an 710600, China)

**Abstract:** 3D point cloud segmentation, as one of the important technology of 3D scene understanding, has aroused people's widespread interest. And it has important research value and broad application prospect. The latest research progress of 3D point cloud segmentation technology based on deep learning is sorted out. Firstly, eight indoor and outdoor common datasets frequently utilized in 3D point cloud segmentation are introduced. Then, the existing semantic segmentation, instance segmentation and part segmentation mainly via deep learning are explained and analyzed in depth, and the effectiveness of some methods is compared based on quantitative data. Finally, we summarize the shortcomings of the existing methods from ten aspects, and put forward the work prospects pertinently.

**Keywords:** deep learning; 3D point cloud; datasets; semantic segmentation; instance segmentation; part segmentation

## 0 引言

近年来, 3D 数据在自动驾驶<sup>[1]</sup>、增强现实<sup>[2]</sup>、机器人、遥感和医疗等领域被广泛使用, 推动了 3D 场景理解与分析技术的发展. 相比于 2D 数据, 3D 数据丰富的内部结构信息能够更加充分地表征复杂场景. 点云作为 3D 数据的主要表示形式, 可直接由可测量距离的三维传感器获取, 也可以从立体多视点图像中生成. 由于点云具有无序性和密度不均匀性等缺点, 难以直接应用常规卷积操作进行处理, 因此如何有效利用点云认知客观世界是一项有挑战性的工作.

3D 点云分割是点云场景理解与分析的重要支撑性技术. 传统的点云分割方法包括基于边缘检测的算法<sup>[3-4]</sup>、基于区域增长的算法<sup>[5-6]</sup>、基于特征聚类的算法<sup>[7-8]</sup>、基于模型拟合的算法<sup>[9-10]</sup>以及基于图形的算法<sup>[11-12]</sup>等. 由于这些方法过分依赖于人工设计

特征, 成本高、计算量大、普适性差, 在面向大数据集时表现不佳. 随着基于多视图的深度学习模型<sup>[13]</sup>和基于体素化的 3D 卷积神经网络<sup>[14]</sup>被提出, 深度学习开始广泛应用于 3D 点云数据的处理中. 随着相关研究的深入, 一些专门为 3D 点云分割设计的深度学习体系框架<sup>[15-19]</sup>被提出, 并显示出良好的效果, 但针对的数据集规模和应用场景有限. 近年来, 一些研究机构相继提供出较为新颖的室内和室外数据集, 如 ScanNet<sup>[20]</sup>、S3DIS<sup>[21]</sup>、Semantic3D<sup>[22]</sup>、SemanticKITTI<sup>[23]</sup>、ShapeNet Part<sup>[24]</sup>、Paris-Lille-3D<sup>[25]</sup>、PartNet<sup>[26]</sup>、SceneNN<sup>[27]</sup>等. 这些具有丰富应用场景的数据集提升了语义认知的难度, 相应地促使许多更加鲁棒和高效的方法被提出, 整体上促进了 3D 点云分割技术的进一步发展. 与现有的一些只关注于 3D 点云分割中某一类方法的综述<sup>[28-30]</sup>不同, 本文

收稿日期: 2021-09-22; 录用日期: 2022-05-31.

基金项目: 国家自然科学基金项目(61971339, 61471161); 陕西省自然科学基金重点项目(2018JZ6002).

<sup>†</sup>通讯作者. E-mail: lujian@xpu.edu.cn.

\*本文附带电子附录文件, 可登录本刊官网该文“资源附件”区自行下载阅览.

从语义分割、实例分割和部件分割3个角度对近年来3D点云分割技术的发展进行了全面性的综述。

本文的安排如下:第1节介绍目前主要的公开数据集及相应的评价指标;第2节~第4节分别介绍3D点云语义分割、实例分割和部件分割方法及表现;第5节从10个方面分析现有方法存在的缺陷,并提出可能的解决路径。

## 1 数据集和评价指标

建构有效且多样的数据集意义在于:首先,精心设计的数据集及其评价指标能够用于不同分割方法性能的评估和比较;其次,具有多样性数据和场景的数据集群能够用于评价某一分割方法的适应性和鲁棒性;再次,具有大数据和新型场景应用要求的数据集可以有力推进理论研究的深入,促进新方法产生。

### 1.1 数据集

为了促进3D点云分割技术的研究,一些机构公开了几种可靠的数据集。表1给出了3D点云分割研究中常用的8种数据集,其中ScanNet、S3DIS、Semantic3D、SemanticKITTI和ShapeNet part数据集近年来被研究者广泛使用。

表1 3D点云分割常用公开数据集

数据集	年份	应用场景	标注方式	类别	数据量
SceneNN	2016	室内场景	Mesh	—	—
ShapeNet part	2016	物体部件	Pointwise	16	16 881 shapes
S3DIS	2016	室内场景	Pointwise	13	700 million
ScanNet	2017	室内场景	Voxel	20	242 million
Semantic3D	2017	城市街道	Pointwise	8	4 009 million
Paris-Lille-3D	2018	城市街道	Pointwise	50	143.1 million
PartNet	2018	物体部件	Pointwise	24	26 671 shapes
SemanticKITTI	2019	户外场景	Pointwise	25	4 549 million

ScanNet<sup>[20]</sup>数据集是一个由美国普林斯顿大学和斯坦福大学共同开发的实例级室内RGB-D数据集,提供了来自70个独立室内场景的1 513次3D扫描场景,包含超过250万个RGB-D图像。在语义分割任务中,该数据集被标记为20类带注释的三维体素化对象,每类对应一个家具类别。S3DIS<sup>[21]</sup>数据集是一个包含2.5亿个点的大规模的室内RGB-D数据集,该数据采集于3个建筑物中的6个室内区域,包括教室和办公室等,其总覆盖面积超过6 000 m<sup>2</sup>,所标注的物体分为结构单元和活动单元共13类(如天花板、地板、墙、门等)。Semantic3D<sup>[22]</sup>数据集是目前最大的可用3D点云户外场景分割数据集,该数据集有8个具体的语义类,涵盖了广泛的户外场景,包括教堂、街道、铁轨、广场、城堡等,目前大约有40亿个

高质量手工标记的点。SemanticKITTI<sup>[23]</sup>数据集是基于KITTI视觉基准的大规模里程数据集,显示了德国卡尔斯鲁厄内部城市交通、住宅区、公路场景和农村道路。原始的测量数据集包括22个序列,将序列00到10作为训练集,将11到21作为测试集。该数据集共包含28个类别,在标注数据过程中使用定位与地图构建系统进行闭环检测和点云的配准、重建,确保地图的一致性。ShapeNet part<sup>[24]</sup>数据集包含55类共51 300个3D模型,是一个具有丰富标注的数据集。ShapeNet part是ShapeNet数据集的子集,由美国斯坦福大学和芝加哥丰田技术学院共同开发,包含16类物体16 881个形状、31 693个网格,零部件类别共有50种。

### 1.2 评价指标

评价指标用于分割算法之间的性能比较,包括均交并比(mean intersection over union, mIoU)、总体分割精度(overall accuracy, OA)、平均精确率(mean accuracy precision, mAP)、参数量(parameters)和每秒的浮点运算数(FLOPs)等,具体公式见表2。其中: $N$ 表示总样本数, $C$ 表示分割类别数,TP、TN、FP、FN分别表示为真阳性、真阴性、假阳性和假阴性。

表2 点云分割常用的评价指标

评价指标	计算公式
accuracy	$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$
mAcc	$\text{mAcc} = \frac{1}{C+1} \sum_{i=0}^n \text{accuracy}$
IoU	$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}$
mIoU	$\text{mIoU} = \frac{1}{C+1} \sum_{i=0}^C \text{IoU}$
OA	$\text{OA} = \frac{1}{N} \sum_C \text{TP}_i$
AP	$\text{AP} = \frac{1}{C} \sum_C \left( \frac{1}{ \text{thresholds} } \sum_i \frac{\text{TP}(i)}{\text{TP}(i) + \text{FP}(i)} \right)$
mAP	$\text{mAP} = \frac{1}{C} \sum_{i=1}^C \text{AP}_C$
parameters	—
FLOPs	—

## 2 3D点云语义分割

3D点云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2D语义分割相比,3D语义分割能够更细致地区分空间物体。然而,由于原始点云的无序性、密度不均匀性等,无法直接将其输入传统的卷积神经网络(convolutional neural networks, CNN)进行处理,增加了分割任务的难度。根据点云数据有无标签,将3D语义分割方法划分为

有监督分割方法和无(弱)监督分割方法. 从3D数据结构化差异的角度, 将基于有监督的深度学习3D点云语义分割方法又分为基于体素化的方法、基于多视图的方法和基于点云的方法. 基于体素化的方法和基于多视图的方法都需要进行数据预处理, 容

易造成部分信息丢失. 目前主流的基于点云的方法直接对点云信息进行学习, 主要包括逐点MLP方法、基于点卷积的方法、基于RNN的方法和基于图的方法. 图1按时间顺序给出了语义分割方法发展脉络, 图中不同颜色代表不同的语义分割方法类别.

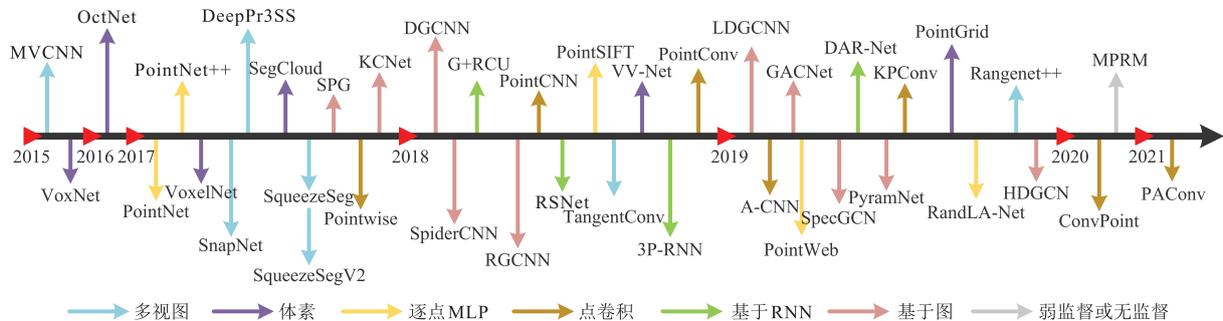


图1 3D点云语义分割方法发展脉络

## 2.1 基于体素化的方法

基于体素化方法<sup>[14,31-32]</sup>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(occupancy voxels), 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. 该方法有效地促使无序点云规则化, 但存在较高的计算和内存成本, 且由于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.

针对网格设置固定化所带来的限制, 研究人员提出了改进方法. SegCloud<sup>[19]</sup>通过三线性插值将体素级3D全卷积的神经网络与条件随机场联合, 能够学习到3D数据的细粒度预测, 降低错误划分的可能, 但难以区分颜色相似的目标, 且体素化的实现需要附加一定的成本代价. 为此, PointGrid<sup>[33]</sup>采用随机采样和填充“0”的方式保证在每个体素单元中采样恒定数量且具有结构化分布的若干个点, 有效克服了VoxNet<sup>[14]</sup>等方法在体素规模上的限制, 同时降低了SegCloud存在的成本过高的问题. 虽然其通过学习更高阶的局部近似函数能够更好地表示局部几何形状的细节, 但对几何上下文信息获取不足.

为了进一步减少不必要的内存占用及计算消耗, Kd-Net<sup>[34]</sup>、OctNet<sup>[35]</sup>引入了K-D树和八叉树结构构建计算图和共享可学习的参数对点云空间进行划分, 以划分结果为底层自下而上地逐层提取特征. 该方法可以自适应地根据点云的稠密程度分配计算和内存资源, 但对噪声点比较敏感, 且实现难度较大. VV-Net<sup>[36]</sup>利用基于内核的插值变分自动编码器和径向基函数处理局部连续表示, 可以捕获每个体素中点的分布, 丰富局部几何表示. 该方法采用与分

组卷积方法联合的方式, 有效降低了计算量, 但由于通道之间关联性不足影响分割精度.

虽然上述方法部分地改善了体素化方法存在的问题, 但总体上看, 运算效率低, 需要使用消耗较大内存的高分辨率网格以平衡点云数据稀疏性的限制, 在体素化过程中不可避免地损失几何信息导致分割精度不佳等问题, 始终限制体素化方法的进一步发展和应用.

## 2.2 基于多视图的方法

与基于体素化的方法不同, 多视图方法首先将一个3D对象按多个角度投影成多个2D视图并提取各视图特征, 然后融合这些特征以预测结果. 如何将多个角度的视图特征聚合成一个有区分力的全局表示是一个关键挑战. MVCNN<sup>[13]</sup>模拟相机从若干不同角度获取3D物体的2D图像, 再利用预训练CNN模型提取特征, 最后通过全局最大池化将不同视角的特征聚合达到形状识别的目的. 该方法所提取的特征冗余性高, 且没有很好地考虑多视角下特征之间的关系, 限制了特征描述子的可区分力.

受图像语义分割的启发, Lawin等<sup>[37]</sup>以虚拟相机将3D点云从多个视角投影至2D平面上, 将在每个视角生成的关于颜色、深度和表面法线信息的3个图像进行合成, 使用多流全卷积网络对合成图像进行像素级分数预测. 与体素化方法相比, 该方法以投影抽取信息的方式降低了内存占用, 但需要付出空间几何信息丢失的代价. SnapNet<sup>[38]</sup>通过深度分割网络将由多个虚拟相机创建的RGB图与深度信息合成的图像进行像素级的分割, 以残差校正方式融合特征, 分割效果更佳, 但没有充分利用空间上下文信息. 在SnapNet的基础上, SnapNet-R<sup>[39]</sup>以直接处理多个视

图的方法对立体图像中得到的RGB-D图像进行2D标记,并利用SnapNet密集标注3D点,能够获取更多的空间几何信息,且模型比SnapNet更容易实现,但对于目标边界处的分割效果仍需改善.

基于SqueezeNet<sup>[40]</sup>提出的SqueezeSeg<sup>[41]</sup>和SqueezeSegV2<sup>[42]</sup>使用球面投影获取点云信息.SqueezeSeg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.SqueezeSegV2在SqueezeSeg的基础上加入上下文聚合模块来减轻噪声带来的不利影响,并利用无监督域自适应训练管道来解决不同领域的的数据迁移问题,显著减少了合成数据与真实数据之间的分布差距.RangeNet++<sup>[43]</sup>利用球面投影方式处理输入点云,并采用2D CNN学习视图特征,结合学习激光扫描得到的距离得出语义标签.与上述方法相比,该方法有效改善了多视图固有的信息丢失造成的不利影响,更加准确和快速.

总体上,基于多视图的分割方法容易受虚拟视角的影响没有充分利用点云的几何结构信息,由2D信息重塑3D信息的方法必然会导致信息丢失,使得分割精度不佳.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,但该方法仍然不能解决多视图方法固有的对遮挡等场景无力处理的困境.

### 2.3 基于点云的方法

为了摆脱前述两种方法在预处理过程中损失结构性信息这一限制,学者们开始研究能够对3D点云

数据进行直接处理的方法.

#### 2.3.1 基于多层感知机MLP的方法

PointNet<sup>[15]</sup>直接将3D点云作为输入,引入T-Net结构实现点云的旋转不变,利用共享的多层感知机(multilayer perceptron, MLP)提取输入的每个点的特征,通过最大池化将所有点的信息聚合得到全局特征.该方法有效改善了原始点云旋转不变性差的问题,广泛应用于分类、部件分割和语义分割等任务,但存在过分关注全局特征而忽略局部特征的问题,没有考虑点与点之间的结构信息,也没有充分考虑到点云密度不均匀所造成的不利影响等问题,难以适应复杂场景.

针对以上问题,Qi等<sup>[16]</sup>在PointNet基础上进一步提出了PointNet++.如图2所示,该方法通过层级下采样来捕捉局部几何细节信息.此外,该方法还提出了多尺度分组和多分辨率分组的策略,以克服由于点云数据密度不均匀引起的稀疏点信息可能被忽视问题.虽然有效解决了PointNet在采样和特征提取方法上存在的问题,但仍然没有充分考虑如何有效利用点与点之间的结构信息,且所采用的 $K$ 近邻搜索方法可能使得采样点集中在一个方向上,容易出现标签错分.为了改善由于方向性限制造成的标签错分的问题,PointSIFT<sup>[44]</sup>对来自8个空间方向的信息进行堆叠和编码,借鉴了PointNet++的最远点采样策略来聚合局部特征.由于增加了参数,该方法计算量增大,同时其所提取的局部特征对于方向也过于敏感.

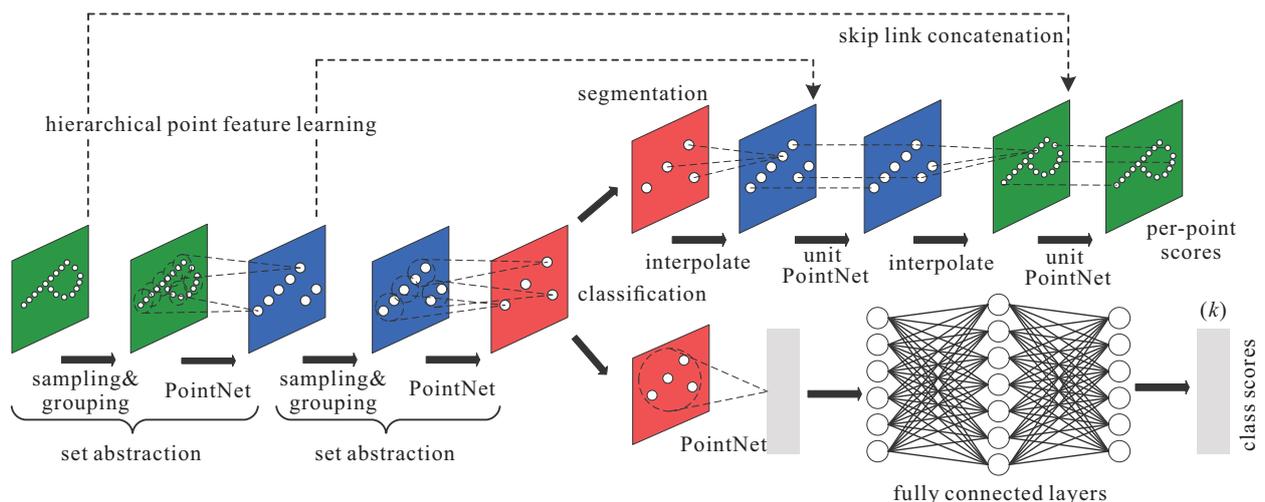


图2 PointNet++的网络架构

为了充分适应大规模点云场景,RandLA-Net<sup>[45]</sup>被提出.在该方法中,首先引入局部空间编码单元以补偿由于随机采样丢失的局部结构信息;其次利用注意力池化学习和聚合邻域点特征的方法强化重要

信息,并使用扩张残差块增加每个点的感受野以提取更丰富的特征.该方法具有优异的内存效率以及计算效率,能够保留较多细节几何结构信息,适用于处理大规模点云数据.但目前由于数据采集噪声的影

响,该方法在分割精度上的提升有限.

### 2.3.2 点卷积的方法

为了更好地通过局部拓扑结构丰富上下文表示, Hua等<sup>[46]</sup>利用逐点卷积获取点的局部信息实现语义分割,但点固定化使得网络灵活性较差. 为解决逐点卷积运算过于繁重的问题,文献[47]提出了一种包含刚性卷积和可变形卷积的核点卷积方法 **KPConv**. 刚性卷积首先取半径邻域内的点作为输入,通过计算核心点与其相邻点之间的距离确定每个核点的权值矩阵,然后利用核点卷积得出每个点的特征,将这些特征加权求和得到下采样点. 可变形卷积对刚性卷积获取的核点位置进行优化,其首先根据输入计算出一个用来优化核点位置的偏移量,先对每个核点的位置进行调整,再执行核点卷积操作. 由于 **KPConv** 中核点数是可以任意设置的,比逐点卷积等固定卷积核的方法更灵活. 类似地, **SpiderCNN**<sup>[48]</sup>、**ConvPoint**<sup>[49]</sup> 和 **A-CNN**<sup>[50]</sup> 等方法也都通过构造一种新的卷积方式对3D点云进行分割,能够较好地获取点云结构信息. 受 **KPConv** 和 **PointConv** 构造卷积核的启发,文献[51]提出了一种基于动态内核组装的位

置自适应卷积方法 **PAConv**. 首先在权重库中动态组合基本权重矩阵来构造卷积核,然后通过 **ScoreNet** 的动态内核组装策略减轻内存和计算负担. 该方法网络设计更简单,与 **PointConv** 和 **SpiderCNN** 方法相比,占用内存更小,分割效率更高.

点云数据的无序性和密度不均匀性,使得直接对其应用卷积运算的结果容易受输入点顺序的影响. 为了解决上述问题, **PointCNN**<sup>[52]</sup> 使用  $\chi$  变换矩阵对特征矩阵进行重新加权和排列改善无序性的影响,有效降低了时间复杂度和空间复杂度. 在理论上该算法可以有效解决形状信息丢失的问题,且分割效果不易受输入点顺序影响,但实际效果并未达到预期. 文献[53]提出了一种针对非均匀3D数据的新卷积方法 **PointConv**. 如图3所示,该方法将卷积核视为由权重和密度函数组成的3D点的局部坐标的非线性函数,对于给定点,通过核密度估计确定密度函数,利用 **MLP** 对密度进行重加权以确定权重函数. 该方法能够有效处理点云的不均匀性带来的不利影响,使得模型的分割性能显著提升,但邻域特征获取不充分,且时间和空间复杂度较大.

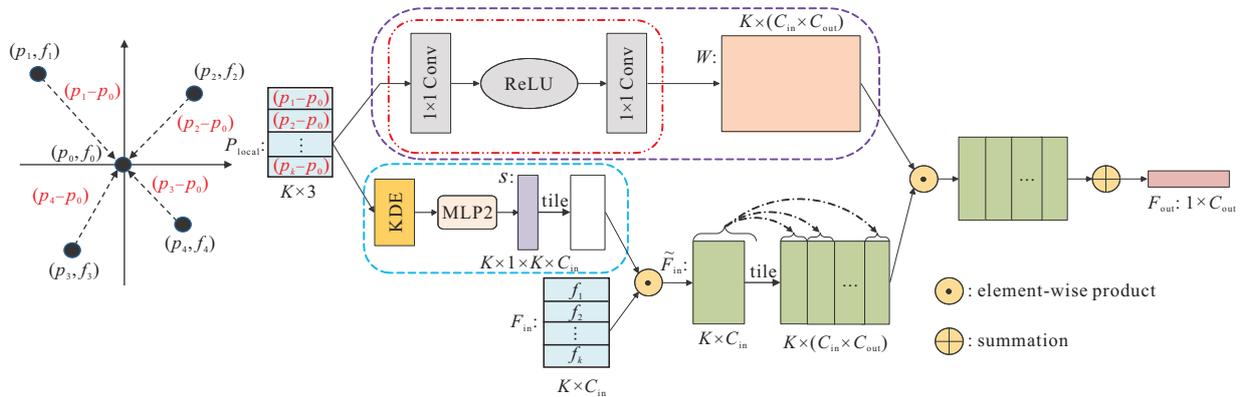


图3 PointConv的网络架构

虽然上述方法有效提升了3D点云分割效果,但均存在由于忽略目标之间的边界信息而使得分割效果不佳的问题. 为此, Gong等<sup>[54]</sup>采用具有边界感知能力的几何编码模块和边界预测模块来预测边界,充分挖掘边界信息,有效降低了目标分类错误的概率,但还是难以避免训练过程中较高的时间和空间复杂度.

### 2.3.3 基于RNN的方法

为了从点云中捕获相邻点的结构特征,一些学者提出将循环神经网络(recurrent neural network, RNN)应用于3D点云语义分割模型. Engemann等<sup>[17]</sup>将用 **PointNet** 中的点的块转换为多尺度块或网格块以获得输入级上下文,然后将 **PointNet** 提取的逐块特征依

次做合并或循环合并,以获得输出级的上下文. 循环合并能够将有关场景的信息保留,很好地提升学习效率,但该方法学习到的局部特征不足. 为了更好地解决局部几何特征提取不充分以及相邻点之间的关系信息获取不足的问题, **3P-RNN**<sup>[55]</sup> 利用逐点金字塔池化捕获不同尺度的局部上下文信息,并使用双向分层 **RNN** 融合更大范围的空间相关性数据. 该方法在室内外数据集上均取得了较好的效果,泛化能力强,但对于一些相似的语义类的区分有限,比如门和墙.

受到 **PointSIFT** 的启发, **RSNet**<sup>[56]</sup> 提出了一种将切片池化层、**RNN** 层和切片反池化层组合的轻量级局部依赖度建模模块. 如图4所示,该方法将输入的无序点云特征从  $x, y, z$  三个方向切片,利用切片池化

层将无序点特征集转换为特征向量序列并提取特征,采用双向RNN学习切片间的相关性,最后利用切片反池化层将序列中的特征分配回各个原始点.虽然该方法对于一些平面目标如桌子等有较高的分割精度,但将点云切片化会严重影响对点之间相互关系的学习,同时也没有明确每个点的局部特征学习邻域.

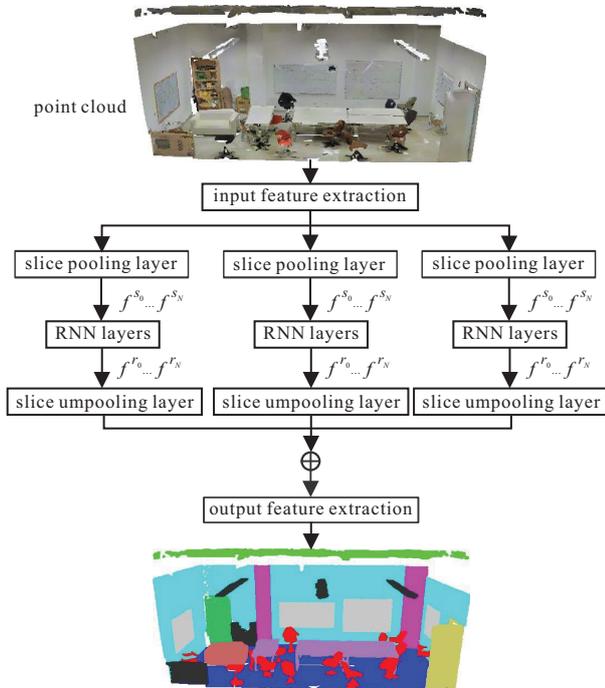


图4 RSNet的网络架构

### 2.3.4 基于图卷积的方法

基于图卷积的点云语义分割将点云中的每个点视为图的顶点,与邻域点生成图有向边,在空间域或频谱域中学习点、边特征,以捕获3D点云的局部几何结构特征.

空间域的图卷积方法充分利用了每个节点及其邻近节点信息.为了有效利用丰富的边缘特征提升大规模点云场景的分割精度,SPG<sup>[57]</sup>首先将点云划分为几何上简单但富有意义的超点集合,由超边连接相邻的超点形成超点图,将每个超点都嵌入到一个PointNet网络中,再将沿超边传递的信息在门控循环单元中细化以生成最终标签.该方法详细描述了相邻形状之间的关系,适用于处理大规模点云场景,但是将目标进行划分的步骤较难实现,容易造成分类错误.为了改善使用PointNet而存在的忽略邻域点之间的相关性的问题,DGCNN<sup>[58]</sup>提出了一个用于学习边缘特征的边缘卷积(EdgeConv),通过构建局部邻域图和对每条邻边进行EdgeConv操作,动态更新层级之间的图结构.EdgeConv可以捕捉到每个点与其邻域点的距离信息,却忽视了相邻点之间向量的方向信息,损失了部分结构信息.基于DGCNN,一系列图卷

积方法如LDGCNN<sup>[59]</sup>、HDGCN<sup>[60]</sup>、GACNet<sup>[61]</sup>等相继被提出.LDGCNN<sup>[59]</sup>将动态图不同层之间的特征级联,使用KNN和共享参数的MLP提取中心点及其邻域中的局部特征,捕获丰富的边缘向量信息.受Xception<sup>[62]</sup>可分离卷积策略的启发,为显著减少参数和计算负担,HDGCN<sup>[60]</sup>使用多个基于深度图卷积和逐点卷积的DGConv块改善分割效果.GACNet<sup>[61]</sup>受注意力机制的启发,通过为每个相邻点和通道特征分配不同的权重强调学习重要特征,但受限于点样本数过多且分布不均匀,权重分配难度较大,效果提升有限.

为了进一步捕捉高维空间内局部点之间的关系,KCNet<sup>[63]</sup>基于PointNet提出了一种用核相关和图池化联合挖掘点云局部结构信息的机制.将点集内核定义为一组可学习的3D点,由核相关性度量每个数据点的最近邻点与点集核之间的相似性,然后将最近邻点与原始点进行连接以获取局部几何结构特征图,再将局部特征图池化以探索局部特征结构.PyramNet<sup>[64]</sup>利用图嵌入模块将点云与图关联起来,用协方差矩阵来表征点与点之间的相关性,提高了对局部特征的表达能力.同时,使用金字塔注意力模块提取具有不同语义强度的特征,提高了局部特征提取的准确性.

谱域图方法将卷积定义为频谱滤波,其通过将图信号和图拉普拉斯特征向量相乘实现.SpecGCN<sup>[65]</sup>针对PointNet++忽略相邻点之间的关系信息的问题提出了一种局部谱图卷积方法,通过在局部图上使用谱图卷积获取每个点的邻域结构信息,使用图池化递归聚类谱坐标以聚合节点特征.RGCNN<sup>[66]</sup>基于谱图理论,将点云中的点特征看作是图上的信号,用切比雪夫多项式近似定义了对图的卷积.RGCNN通过在自身的每一层中更新图拉普拉斯矩阵自适应地捕获动态图结构信息,虽然相比于传统谱图卷积降低了计算复杂度,但计算复杂度受KNN的K取值的影响波动,无法适用于大规模数据集.

## 2.4 弱(无)监督语义分割

为了使弱监督方法能达到与全监督方法相近的分割精度和效果,通常采用加入先验约束或建构上下文关系挖掘和信息整合模型的方法.

为了使弱监督分割具有较强的上下文建模能力,Xu等<sup>[67]</sup>提出在训练阶段使用学习梯度近似和利用附加的空间和颜色平滑度约束实现对极少量点的标记,利用不完全监督分支和不确切监督分支对有标记的点的特征进行学习,同时引入Siamese自监督分

支学习无标记的大部分点. 该方法仅对占比为10%的标记数据进行特征学习, 却取得了近似全监督分割的效果. 为了充分捕获上下文信息, MPRM<sup>[68]</sup>引入了一个多路径区域挖掘模块, 从空间关系的各个维度挖掘每个类别的定位线索. 该模块使用3个不同注意力模块分支, 分别用于收集沿空间域的远程上下文信息、探索通道相互依赖性、聚合上下文特征, 然后在各分支中提取输出特征映射以获取点类激活映射特征图(point class activation map), 应用逐元素最大值法

获得特征图的伪点标签, 利用该伪点标签训练分割网络. 该方法仅使用部分标记点就可以达到接近监督学习方法PointNet的分割精度. 总体而言, 目前对弱监督方法的研究还处在初级阶段, 分割精度有限.

## 2.5 3D点云语义分割方法对比与结论

为了更好地展现各模型的有效性, 本文汇总了现有主流方法的分割效果并进行了分析. 表3显示了24个模型在4个基准点云数据集上的语义分割结果, 其中“-”表示该方法未提供相应的结果.

表3 不同模型在S3DIS、Semantic3D、ScanNet和SemanticKITTI数据集上的语义分割结果 %

method	S3DIS				Semantic3D		ScanNet	SemanticKITTI
	area5(OA)	area5(mIoU)	6-fold(OA)	6-fold(mIoU)	red.(OA)	red.(mIoU)	mIoU	mIoU
SegCloud <sup>[19]</sup>	—	48.9	—	—	88.1	61.3	—	—
VV-Net <sup>[36]</sup>	—	—	87.8	78.2	—	—	—	—
SnapNet <sup>[38]</sup>	—	—	—	—	88.6	59.1	—	—
TangentConv <sup>[69]</sup>	82.5	52.8	—	—	89.3	66.4	40.9	40.9
SqueezeSegV2 <sup>[43]</sup>	—	—	—	—	—	—	—	39.7
RangeNet++ <sup>[44]</sup>	—	—	—	—	—	—	—	52.2
PointNet <sup>[15]</sup>	—	41.1	78.6	47.6	—	—	—	14.6
PointNet++ <sup>[16]</sup>	—	—	81.0	54.5	—	—	33.9	20.1
PointSIFT <sup>[44]</sup>	—	—	88.7	70.2	—	—	41.5	—
PointWeb <sup>[70]</sup>	87.0	60.3	87.3	66.7	—	—	—	—
RandLA-Net <sup>[45]</sup>	—	—	87.2	68.5	94.8	77.4	—	50.3
PointCNN <sup>[52]</sup>	85.9	57.3	88.1	65.4	—	—	45.8	—
PUConv <sup>[71]</sup>	—	—	—	—	—	—	57.6	—
KPConv <sup>[47]</sup>	—	67.1	—	70.6	92.9	74.6	68.4	—
RSNet <sup>[56]</sup>	—	51.9	—	56.5	—	—	39.4	—
G+RCU <sup>[17]</sup>	—	45.1	81.1	49.7	—	—	—	—
3P-RNN <sup>[55]</sup>	85.7	53.4	86.9	56.3	—	—	—	—
DGCNN <sup>[58]</sup>	—	—	84.1	56.1	—	—	—	—
SPG <sup>[63]</sup>	86.4	58.0	85.5	62.1	94.0	73.2	—	17.4
SSP+SPG <sup>[72]</sup>	87.9	61.7	87.9	68.4	—	—	—	—
GACNet <sup>[66]</sup>	87.8	62.9	—	—	91.9	70.8	—	—
PAG <sup>[73]</sup>	86.8	59.3	88.1	65.9	—	—	—	—
HDGCN <sup>[60]</sup>	—	59.3	—	66.9	—	—	—	—
HPEIN <sup>[74]</sup>	87.2	61.9	88.2	67.8	—	—	61.8	—

从表3可以看出, S3DIS数据集的使用最为广泛, 但由于S3DIS数据集的数据量庞大, 一些场景的分割效果不佳, mIoU值整体较低. 相比较而言, 基于体素化的方法VV-Net在S3DIS数据集表现突出, 在6-fold上mIoU值达到了78.2%.

SnapNet、TangentConv、SegCloud、GACNet、KPConv、RandLA-Net和SPG等均使用了大规模城市场景数据集Semantic3D验证算法效能, 从表3可以看出以上算法在Semantic3D数据集上分割效果均表现良好. RandLA-Net采用了随机采样方法, 且为防止结

构信息丢失, 引入了局部编码模块和注意力池化, 非常适用于大规模场景, 其在几亿点的场景下, mIoU可达到77.4%, OA更是高达94.8%, 是目前运用于大场景分割中最有效的网络.

一些点卷积的方法和多视图的方法很多都选择使用真实场景下的室内数据集ScanNet验证方法的有效性. ScanNet在生成标签时对每一个场景进行了体素化使得部分结构信息损失, 因此分割效果普遍不好. 基于点卷积的方法KPConv由于可以以不定数量的点云数据直接训练模型参数而使得网络泛化能

力好,相比其他方法分割效果更佳,其mIoU达到了68.4%。

一些预处理方法如SqueezeSeg、SqueezeSegV2等对基于汽车雷达的大型户外场景数据集SemanticKITTI进行分割,分割效果较差。目前在使用了SemanticKITTI数据集的方法中,RandLA-Net的分割表现最佳,mIoU值达到了50.3%,再次验证了该方法面对大规模场景的有效性。

### 3 3D点云实例分割

3D点云实例分割是场景理解的一项具有挑战性的任务,不仅需要区分场景中细粒度更高的各类对象,还需要预测各类对象的属性。与语义分割相比,实例分割需要更精确和更高细粒度的点推理。近年来,3D点云实例分割引起广泛关注,许多方法被提出,具体可以归纳为基于Proposal(分割目标)的实例分割方法和Proposal-free的实例分割方法两类。图5为近年实例分割方法发展时间轴图。

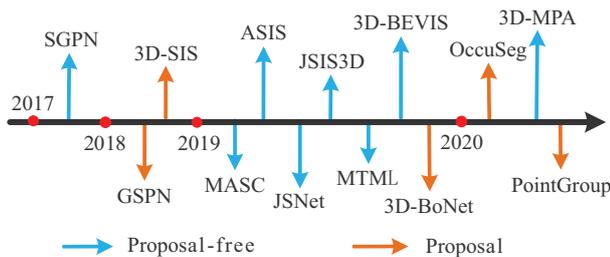


图5 具有代表性的3D点云实例分割方法脉络

#### 3.1 基于Proposal的实例分割

基于Proposal的方法将实例分割任务分解成3D目标检测和实例掩码预测两个子任务。GSPN<sup>[75]</sup>采用一种将先验网络、识别网络和生成网络相结合进行学习 and 预测的综合分析策略,通过重建场景中噪声观测的形状生成Proposal,再将其引入到基于区域的PointNet (R-PointNet)中实现灵活的Proposal优化和实例分割生成。该方法大大减少了与目标不对应的盲框Proposal,但对分割目标的细化需要昂贵的非极大值抑制操作。3D-SIS<sup>[76]</sup>首先将带有颜色信息的2D图像特征映射成体素表示,之后与带有几何信息的体素融合,继而在目标区域上预测边界框位置、对象类标签和实例掩码。该方法虽然取得了较高的分割精度,但检测网络与实例分割网络之间没有共享结构与权重,且数据处理过程复杂,无法适用于大规模场景。为此,3D-BoNet<sup>[77]</sup>提出一种基于边界框回归的点云实例分割方法。首先提取全局特征和点的特征,将全局特征直接用于预测边界框,在边界框预测结束后,一些点会落入对应的边界框中,再根据点的特征

与边界框预测分数对框内的点进行二分类,从而分割出实例目标。该方法明确地预测目标边界框,最终得出的实例具有高目标性,但两阶段的训练步骤使网络运算负担较大。

为了进一步获取更高级别的点特征,3D-MPA<sup>[78]</sup>首先采用稀疏卷积作为主干网络提取每个点的特征,然后利用图卷积提取相邻的Proposal之间的特征关系,同时对Proposal的特征进行细化,最后通过特征聚类得到最终的分割结果。该方法分割性能优于PointGroup<sup>[79]</sup>,但图卷积和PointNet网络的叠加使用,使得该网络时间和空间复杂度较大,且难以进一步优化。

总之,基于Proposal的方法虽然直观、分割精度高,但通常需要两阶段的训练,并依赖于使用非极大值抑制等后处理操作对密集的候选框进行选择,比较耗时且计算量大。

#### 3.2 Proposal-free的实例分割

Proposal-free的实例分割方法通过摒弃了基于Proposal的实例分割方法中的目标检测的环节进一步简化了分割任务,采取提取数据特征的方法进一步得到实例分割结果,或者借助于语义分割结果完成实例分割。

SGPN<sup>[80]</sup>首先使用PointNet提取点云的特征,以结构上的相似性矩阵、置信度分支和语义预测分支3个功能模块,分别用于产生点聚类Proposal、选择可用的Proposal和给每个点预测一个类别标签,再使用非极大值抑制噪声并对重叠剪枝,以生成预测结果。该方法能够处理生成Proposal时产生的冗余,减少噪声点,但相似矩阵的构造需要大量的内存消耗,且通过相似性方法获取的不全是有用信息,难以扩展到大规模数据。针对SGPN存在由相似矩阵引起的无效学习问题,MASC<sup>[81]</sup>利用U-Net和子流形稀疏卷积处理体素化的点云,对每个体素进行语义分割级的类别预测,然后对不同尺度下相邻体素进行关联性预测。该方法通过预测多个尺度上相邻点之间的相似性用以代替SGPN的相似性矩阵,相比起来,该方法更高效,但由于没有挖掘各个尺度对网络性能的影响,且模型依赖于SparseConvNet<sup>[82]</sup>网络,使得网络性能提升空间有限。

为了进一步对复杂场景进行实例分割,MTML<sup>[83]</sup>使用改进的SSCNet网络<sup>[84]</sup>处理输入的体素网格,通过学习抽象的嵌入特征,将具有相同实例标签的体素在特征空间中紧密聚合,将具有不同目标实例的体素远离,然后估计每个体素的方向信息以学

习实例信息. 由于该方法主要针对体素场景, 普适性较差, 容易丢失几何信息. Zhang等<sup>[85]</sup>在点云实例分割中引入多元高斯分布对每个点进行相似性度量, 并结合不确定性估计引入用于聚类的logDice损失函数处理数据, 有效提升了模型对同类别细粒实例的区分能力. PointGroup<sup>[79]</sup>通过探索目标实例之间的空隙对点进行分组, 利用双分支网络来提取点特征、预测点语义标签和点偏移量, 将基于原始点坐标和偏移位移的点坐标聚类, 实现目标分离, 最后由ScoreNet网络来评估候选实例, 优化分割结果. 该方法有效获取了更多有用信息, 但去除冗余的非极大值抑制法不利于有效数据的保留.

为了有效地融合语义分割和实例分割两个任务的特征以实现互利寻优, ASIS<sup>[86]</sup>和JSNet<sup>[87]</sup>构建了一种由共享编码器和两个并行解码器组成的基准网络, 同时处理实例分割和语义分割两个任务. 通过一种联合实例语义分割模块将语义特征转入实例嵌入空间, 再将转换后的特征进一步与实例特征融合来提升实例分割效果. 该方法使得两个任务能够互利共赢, 但特征融合会增加模型负担, 且对上下文信息获取不充分. JSIS3D<sup>[88]</sup>和3D-BEVIS<sup>[89]</sup>均同时进行语义和实例分割, 但没有挖掘和利用语义分割和实例分割之间的关系进行互利, 使得分割性能提升幅度较小.

综上, Proposal-free的方法不需要预先生成Proposal, 计算量小, 但由于未明确检测目标边界, 容易导致丢失有用信息, 总体上分割效果不佳.

### 3.3 实例分割方法对比及结论

表4显示了在ScanNet(v2)数据集上的3D点云实例分割结果.

表4 ScanNet(v2)数据集上的3D点云实例分割结果 %

ScanNet(v2)	年份	验证集			隐藏测试集		
		mAP @0.25	mAP @0.5	mAP	mAP @0.25	mAP @0.5	mAP
SGPN <sup>[80]</sup>	2017	22.2	11.3	—	39.0	14.3	4.9
3D-BEVIS <sup>[89]</sup>	2019	—	—	—	40.1	24.8	11.7
3D-SIS <sup>[76]</sup>	2019	35.7	18.7	—	55.8	38.2	16.1
GSPN <sup>[75]</sup>	2018	53.4	37.8	19.3	54.4	30.6	15.8
MTML <sup>[83]</sup>	2019	55.4	40.2	20.3	73.1	54.9	28.2
3D-BoNet <sup>[77]</sup>	2019	—	—	—	68.7	48.8	25.3
MASC <sup>[81]</sup>	2019	—	—	—	—	44.7	—
3D-MPA <sup>[78]</sup>	2020	72.4	59.1	35.3	73.7	61.1	35.5
OccuSeg <sup>[90]</sup>	2020	—	—	—	73.9	63.4	44.3
PointGroup <sup>[82]</sup>	2020	—	—	—	—	63.6	—

在ScanNet(v2)数据集上进行3D点云实例分割, 以重叠25%时的平均精度(mAP@0.25)、重叠50%时的平均精度(mAP@0.5)和mAP作为其评价指标. 3D-MPA<sup>[82]</sup>用在Proposal上学到的高级特征对Proposal进行聚类实现简化各目标区域, 能够在简化区域产生过程的同时很好地处理离群的目标区域, 计算更高效, 在验证集上表现最优, 其mAP@0.25、mAP@0.5和mAP分别达到了72.4%、59.1%、35.3%. OccuSeg<sup>[90]</sup>采用了一种将输入体素转换为超体素的策略. 与2D空间中的超像素相比, 由于具有几何连续性和局部凸度约束, 3D空间中的实例边界更易于识别, 因此超体素能够更好地分离不同的实例. 该方法在3个度量上都有很大的优势, 特别是mAP值在隐藏测试集上高达44.3%, 说明了其所采用策略的有效性.

## 4 3D点云部件分割

部件分割是一项具有挑战性的3D数据理解和分析任务, 其目标是预测给定形状中每个部件的类别标签. 基于深度学习的部件分割方法的分割精度相比传统方法有很大提升, 但仍然还有一些问题如如何有效获取具有辨别性的细粒度特征等亟待解决, 充分挖掘和融合多层次信息以增强模型的识别能力成为3D点云部件分割研究的重要方向.

受2D图像分割的启发, Kalogerakis等<sup>[91]</sup>基于图像的全卷积网络(fully convolutional networks, FCN)和基于表面的条件随机场(conditional random field, CRF)进行端到端的3D点云部件分割. 首先在不同尺度上计算能够最大限度地覆盖所查看的形状表面的视角, 再以这些视角获取阴影图像和深度图像输入FCNs网络, 将输出的标签置信度映射到Image2Surface投影层聚合并反投影回3D形状表面上, 借助CRF促使目标表面的标记保持一致. 该方法采用基于视图方法, 无法充分获取形状的深层结构特征, 计算量较大. 为了能够在有限分辨率下提取更多具有鉴别性的特征, VoxSegNet<sup>[92]</sup>将形状体素化, 利用由堆叠的残差块组成的空间密集提取模块从稀疏的体素数据中提取多尺度特征, 利用注意力特征聚合模块对从不同抽象层次提取的特征进行聚合. 该方法有效缓解了多视图方法造成的几何信息丢失问题, 但从稀疏数据中获取可区分性特征仍然较难, 因此如何在分辨率受限的条件下进行更精细的特征提取值得进一步研究.

受全卷积网络<sup>[93]</sup>的启发, Wang等<sup>[94]</sup>首先提取平均测地距离<sup>[95]</sup>、形状上下文<sup>[96]</sup>和自旋图像<sup>[97]</sup>3

类常用的形状几何特征,利用 SFCN(shape fully convolutional networks)网络,通过跳跃连接将深度粗层的结构信息与浅层细致的外观信息结合以产生精确的分割结果,再依据特征之间互补性和多标签图剪枝方法对分割结果进行优化.与上述方法相比,该方法泛化能力更好,但计算中设计的形状必须是流形网格,故无法用于处理较大规模场景. BAE-NET<sup>[98]</sup>引入分支自动编码器来学习特定部件形状的紧凑表示,将学习到的特征送到解码器处理,并利用最大池化对输出特征聚合.该方法可以应对很多类3D物体形状,但对初始参数比较敏感,且对相似的部件容易混淆,阻碍了其对于更细粒度部件的有效分割.

ShapeNet part 是一个由3D CAD模型对象表示的丰富注释的大型部件存储库,在部件分割任务中广泛应用.表5显示了不同方法在ShapeNet part 数据集上的分割结果对比.从表5可以发现,选用ShapeNet part 数据集进行验证的各类算法,mIoU值基本在80%以上.V3DPC<sup>[99]</sup>结合Delaunay三角剖分图绘制算法和多尺度U-Net集成的方法,保留了点云中各点之间的结构信息,instance mIoU值达到了88.8%,分割效果表现最佳.整体而言,目前算法对细粒度的3D物体有较好的识别效果,对部件的分割结果更加接近于Ground Truth.

表5 在ShapeNet part 数据集上比较部件分割结果 %

method	年份	class mIoU	instance mIoU
PointNet <sup>[15]</sup>	2017	77.4	83.7
PointNet++ <sup>[16]</sup>	2017	81.9	85.1
Kd-Net <sup>[34]</sup>	2017	77.4	82.3
PointCNN <sup>[52]</sup>	2018	84.6	86.1
RS-Net <sup>[56]</sup>	2018	81.4	84.9
KCNet <sup>[63]</sup>	2018	82.2	84.7
DGCNN <sup>[58]</sup>	2018	82.3	85.1
RGCNN <sup>[66]</sup>	2018	79.5	84.3
VoxSegNet <sup>[92]</sup>	2018	84.8	87.5
PointGrid <sup>[33]</sup>	2018	82.2	86.4
SpiderCNN <sup>[48]</sup>	2018	82.4	85.3
LDGCNN <sup>[59]</sup>	2019	81.6	85.1
PyramNet <sup>[64]</sup>	2019	81.6	83.9
PointConv <sup>[53]</sup>	2019	82.8	85.7
$\Psi$ -CNN <sup>[100]</sup>	2019	83.4	86.8
RS-CNN <sup>[101]</sup>	2019	84.0	86.2
KPCov <sup>[47]</sup>	2019	85.1	86.4
V3DPC <sup>[99]</sup>	2020	84.6	88.8
PAConv <sup>[51]</sup>	2021	84.6	86.1

## 5 总结与展望

本文将近几年3D点云语义分割、实例分割与部件分割方法的研究进展进行了归纳.相比于传统方法<sup>[102]</sup>,基于深度学习的方法很大程度提升了分割效果,但从目前的研究成果上看,仍然有很多问题尚未得到很好地解决.本文在总结目前存在的问题和挑战的基础上提出可能的解决方法和思路.

1) 点云的无序性和密度不均匀性使得3D点云分割具有很强的挑战性,现有的方法均存在一定的局限性.

### ① 点云的无序性.

根本上,点云是由一系列带有位置坐标信息的点组成的,其无序性首先表现为不具备逐点视角上的排列不变性,其次表现为不具备结构视角上的旋转不变性.

在改善排列不变性方面,目前的主流方法是先通过如对称函数(PointNet)、图卷积(DGCNN)、 $\chi$ -变换卷积算子(PointCNN)等将数据有序化,以保证输入分割模型的数据具有排列不变性,但仍存在局部信息学习不足、计算复杂度增加等问题. PCT<sup>[103]</sup>和PT<sup>[104]</sup>引入了自然语言处理中的一种能够使数据顺序化的Transformer<sup>[105]</sup>机制,并结合Attention机制,在分类分割数据集上均表现良好,但增加了模型的计算负担. Transformer-XL<sup>[106]</sup>机制进一步改进了Transformer机制,将输入特征进行切分,使用循环机制来学习连续段之间的长期依赖关系,使输入点云数据具有排列不变性的同时有效降低了模型复杂度.可以考虑在已有的针对点云处理的算法中普遍融入该机制.

在改善旋转不变性方面,基于多视图的方法和基于体素化的方法虽然能够改善点云的非结构化问题,但容易造成有用信息丢失和计算量增加,分割性能提升受限;基于图卷积的方法通过在点与点之间构造图模型来解决点云的非结构化问题,能够避免结构信息丢失的问题,但构造图的方法明显增加了计算和存储负担.轻量级网络EfficientNet<sup>[107]</sup>提出一种复合扩张方法,通过对模型的宽度、深度和图片分辨率的同时缩放,减少了参数量,轻量化了模型,取得了惊艳的效果.可以考虑将图卷积算法与该方法相结合,在解决网络无序性问题的同时有效避免计算负担.

### ② 点云的密度不均匀性.

3D点云数据的密度不均匀性可能导致采样过程中稀疏点处关键几何信息被忽略,处理密度不均

匀性带来的不利影响是提高点云分割效率的重要途径之一。目前主流的处理方法是通过调整采样密度(PointNet++、PointConv)或改变采样方法(RandLA-Net)来改善密度不均匀特性,虽然在一定程度上缓解了有用信息缺失的问题,但这些方法对密集区域更有效,可能存在重要几何信息的稀疏处的数据被忽略或者利用较少。SSC<sup>[108]</sup>将稀疏卷积(SC)和有效稀疏卷积(VSC)作用于稀疏点,使用VSC卷积、分层SC卷积和稀疏池化操作的组合来构建VGG、ResNet和DenseNet等流行卷积网络的稀疏卷积,有效避免了稀疏点信息的丢失。可以考虑将这种3D稀疏卷积与现有的基于点的方法相结合,充分提取稀疏数据中的信息。

2) 大多数现有的语义分割和实例分割存在模型结构复杂、成本高昂的问题,未来应该多考虑网络模型的轻量化。

现有的大多数语义分割和实例分割方法均存在模型结构复杂且训练参数多的问题,学者们针对语义分割任务提出了一些轻量化模型的方法,如在RandLA-Net方法中提出一种将随机采样与局部特征聚合模块结合的机制,可以使网络模型轻量化,同时有效提升分割精度。相比之下,目前针对实例分割网络的轻量化研究更是处于起步阶段。在2D图像处理中,采用Group Convolution<sup>[109]</sup>和Channel Shuffle<sup>[110]</sup>结合的方式可以降低参数量,提升特征的代表能力,同时允许信息在通道之间流动。可以考虑借鉴该方法,将类似机制与3D点云语义分割和实例分割算法结合,在降低模型复杂度的同时提升通道之间的相互联系,获取更多上下文信息,在不损失甚至提升分割精度的同时轻量化网络。

3) 大多数现有的语义分割模型难以保证实时性,限制了其在各个实际场景中的应用,可以考虑融合实时性较好的网络以改善现有模型的实时性。

无人驾驶等一些实际应用领域需要实时感知场景中的目标,但目前很少有算法能够满足实时性<sup>[111]</sup>要求,已有高精度的大模型都难以达到实时数据处理的要求。SqueezeSeg基于轻量级的卷积神经网络SqueezeNet实现了对3D点云场景的实时语义分割,运算速度快且稳定,在大规模数据集KITTI上单帧检测时间降低到20ms,但是由于需要预先将3D点云经过球面投影得到前视图,投影过程中容易导致目标细节信息丢失而影响分割精度。对此,可以考虑将SqueezeNet网络和基于点的方法结合,借助

SqueezeNet使网络轻量化,同时利用基于点的方法尽可能保留细节几何信息,在保证准确率的同时提高实时性。

4) 人工标注大规模数据成本高昂,可以考虑采用弱监督或无监督方法降低标注难度。

目前公开可用的数据集有限的语义标注和相对较小的采集空间范围限制了其在大场景3D点云细粒度语义理解方面的发展,发展覆盖大场景和大数据的大规模数据集成为进一步深入研究的必然需求。而逐点数据标注成本高昂,不适应真实大规模场景。一些学者提出将弱监督或者无监督的分割技术应用于3D点云语义分割中,使网络在大规模的无标记/较少标记的数据集上应用,以规避标注工序,但目前还处于初级发展阶段。在未来工作中,可以考虑基于自监督的思想,将目前已有的分割算法与自监督等变注意力机制(self-supervised equivariant attention mechanism, SEAM)<sup>[112]</sup>相结合进行弱监督3D点云分割,借助自监督学习节省掉标注成本。

5) 现有采样策略容易引起细节信息丢失、计算代价高等问题,可以考虑通过优化采样方法加以改善。

现有的大多数算法所采用的采样策略存在计算代价高、内存占用大、细节信息易丢失的问题,难以处理大规模点云场景。比如PointNet++、PointConv、PointCNN等采用非学习式的最远点采样的方法,使用较少的特征点即可覆盖整个空间特征,但存在计算复杂度高和可扩展性差的问题;RandLA-Net采用随机采样的方法降低了时间复杂度和内存消耗,但难以避免信息丢失。相比上述方法,SampleNet<sup>[113]</sup>提出一种可微近似点云采样方法,比最远点采样的方法采样更均匀、结构信息更丰富,而且降低了信息丢失的概率,目前在分类、重建等任务上均取得了不错的结果。可以考虑将该方法或机制引入到3D点云分割任务中,改善采样方法从而促进分割效果的提升。

6) Proposal-free的实例分割目标边界模糊,容易导致边界处目标分割错误,可以考虑引入边界感知方法加以改善。

现有Proposal-free的实例分割方法没有明确检测目标边界,面对多个目标的杂乱场景和类间模糊场景会产生不准确的边界定位和严重的分类错误。在2D图像领域,Zhao等<sup>[114]</sup>引入了边界感知和语义感知的两个轻量级的模块,解决了多类目标部件特征提取时存在的点和标签混淆问题。该方法网络规模较

小且易于实现,可以考虑借鉴该方法的思想,加入边界感知机制来区分边界处点的标签,避免目标被错误分割.

7) 如何有效解决噪声对分割效果的不利影响仍然是一项有挑战性的任务,可以考虑采用预先滤波的方法加以处理.

从对场景的可视化结果可以看出,由激光雷达等扫描得到的3D点云数据存在大量的噪声<sup>[115]</sup>,尤其是在室外场景中是不可避免的.在对输入点云进行采样时,有一些特殊类别很容易采集到很多的噪声点,使得这些类别的分割精度偏低甚至接近于零,如S3DIS数据集中的beam类.对噪声的处理一般在点云特征编码阶段,为了能够尽量降低噪声对分割效果的影响,可以考虑通过预先对离群点、噪声点等异常点仿照体素滤波的方法<sup>[116]</sup>进行下采样处理,尽可能将异常点从场景中分离出去,将清洗过的数据再输入分割网络中.

8) 目前语义分割任务缺少序列数据集,使得基于视频的语义分割研究发展缓慢,需要鼓励构建基于视频序列的数据集以适应实际需求.

3D视频中蕴藏着丰富的现实世界信息,从视频数据中提取到的时间特征和空间特征能够为点云语义分割提供更高维度的有关环境的精确几何信息,帮助主体更好地理解 and 识别环境.如自动驾驶等需要对其附近的建筑和物体具有细粒度的理解能力的应用领域,就需要处理大量视频数据.然而,目前可用的3D视频数据集很少,与实际需求严重脱轨,基于视频的语义分割研究发展缓慢.SemanticKITTI是已有的最大的一个包含28类的4D点云序列数据集,但实际应用中可能出现的类别种类远超其所囊括.为了进一步促进算法的研究和应用,需要进一步采集和标注含有更多场景和类别的数据量更丰富的点云序列数据集,这将成为推动基于视频的3D点云分割算法研究进一步发展的基础设施.

9) 现有的大多数语义分割模型无法捕获时间信息,难以充分挖掘4D点云场景数据信息,可以考虑借鉴时序上下文融合模块等予以改善.

在构建了上述大规模4D点云序列数据集的基础上,如何充分捕获时空信息将是另一个重要的挑战,目前如何利用4D点云场景信息分割目标还未得到充分的研究.使用现有的3D点云语义分割模型处理4D点云数据时,会因为未利用时间信息而无法得到满意的4D场景目标的分割精度.

SpSequenceNet<sup>[117]</sup>引入跨帧全局注意模块和跨帧局部插值模块来捕捉4D点云序列中的时间信息,在SemanticKITTI数据集上的分割精度高于仅提取空间特征的方法,但在高频帧下的计算量较大.针对2D图像的视频语义分割方法VSPW<sup>[118]</sup>引入了一种时序上下文融合(temporal context blending, TCB)模块,通过提取时间-空间上下文信息来增强特征,从而有效提升了分割精度.可以考虑在现有3D点云语义分割模型中引进类似机制兼顾挖掘时间和空间信息.

10) 现有的语义分割单一方法普遍存在结构信息易丢失、计算代价高等缺陷,可以考虑通过方法融合的方式进行补偿.

结合前文内容可知,预处理数据的方法或直接处理数据的方法在单一使用时均存在一定的局限性.前者容易造成信息丢失;后者是目前广泛使用、效果最佳的一种方法,但其下采样过程需要查找邻域点,减缓了运算速度.一些学者提出将基于点的方法和基于体素的方法融合的网络FusionNet<sup>[119]</sup>,形成独特的基于体素的“mini-PointNet”进行点云表示和学习,可以同时实现邻域体素特征聚合和细粒度的逐点特征学习,有效利用基于体素化方法对点云规则化的处理能力和基于点的方法的上下文表征能力提高分割精度.与单一基于体素化的网络相比,该方法可以获得更准确的逐点预测结果;与逐点卷积方法相比,该方法以更小的代价实现了更有效的特征聚合.在未来的研究中,可以考虑通过构造类似的方法融合网络,优势互补地总体提升分割效果.

## 参考文献(References)

- [1] Qi C R, Liu W, Wu C, et al. Frustum PointNets for 3D object detection from RGB-D data[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 918-927.
- [2] Park Y, Lepetit V, Woo W. Multiple 3D object tracking for augmented reality[C]. The 7th IEEE/ACM International Symposium on Mixed and Augmented Reality. Cambridge, 2008: 117-120.
- [3] Xi X H, Wan Y P, Wang C. Building boundaries extraction from points cloud using an image edge detection method[C]. 2016 IEEE International Geoscience and Remote Sensing Symposium. Beijing, 2016: 1270-1273.
- [4] Ni H, Lin X G, Ning X, et al. Edge detection and feature line tracing in 3D-point clouds by analyzing geometric properties of neighborhoods[J]. Remote Sens, 2016, 8: 710.
- [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: 88-100.
- [6] Fan S N, Huang N, Fang P F, et al. A 3D point cloud segmentation method based on local convexity and dimension features[C]. 2018 Chinese Control and Decision Conference (CCDC). Shenyang, 2018: 5012-5017.
- [7] Klasing K, Wollherr D, Buss M. A clustering method for efficient segmentation of 3D laser data[C]. *IEEE International Conference on Robotics and Automation*. Pasadena, 2008: 4043-4048.
- [8] Hao W, Wang Y, Ning X, et al. Automatic building extraction from terrestrial laser scanning data[J]. *Advances in Electrical and Computer Engineering*, 2013, 13(3): 11-16.
- [9] Schnabel R, Wahl R, Klein R. Efficient RANSAC for point-cloud shape detection[J]. *Computer Graphics Forum*, 2007, 26(2): 214-226.
- [10] Wang Y M, Shi H B. A segmentation method for point cloud based on local sample and statistic inference[C]. *Geo-Informatics in Resource Management and Sustainable Ecosystem*. Berlin: Springer, 2014: 274-282.
- [11] Strom J, Richardson A, Olson E. Graph-based segmentation for colored 3D laser point clouds[C]. 2010 *IEEE/RSJ International Conference on Intelligent Robots and Systems*. Taipei, 2010: 2131-2136.
- [12] Yang J Y, Gan Z Q, Li K, et al. Graph-based segmentation for RGB-D data using 3D geometry enhanced superpixels[J]. *IEEE Transactions on Cybernetics*, 2015, 45(5): 913-926.
- [13] Su H, Maji S, Kalogerakis E, et al. Multi-view convolutional neural networks for 3D shape recognition[C]. 2015 *IEEE International Conference on Computer Vision*. Santiago, 2015: 945-953.
- [14] 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 (IROS)*. Hamburg, 2015: 922-928.
- [15] Charles R Q, Hao S, Mo K C, et al. PointNet: Deep learning on point sets for 3D classification and segmentation[C]. 2017 *IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, 2017: 77-85.
- [16] Qi C R, 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, 2017: 5099-5108.
- [17] Engelmann F, Kontogianni T, Hermans A, et al. Exploring spatial context for 3D semantic segmentation of point clouds[C]. 2017 *IEEE International Conference on Computer Vision Workshops*. Venice, 2017: 716-724.
- [18] Cheng H Z, Lu J, Luo M X, et al. PTANet: Triple attention network for point cloud semantic segmentation[J]. *Engineering Applications of Artificial Intelligence*, 2021, 102: 104239.
- [19] Tchapmi L, Choy C, Armeni I, et al. SEGCloud: Semantic segmentation of 3D point clouds[C]. 2017 *International Conference on 3D Vision (3DV)*. Qingdao, 2017: 537-547.
- [20] Dai A, Chang A X, Savva M, et al. ScanNet: Richly-annotated 3D reconstructions of indoor scenes[C]. 2017 *IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, 2017: 2432-2443.
- [21] Armeni I, Sener O, Zamir A R, et al. 3D semantic parsing of large-scale indoor spaces[C]. 2016 *IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, 2016: 1534-1543.
- [22] Hackel T, Savinov N, Ladicky L, et al. Semantic3D. Net: A new large-scale point cloud classification benchmark[J/OL]. 2017, arXiv: 1704.03847.
- [23] Behley J, Garbade M, Milioto A, et al. SemanticKITTI: A dataset for semantic scene understanding of LiDAR sequences[C]. 2019 *IEEE/CVF International Conference on Computer Vision (ICCV)*. Seoul, 2019: 9296-9306.
- [24] Yi L, Kim V G, Ceylan D, et al. A scalable active framework for region annotation in 3D shape collections[J]. *ACM Transactions on Graphics*, 2016, 35(6): 1-12.
- [25] Roynard X, Deschaud J E, Goulette F. Paris-Lille-3D: A large and high-quality ground-truth urban point cloud dataset for automatic segmentation and classification[J]. *The International Journal of Robotics Research*, 2018, 37(6): 545-557.
- [26] Mo K, Zhu S, Chang A X, et al. PartNet: A large-scale benchmark for fine-grained and hierarchical part-level 3D object understanding[C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach, 2019: 909-918.
- [27] Hua B S, Pham Q H, Nguyen D T, et al. SceneNN: A scene meshes dataset with annotations[C]. 2016 *Fourth International Conference on 3D Vision (3DV)*. Stanford, 2016: 92-101.
- [28] 张佳颖, 赵晓丽, 陈正. 基于深度学习的点云语义分割综述[J]. *激光与光电子学进展*, 2020, 57(4): 28-46. (Zhang J Y, Zhao X L, Chen Z. Review of semantic segmentation of point cloud based on deep learning[J]. *Laser & Optoelectronics Progress*, 2020, 57(4): 28-46.)
- [29] Li Y, Ma L F, Zhong Z L, et al. Deep learning for LiDAR point clouds in autonomous driving: A review[J]. *IEEE*

- Transactions on Neural Networks and Learning Systems, 2021, 32(8): 3412-3432.
- [30] Guo Y L, Wang H Y, Hu Q Y, et al. Deep learning for 3D point clouds: A survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(12): 4338-4364.
- [31] Zhou Y, Tuzel O. VoxelNet: End-to-end learning for point cloud based 3D object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 4490-4499.
- [32] Qi C R, Su H, NieSSner M, et al. Volumetric and multi-view CNNs for object classification on 3D data[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, 2016: 5648-5656.
- [33] Le T, Duan Y. PointGrid: A deep network for 3D shape understanding[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 9204-9214.
- [34] Klokov R, Lempitsky V. Escape from cells: Deep kd-networks for the recognition of 3D point cloud models[C]. 2017 IEEE International Conference on Computer Vision. Venice, 2017: 863-872.
- [35] Riegler G, Ulusoy A O, Geiger A. OctNet: Learning deep 3D representations at high resolutions[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, 2017: 6620-6629.
- [36] Meng H Y, Gao L, Lai Y K, et al. VV-Net: Voxel VAE net with group convolutions for point cloud segmentation[C]. 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, 2019: 8499-8507.
- [37] Lawin F J, Danelljan M, Tosteberg P, et al. Deep projective 3D semantic segmentation[C]. Computer Analysis of Images and Patterns. Cham: Springer, 2017: 95-107.
- [38] Boulch A, Guerry J, le Saux B, et al. SnapNet: 3D point cloud semantic labeling with 2D deep segmentation networks[J]. Computers & Graphics, 2018, 71: 189-198.
- [39] Guerry J, Boulch A, le Saux B, et al. SnapNet-R: Consistent 3D multi-view semantic labeling for robotics[C]. 2017 IEEE International Conference on Computer Vision Workshops. Venice, 2017: 669-678.
- [40] 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/OL]. 2016, arXiv: 1602.07360.
- [41] 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. Brisbane, 2018: 1887-1893.
- [42] 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, 2019: 4376-4382.
- [43] 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 (IROS). Macao, 2019: 4213-4220.
- [44] Jiang M, Wu Y, Lu C. Pointsift: A sift-like network module for 3D point cloud semantic segmentation[J/OL]. 2018, arXiv: 1807.00652.
- [45] Hu Q Y, Yang B, Xie L H, et al. RandLA-Net: Efficient semantic segmentation of large-scale point clouds[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, 2020: 11105-11114.
- [46] Hua B S, Tran M K, Yeung S K. Pointwise convolutional neural networks[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 984-993.
- [47] Thomas H, Qi C R, Deschard J E, et al. KPConv: Flexible and deformable convolution for point clouds[C]. 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, 2019: 6410-6419.
- [48] Xu Y F, Fan T Q, Xu M Y, et al. SpiderCNN: Deep learning on point sets with parameterized convolutional filters[C]. Proceedings of the European Conference on Computer Vision (ECCV). Munich, 2018: 87-102.
- [49] Boulch A. ConvPoint: Continuous convolutions for point cloud processing[J]. Computers & Graphics, 2020, 88: 24-34.
- [50] Komarichev A, Zhong Z, Hua J. A-CNN: Annularly convolutional neural networks on point clouds[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, 2019: 7421-7430.
- [51] Xu M T, Ding R Y, Zhao H S, et al. PAConv: Position adaptive convolution with dynamic kernel assembling on point clouds[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, 2021: 3172-3181.
- [52] Li Y, Bu R, Sun M, et al. PointCNN: Convolution on x-transformed points[C]. Advances in neural information processing systems. Montreal, 2018: 820-830.
- [53] Wu W X, Qi Z A, Li F X. PointConv: Deep convolutional networks on 3D point clouds[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, 2019: 9613-9622.
- [54] Gong J Y, Xu J C, Tan X, et al. Boundary-aware geometric encoding for semantic segmentation of point clouds[J]. Proceedings of the AAAI Conference on

- Artificial Intelligence, 2021, 35(2): 1424-1432.
- [55] Ye X Q, Li J M, Huang H X, et al. 3D recurrent neural networks with context fusion for point cloud semantic segmentation[C]. Proceedings of the European Conference on Computer Vision (ECCV). Munich, 2018: 403-417.
- [56] Huang Q G, Wang W Y, Neumann U. Recurrent slice networks for 3D segmentation of point clouds[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 2626-2635.
- [57] Landrieu L, Simonovsky M. Large-scale point cloud semantic segmentation with superpoint graphs[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 4558-4567.
- [58] Wang Y, Sun Y B, Liu Z W, et al. Dynamic graph CNN for learning on point clouds[J]. ACM Transactions on Graphics, 2019, 38(5): 1-12.
- [59] Zhang K G, Hao M, Wang J, et al. Linked dynamic graph CNN: Learning through point cloud by linking hierarchical features[C]. The 27th International Conference on Mechatronics and Machine Vision in Practice (M2VIP). Shanghai, 2021: 7-12.
- [60] Liang Z D, Yang M, Deng L Y, et al. Hierarchical depthwise graph convolutional neural network for 3D semantic segmentation of point clouds[C]. International Conference on Robotics and Automation (ICRA). Montreal, 2019: 8152-8158.
- [61] Wang L, Huang Y C, Hou Y L, et al. Graph attention convolution for point cloud semantic segmentation[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, 2019: 10288-10297.
- [62] Chollet F. Xception: Deep learning with depthwise separable convolutions[C]. IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, 2017: 1800-1807.
- [63] Shen Y R, Feng C, Yang Y Q, et al. Mining point cloud local structures by kernel correlation and graph pooling[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 4548-4557.
- [64] Kang Z, Li N. PyramNet: Point cloud pyramid attention network and graph embedding module for classification and segmentation[J/OL]. 2019, arXiv: 1906.03299.
- [65] Wang C, Samari B, Siddiqi K. Local spectral graph convolution for point set feature learning[C]. Proceedings of the European Conference on Computer Vision (ECCV). Munich, 2018: 52-66.
- [66] Te G S, Hu W, Zheng A M, et al. RGCNN: Regularized graph CNN for point cloud segmentation[C]. MM'18: Proceedings of the 26th ACM International Conference on Multimedia. Seoul, 2018: 746-754.
- [67] Xu X, Lee G H. Weakly supervised semantic point cloud segmentation: Towards  $10 \times$  fewer labels[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, 2020: 13703-13712.
- [68] Wei J C, Lin G S, Yap K H, et al. Multi-path region mining for weakly supervised 3D semantic segmentation on point clouds[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, 2020: 4383-4392.
- [69] Tatarchenko M, Park J, Koltun V, et al. Tangent convolutions for dense prediction in 3D[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 3887-3896.
- [70] Zhao H S, Jiang L, Fu C W, et al. PointWeb: Enhancing local neighborhood features for point cloud processing[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, 2019: 5560-5568.
- [71] Lu J, Cheng H Z, Luo M X, et al. PUConv: Upsampling convolutional network for point cloud semantic segmentation[J]. Electronics Letters, 2020, 56(9): 435-438.
- [72] Landrieu L, Boussaha M. Point cloud oversegmentation with graph-structured deep metric learning[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, 2019: 7440-7449.
- [73] Pan L, Chew C M, Lee G H. Pointatrousgraph: Deep hierarchical encoder-decoder with atrous convolution for point clouds[J/OL]. 2019, arXiv: 1907.09798.
- [74] Jiang L, Zhao H, Liu S, et al. Hierarchical point-edge interaction network for point cloud semantic segmentation[C]. Proceedings of the IEEE International Conference on Computer Vision. Seoul, 2019: 10433-10441.
- [75] Yi L, Zhao W, Wang H, et al. GSPN: Generative shape Proposal network for 3D instance segmentation in point cloud[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, 2019: 3942-3951.
- [76] Hou J, Dai A, NieSSner M. 3D-sis: 3D semantic instance segmentation of RGB-D scans[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach: 2019: 4416-4425.
- [77] Yang B, Wang J, Clark R, et al. Learning object bounding boxes for 3D instance segmentation on point clouds[C]. Advances in Neural Information Processing Systems. Jaipur, 2019: 6740-6749.
- [78] Engelmann F, Bokeloh M, Fathi A, et al. 3D-MPA:

- Multi-proposal aggregation for 3D semantic instance segmentation[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, 2020: 9028-9037.
- [79] Jiang L, Zhao H S, Shi S S, et al. PointGroup: Dual-set point grouping for 3D instance segmentation[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, 2020: 4866-4875.
- [80] Wang W Y, Yu R, Huang Q G, et al. SGPN: Similarity group proposal network for 3D point cloud instance segmentation[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 2569-2578.
- [81] Liu C, Furukawa Y. MASC: Multi-scale affinity with sparse convolution for 3D instance segmentation[J/OL]. 2019, arXiv: 1902.04478.
- [82] Graham B, Engelcke M, van der Maaten L. 3D semantic segmentation with submanifold sparse convolutional networks[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 9224-9232.
- [83] Lahoud J, Ghanem B, Oswald M R, et al. 3D instance segmentation via multi-task metric learning[C]. IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, 2019: 9255-9265.
- [84] Song S R, Yu F, Zeng A, et al. Semantic scene completion from a single depth image[C]. IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, 2017: 190-198.
- [85] Zhang B, Wonka P. Point cloud instance segmentation using probabilistic embeddings[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, 2021: 8879-8888.
- [86] Wang X L, Liu S, Shen X Y, et al. Associatively segmenting instances and semantics in point clouds[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, 2019: 4091-4100.
- [87] Zhao L, Tao W B. JSNet: Joint instance and semantic segmentation of 3D point clouds[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12951-12958.
- [88] Pham Q H, Nguyen T, Hua B S, et al. JSIS3D: Joint semantic-instance segmentation of 3D point clouds with multi-task pointwise networks and multi-value conditional random fields[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, 2019: 8819-8828.
- [89] Elich C, Engelmann F, Schult J, et al. 3D-BEVIS: Birds-eye-view instance segmentation[J/OL]. 2019, arXiv: 1904.02199.
- [90] Han L, Zheng T, Xu L, et al. OccuSeg: Occupancy-aware 3D instance segmentation[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, 2020: 2937-2946.
- [91] Kalogerakis E, Averkiou M, Maji S, et al. 3D shape segmentation with projective convolutional networks[C]. IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, 2017: 6630-6639.
- [92] Wang Z J, Lu F. VoxSegNet: Volumetric CNNs for semantic part segmentation of 3D shapes[J]. IEEE Transactions on Visualization and Computer Graphics, 2020, 26(9): 2919-2930.
- [93] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]. IEEE Conference on Computer Vision and Pattern Recognition. Boston, 2015: 3431-3440.
- [94] Wang P Y, Gan Y, Shui P P, et al. 3D shape segmentation via shape fully convolutional networks[J]. Computers & Graphics, 2018, 70: 128-139.
- [95] Hilaga M, Shinagawa Y, Kohmura T, et al. Topology matching for fully automatic similarity estimation of 3D shapes[C]. Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques. New York, 2001: 203-212.
- [96] Belongie S, Malik J, Puzicha J. Shape matching and object recognition using shape contexts[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(4): 509-522.
- [97] Johnson A E, Hebert M. Using spin images for efficient object recognition in cluttered 3D scenes[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1999, 21(5): 433-449.
- [98] Chen Z, Yin K, Fisher M, et al. BAE-NET: Branched autoencoder for shape co-segmentation[C]. Proceedings of the IEEE International Conference on Computer Vision. Seoul, 2019: 8490-8499.
- [99] Lyu Y C, Huang X M, Zhang Z M. Learning to segment 3D point clouds in 2D image space[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, 2020: 12252-12261.
- [100] Lei H, Akhtar N, Mian A. Octree guided CNN with spherical kernels for 3D point clouds[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, 2019: 9623-9632.
- [101] Liu Y C, Fan B, Xiang S M, et al. Relation-shape convolutional neural network for point cloud analysis[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, 2019: 8887-8896.
- [102] 陈辉, 黄晓铭, 刘万泉. 基于动态网格 $k$ 邻域搜索的激光点云精简算法[J]. 控制与决策, 2020, 35(12):

- 2986-2992.  
(Chen H, Huang X M, Liu W Q. Laser point cloud simplification algorithm based on dynamic grid  $k$ -nearest neighbors searching[J]. Control and Decision, 2020, 35(12): 2986-2992.)
- [103] Guo M H, Cai J X, Liu Z N, et al. PCT: Point cloud transformer[J]. Computational Visual Media, 2021, 7(2): 187-199.
- [104] Zhao H, Jiang L, Jia J, et al. Point transformer[J/OL]. 2020, arXiv: 2012.09164.
- [105] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]. Advances in Neural Information Processing Systems. Long Beach, 2017: 5998-6008.
- [106] Dai Z, Yang Z, Yang Y, et al. Transformer-xl: Attentive language models beyond a fixed-length context[J/OL]. 2019, arXiv: 1901.02860.
- [107] Humphrey E J, Bello J P. Rethinking automatic chord recognition with convolutional neural networks[C]. The 11th International Conference on Machine Learning and Applications. Boca Raton, 2012: 357-362.
- [108] Graham B, van der Maaten L. Submanifold sparse convolutional networks[J/OL]. 2017, arXiv: 1706.01307.
- [109] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [110] Zhang X Y, Zhou X Y, Lin M X, et al. ShuffleNet: An extremely efficient convolutional neural network for mobile devices[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 6848-6856.
- [111] 薛俊韬, 马若寒, 胡超芳. 基于 MobileNet 的多目标跟踪深度学习算法[J]. 控制与决策, 2021, 36(8): 1991-1996.  
(Xue J T, Ma R H, Hu C F. Deep learning algorithm based on MobileNet for multi-target tracking[J]. Control and Decision, 2021, 36(8): 1991-1996.)
- [112] Wang Y D, Zhang J, Kan M N, et al. Self-supervised equivariant attention mechanism for weakly supervised semantic segmentation[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, 2020: 12272-12281.
- [113] Lang I, Manor A, Avidan S. SampleNet: Differentiable point cloud sampling[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, 2020: 7575-7585.
- [114] Zhao Y F, Li J, Zhang Y, et al. Multi-class part parsing with joint boundary-semantic awareness[C]. IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, 2019: 9176-9185.
- [115] 鹿天柱, 钱晓超, 何舒, 等. 一种基于深度学习的时间序列预测方法[J]. 控制与决策, 2021, 36(3): 645-652.  
(Lu T Z, Qian X C, He S, et al. A time series prediction method based on deep learning[J]. Control and Decision, 2021, 36(3): 645-652.)
- [116] Xiong B, Jiang W Z, Li D K, et al. Voxel grid-based fast registration of terrestrial point cloud[J]. Remote Sensing, 2021, 13(10): 1905.
- [117] Shi H Y, Lin G S, Wang H, et al. SpSequenceNet: Semantic segmentation network on 4D point clouds[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, 2020: 4573-4582.
- [118] Miao J X, Wei Y C, Wu Y, et al. VSPW: A large-scale dataset for video scene parsing in the wild[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, 2021: 4131-4141.
- [119] Zhang F, Fang J, Wah B, et al. Deep FusionNet for point cloud semantic segmentation[C]. European Conference on Computer Vision. Cham, 2020: 644-663.

## 作者简介

卢健(1978—), 男, 副教授, 博士, 从事人工智能与机器视觉、智能机器人技术等研究, E-mail: lujian@xpu.edu.cn;

贾旭瑞(1992—), 女, 硕士生, 从事深度学习、三维点云分割等研究, E-mail: 18434393587@163.com;

周健(1983—), 男, 讲师, 博士, 从事机器人系统建模、环境智能感知等研究, E-mail: zhoujian0627@sohu.com;

刘薇(1982—), 女, 讲师, 博士, 从事人工智能与机器视觉等研究, E-mail: lw\_xaw@foxmail.com;

张凯兵(1975—), 男, 教授, 博士, 从事影像人工智能与机器视觉等研究, E-mail: zhangkaibing@xpu.edu.cn;

庞菲菲(1984—), 女, 讲师, 博士, 从事环境智能感知、目标定位与跟踪等研究, E-mail: ffpang@xpu.edu.cn.

(责任编辑: 齐 霖)