

基于深度学习的大规模三维点云处理综述^①



王振燕, 孙红岩, 孙晓鹏

(辽宁师范大学 计算机与信息技术学院, 大连 116029)

通信作者: 孙红岩, E-mail: 3256153627@qq.com

摘要: 随着三维视觉的快速发展, 基于深度学习的大规模三维点云实时处理成为研究热点. 以三维空间分布无序的大规模三维点云为背景, 综合分析介绍并对比深度学习实时处理三维视觉问题的最新进展, 对点云分割、形状分类、目标检测等方面算法优势与不足进行详细分析, 给出详细的性能分析与优劣对比, 并对点云常用数据集进行简要介绍, 并给出不同数据集的算法性能对比. 最后, 指出未来在基于深度学习方法处理三维点云问题上的研究方向.

关键词: 深度学习; 目标检测; 目标追踪; 形状分类; 点云分割

引用格式: 王振燕, 孙红岩, 孙晓鹏. 基于深度学习的大规模三维点云处理综述. 计算机系统应用, 2023, 32(2): 1-12. <http://www.c-s-a.org.cn/1003-3254/8743.html>

Survey on Large Scale 3D Point Cloud Processing Using Deep Learning

WANG Zhen-Yan, SUN Hong-Yan, SUN Xiao-Peng

(Department of Computer and Information Technology, Liaoning Normal University, Dalian 116029, China)

Abstract: With the rapid development of 3D vision, large-scale 3D point cloud processing in real time based on deep learning has become a research hotspot. Taking a large-scale 3D point cloud with disordered spatial distribution as the background, this study comprehensively analyzes, introduces and compares the latest progress of deep learning in real-time processing of 3D vision problems. Then, it analyzes in detail and compares the advantages and disadvantages of algorithms in terms of point cloud segmentation, shape classification and target detection. Further, it briefly introduces the common data sets of point clouds and compares the algorithm performance of different data sets. Finally, the study points out the future research direction of 3D point cloud processing based on deep learning.

Key words: deep learning (DL); target detection; target tracking; shape classification; point cloud segmentation

深度学习 (deep learning, DL) 指基于数据预测、并改进其预测结果或行为的方法^[1], 训练阶段以最小化损失函数为引导, 通过梯度下降调整计算模型的权重和偏置; 测试阶段以输入数据和训练好的模型参数计算预测值^[1], 广泛应用于二维目标检测、分割和分类^[2,3]等领域.

随着三维扫描仪、激光雷达、RGB-D 相机 (如 Kinect、RealSense 和 Apple 深度相机) 等三维采集设备性能的提高, 获取具有丰富几何形状和比例特征的三维数据成为现实^[4]. 三维点云数据通常包含几何位置信息、颜色信息和强度信息. 其中, 颜色信息通过相机

获取彩色影像, 再将其颜色信息赋予点云中对应的点; 强度信息通过激光扫描仪接收装置采集到的回波强度.

基于深度学习的三维点云数据处理目前面临以下 3 个方面的挑战: 点云无序性及非结构性, 点云旋转不变性, 以及点云特征的有效提取.

本文主要贡献如下: 概括总结点云分割、形状分类、目标检测等方面的深度学习算法; 给出常用的评价指标及相应数据集; 探索基于深度学习的三维点云处理未来研究方向.

本文结构如下: 第 1 节整理深度学习相关数据集

① 基金项目: 国家自然科学基金 (61472170)

收稿时间: 2022-01-15; 修改时间: 2022-02-17; 采用时间: 2022-03-03; csa 在线出版时间: 2022-11-14

CNKI 网络首发时间: 2022-11-15

及其评价指标;第2节回顾三维点云分割算法;第3节和第4节分别介绍三维点云的形状分类和目标检测;第5节总结全文并讨论发展趋势。

1 数据集和评价指标

本节介绍深度学习常用数据集和性能评估指标.深度学习广泛使用的大规模三维点云数据集包括 ModelNet^[5]、ScanNet^[6]、S3DIS^[7]、SemanticKITTI^[8]等。

1.1 数据集

规模大、标记精度高的数据集不仅有助于提高性能评估的精度,且有利于验证对比算法优劣。

点云分割算法常用的数据集包括:ScanNet^[6]、S3DIS^[7]、SemanticKITTI^[8].其中 ScanNet 和 S3DIS 均为室内场景:ScanNet 数据集共 1 513 个场景数据、21 个对象类别,其中 1 201 个场景用于训练,312 个场景用于测试;S3DIS 数据集由 3 个不同建筑的 5 个大型室内场景组成.SemanticKITTI 为室外场景,包含市区、乡村和高速公路等场景的真实图像数据。

形状分类的常用数据集有 ModelNet^[5],包含 127 915 个三维形状、662 类,其子集 ModelNet10 包含 10 类 4 899 个三维形状,ModelNet40 包含 40 类 12 311 个三维形状。

检测算法常用数据集有:SemanticKITTI^[8]、Oxford^[9]、NYU^[10].Oxford 数据集由 21 711 个训练子地图和 3 030 个测试子地图组成.NYU 数据集包含超过 72k 个训练帧和 8k 个测试帧,每帧包含 36 个带注释的关节。

表 1 总结上述用于点云分割、形状分类和检测算法常用数据集的对象类别,同时提供了使用各个数据集的算法以及相关性能对比。

表 1 不同数据集算法性能对比

数据集	类别	方法	评价指标 (%)		
			OA	mAcc	mIoU
ScanNet	21	SPCR ^[11]	—	75.70	—
		DyCo3D ^[12]	—	58.90	—
SemanticKITTI	8	Im-RangeNet ^[13]	—	—	43.60
S3DIS	13	PCT ^[14]	—	67.65	61.33
		DGCSA ^[15]	—	—	50.10
ModelNet40	40	PCT ^[14]	93.20	—	—
		DGCSA ^[15]	93.40	90.60	—

1.2 评价指标

常用的评价指标包括 $mAcc$ 、 $mIoU$ 和 OA .其中,

$mAcc$ 是正类预测个数与所有预测样本数之比的均值,计算公式如下^[16]:

$$mAcc = \frac{1}{N} \sum_{i=1}^N \frac{TP_i + TN_i}{TP_i + FP_i + FN_i + TN_i} \quad (1)$$

其中, TP_i 表示真预测为真, TN_i 表示真预测为假, FP_i 表示假预测为真, FN_i 表示假预测为假, N 表示类别数。

$mIoU$ 评估预测框与人工标记框的重合程度,定义为^[16]:

$$mIoU = \frac{1}{N} \sum_{i=1}^N \frac{TP_i}{(FN_i + FP_i + TP_i)} \quad (2)$$

OA 是在预测为真的样例中、被正确预测为真的比例,计算公式如下^[16]:

$$OA = \frac{\sum_{i=1}^N TP_i}{\sum_{i=1}^N (TP_i + FP_i)} \quad (3)$$

召回率 (recall, R) 是所有正样本样例中,被正确预测为真的比例,定义为^[16]:

$$R = \frac{\sum_{i=1}^N TP_i}{\sum_{i=1}^N (TP_i + FN_i)} \quad (4)$$

精度-召回率 (precision-recall, PR) 曲线的横坐标是召回率,纵坐标是 OA .若在 R 增长的同时, OA 在高水平范围变化不大,则说明分类器的性能较好。

2 点云分割

点云分割指将整个点云聚类为多部分点云,每部分点云对应独立的物体对象,分为部件级分割、场景级分割和对象级分割。

2.1 部件级分割

2018 年, Li 等提出无序点云的置换不变结构自组织网络 (self-organizing network, SO-Net)^[17],以自组织映射 (self-organizing map, SOM) 模拟点云的空间分布,并以单个特征向量表示输入点云.在特征聚合过程中,基于 SOM 执行点到点的 K 近邻 (K-nearest neighbor, KNN) 搜索以控制感受野重叠。

2019 年, Wang 等提出 EdgeConv, 在保持排列不变性的同时,捕获局部几何结构^[18],对输入点云 EdgeConv 层的每个点,计算大小为 k 的边缘特征集,并聚合每个集合内的特征.基于最后一层 EdgeConv 构造 1D 全局描述符,并与每个点的所有 EdgeConv 输出 (用作局部描述符) 串联,输出 p 个语义标签. Wu 等扩展动态滤波器到新的卷积操作 PointConv,可用于构建点云的深度

卷积网络^[19],以层次结构聚集点云特征,并组合细小区域特征为覆盖大空间范围的抽象特征.该层次结构由特征编码模块组成,包括采样层、分组层和 PointConv.

针对体素分辨率增加而导致计算资源耗尽的问题,Wang 等在 2020 年提出新的体卷积神经网络,在有限分辨率下,提取体素化三维数据编码细节信息的鉴别特征^[20].针对体素化形状,提取空间密集 (spatial dense extraction, SDE) 模块的多尺度特征,保留细节信息;串联聚集模块不同阶段的特征;基于 3 个 $1 \times 1 \times 1$ 卷积层预测每个体素的部件标签.

2021 年,Chen 等提出基于图注意力的点神经网络 (graph attention based point neural network, GAPNet),在多层感知器嵌入图注意力机制以学习局部几何表示^[21].该网络引入 GAPLayer 学习点注意力特征,基于多分支机制充分利用特征,设计邻域注意力池化层提取局部特征.

在部件级分割算法中,文献 [17] 以点云自动编码器作为预训练,以提高网络的分割性能;文献 [18] 在层与层间动态更新关系图、实现点的语义分组;文献 [19] 提出的方法高效且结构可扩展;文献 [20] 结合 SDE 模块和注意力特征聚集模块实现分割;文献 [21] 将 GAP-Layer 和注意力池化层集成到多层感知器结构提取局部上下文特征.

2.2 场景级分割

2017 年,Charles 等首次提出直接处理点云的深度学习框架^[22].该方法主要解决了点云存在的两大问题:一是基于空间变换网络解决点云旋转性问题,应用于输入点云和特征变换部分,二是基于最大池化层解决点云无序性问题,最大池化层对点云的整体提取出全局特征.但其无法捕获局部特征,为解决此问题,Qi 等基于层次网络改进 PointNet 架构,从而提高特征提取能力^[23].由多个集合抽象层组成,包括采样层、分组层和 PointNet 层.采样层从输入点选择一组点,定义局部区域的质心.分组层通过查找质心周围的“相邻”点以构造局部区域集.PointNet 层将局部区域模式编码为特征向量.

2019 年,Zhao 等从点云局部邻域中提取上下文特征,通过自适应特征调整 (adaptive feature adjustment, AFA) 模块提取局部点特征^[24],基于局部区域的点特征构建全连接点网络,通过学习确定相互影响并构造区域集成信息,并融合全局特征和局部特征用于识别.Hu

等提出高效轻量级神经网络 RandLA-Net,基于局部空间编码、注意力池化层、以及扩张残差块构造局部特征聚合模块^[25];该方法基于相邻几何结构和显著增加感受野保留局部复杂结构,利用前馈 MLP 提高计算效率.

2020 年,Li 等提出新的图卷积结构泰勒-高斯混合模型网络 (Taylor Gaussian mixture model network, TGNet),基于多尺度分层提取点特征,保证粗尺度和细尺度信息结合^[26].基于 TGConv 构造语义深层特征,并以共享多层感知器 (multilayer perceptron, MLP) 提取每点特征.在输出层内,基于条件随机场 (conditional random field, CRF) 改善分割结果.

2021 年,Guo 等受 Transformer 的启发,基于传统 Transformer 提出新的点云学习框架点云 Transformer (point cloud Transformer, PCT)^[14].该方法合并原始的位置编码和输入嵌入,基于输入嵌入模块产生可区分特征.为解决自注意力机制忽略局部几何信息的问题,PCT 利用邻近嵌入策略改进点嵌入.

在场景级分割中,文献 [22] 提出适用于处理三维无序点集的深度神经网络;文献 [23] 提出 2 种新的集合抽象层解决非均匀点采样问题;文献 [24] 基于 K 近邻分组层提出 AFA 模块,在局部区域构建全连接网络,在点间传输信息;文献 [25] 采用随机采样方法降低点密度减少计算量;文献 [26] 基于卷积滤波器和系列高斯加权泰勒核函数表示的局部坐标,提取局部相关性;文献 [14] 利用 Transformer 固有的顺序不变性,避免需要定义点云数据的顺序.

2.3 对象级分割

2018 年,Wang 等提出基于深度学习的点云三维对象实例分割和相似群候选网络 (similarity group proposal network, SGPN)^[27],引入相似性矩阵以便从单个网络中生成具有类预测的群方案.首先利用 PointNet/PointNet++ 获取点云的特征,然后在此基础上计算相似性矩阵、置信度矩阵和语义分割矩阵,最后将所有结果整合可得到实例分割结果.

2019 年,Wang 等提出点云实例分割与语义分割相结合的分割框架 (associatively segmenting instances and semantics, ASIS)^[28],学习语义感知的点级实例嵌入,基于语义分割增强实例分割,并融合同一实例点的语义特征,提高点语义预测精度.Pham 等基于多任务逐点网络和多值条件随机场联合执行语义实例分割^[29],通过重叠窗口扫描输入的点云,从窗口中提取三维点,

通过多任务点态网络得到语义标签和实例嵌入,优化多值条件随机场模型,产生最终分割结果. Yang 等提出基于包围盒回归的高效分割算法(3D-BoNet),利用最小化匹配代价函数获得大致的包围盒回归,通过点掩码预测实现最终的实例分割^[30]. 基于实例级包围盒预测分支构造每个实例的包围盒;基于点级掩码预测分支为包围盒内的点预测掩码,用于判断包围盒内的点是实例还是背景. 为更好地训练该网络,提出包围盒关联层和多指标损失函数.

2021年, Zhang 等首次提出半监督点云实例分割体系结构,即扰动一致性正则化半监督点云实例分割(segmentation with perturbation consistency regularization, SPCR)^[11]. 该方法在输入端引入各种扰动方案,使网络具有鲁棒性,且易于推广到不可见和未标记的数据. He 等结合动态卷积提出三维点云实例分割方法,该方法包括稀疏卷积主干、权值生成器和实例解码器^[12]. 设计小型轻量 Transformer 捕获点之间的依赖关系和高级交互.

在对象级分割算法中,文献[27]内存使用效率较高,但不能处理多于 10^5 个点的超大场景;文献[28]提出语义感知的实例分割、实例融合的语义分割,并相互协作;文献[29]可用于语义分割和实例分割;文献[30]不需要任何后处理步骤,计算效率较高;文献[11]基于共享加权网络实现半监督框架,并设计扰动一致性正则化机制,为从未标记数据进行网络学习提供自我监督;文献[12]提出比自底向上策略更健壮的非候选框分割方法.

2.4 算法分析与比较

本节分析与比较上述典型的部件级分割算法、场景级分割算法、以及对象级分割算法.

分割算法的分析与比较如表2所示,比较内容包括类别、方法、年份、贡献. 在部件级分割算法中, VoxSegNet 从体素化三维编码信息提取特征,并基于体素化形状提取多尺度特征,但会导致部分信息丢失;而 SO-Net、DGCNN、PointConv、GAPNet 重点解决直接处理点云数据,用于探索局部上下文信息和注意力机制的学习架构,提高了算法精度,但速度较慢. 在场景级分割算法中, PointNet、PointNet++、RandLA-Net 以共享 MLP 为网络基本单元,利用累积跳跃连接增强语义信息,效率相对较高; TGNet 解决了由输入与邻近坐标或特征间的高几何相关性导致分割性能欠佳的问题,并克服了传统聚集函数丢弃或忽略局部邻域间的结构联系;而 PCT 借助 Transformer 的优势,并结合注意力机制学习特征,具有良好的语义特征学习能

力,在计算效率、内存占用等方面远优于上述算法. 在对象级分割算法中,SGPN、ASIS、JSIS3D、DyCo3D 将实例分割看作语义分割的后续聚类步骤,为每个点学习鉴别嵌入特征,再通过均值偏移等聚类算法将同一实例点聚集到一起,但其目标性较差,且后处理步骤的时间成本较高;而 SRCR 主要研究将实例分割问题转化为三维目标检测和实例掩码预测 2 个子任务,这通常依赖两阶段的训练和昂贵的非极大值抑制等操作对密集候选框进行选择; 3D-BoNet 主要研究单阶段、无锚且端到端的包围盒回归算法,通过显示预测目标包围盒,得到的实例有更好的目标性,且不需要复杂耗时的区域候选框网络和后处理步骤.

分割算法的性能比较如表3所示,比较内容包括类别、方法以及3种评价指标. 通过对比,在直接或间接处理大规模点云数据方面,算法的分割精度差异不大,甚至直接处理大规模点云数据(RandLA-Net)的总体精度仅比间接处理(PointNet)低0.11%,这是因为随机采样降低内存占用和计算成本,且局部特征聚合模块保留有用特征; PointNet 相较于其他分割方法在精度方面不占优势,主要原因在于 PointNet 丢失部分点之间的局部特征,但该方面的研究仍在不断探索以提高精度并且其架构一直启发着后续的研究; TGNet 的总体精度高于其他分割算法,主要因为分层多尺度方法和融合多尺度信息的优势提高了分割性能.

3 形状分类

点云分类是为每个点或组分配语义标签的过程. 本节概括分析介绍基于卷积的和基于图分类的算法,并分析讨论其优势不足.

3.1 基于卷积的方法

传统三维卷积神经网络计算量大、内存占用高. 针对以上问题,2018年, Hermsilla 等提出高效的非均匀采样点云卷积学习方法^[31],包括4个关键技术:利用 MLP 表示卷积核;考虑样本密度函数;使用泊松磁盘抽样作为分层点云学习的可伸缩方法;采用蒙特卡罗(Monte Carlo, MC)积分描述卷积. 2019年, Thomas 等提出核点卷积(kernel point convolution, KPConv)以计算三维空间中点的特征^[32],这是对点云卷积的新设计,且不需利用点云的任何中间表达形式. 该方法设计两种不同类型的核:一种是刚性核,先选择通过求解优化问题放置核,且每个点对其他点施加斥力,这些点均被约束在具有吸引力的球面上,其中一点被约束在中心,

最终周围的点被约束在平均半径为核影响距离的1.5倍内以确保每个核之间的影响力有小幅度的重叠; 一种是可变核, 借鉴可变网络的设计, 在刚性基础上利用网络学习局部位移, 使其适用于点云。

表2 分割算法对比

类别	方法	年份	贡献	弊端
部件级分割	PointConv ^[19]	2019	提出新权函数计算公式, 在局部坐标上训练MLP, 以逼近卷积滤波器的连续权函数和密度函数, 构造排列不变性和平移不变性	PointConv无法将更主流的图像卷积网络架构应用到点云数据中, 如ResNet、DenseNet等
	VoxSegNet ^[20]	2020	引入SDE模块缓解子采样过程和数据稀疏性造成的细节丢失, 基于注意力机制融合SDE多尺度信息	分辨率越大, 提高性能上限越困难, 最终损失部分特征, 无法更好地描述细节结构
	GAPNet ^[21]	2021	提出自我注意力和邻域注意力机制, 基于邻域注意力池化层识别重要特征以增强网络的鲁棒性	由于GAPLayer中多分支机制的使用, 导致计算效率较低
场景级分割	PointNet ^[22]	2017	利用对称函数解决点云顺序问题, 引入T-Net解决点云旋转问题	只能捕获全局特征, 从而忽略局部特征
	RandLA-Net ^[25]	2020	基于随机抽样降低内存占用和计算成本, 引入局部特征聚合模块以保留有用特征	RandLA-Net只实现场景级分割, 无法扩展到大规模点云上实现端到端对象级分割
	PCT ^[14]	2021	提出的PCT适合于非结构化、无序且不规则的点云数据, 并设计带有隐式拉普拉斯算子和归一化修正的偏移注意力, 与自注意力相比, 更适用于点云学习	该网络处理大型点云的效果有待补充
对象级分割	ASIS ^[28]	2019	基于baseline网络构造逐点语义预测和输出点级实例嵌入, 属于同一实例的点的嵌入较为接近, 分属不同实例的点的嵌入较为分离	后处理步骤的时间成本较高
	3D-BoNet ^[30]	2019	设计新颖的包围盒关联层, 并利用多指标损失函数以监督包围盒预测分支	该方法在分割后的小块上进行测试, 并未直接在大规模点云上训练和测试
	SRCR ^[11]	2021	提出两个包含几何信息和语义信息的一致性项以更好地预测不同扰动下的不变约束	该方法依赖两阶段训练和昂贵的非极大值抑制等操作对候选框进行选择
	DyCo3D ^[12]	2021	提出数据驱动的方法产生卷积核以应用于实例, 设计轻量级的转换器扩大感受野、捕获非局部依赖性	该网络后处理操作的计算时间较高

表3 不同分割算法在 S3DIS 数据集上的性能对比 (%)

类别	方法	区域5			6折		
		OA	mIoU	mAcc	OA	mIoU	mAcc
部件级分割	DGCNN ^[18]	—	—	—	84.10	56.10	—
	PointNet ^[22]	—	—	—	78.62	47.71	—
	PointWeb ^[24]	86.97	60.28	66.64	87.31	66.73	76.19
场景级分割	TGNet ^[26]	88.50	57.80	—	—	—	—
	RandLA-Net ^[25]	—	—	—	87.20	68.50	81.50
	PCT ^[14]	—	61.33	67.65	—	—	—

2020年, Wang等将规则网格的深度学习推广到不规则三维点云, 并提出新的结构感知卷积 (structure-aware convolution, SAC)^[33], 利用三维核匹配点云邻域, 其中核可看作可学习的三维点构成的“几何模板”。首先, 根据点的空间距离找到其相邻点集; 然后训练得到三维核并匹配; 如果可以完全匹配, 则激活对应的通道。2021年, Tang等提出融合互补卷积神经网络的新方法, 称为多分支卷积神经网络 (multi-branch convolutional neural networks, MBCNN), 在保持高效率的同时实现当前最先进的性能^[34]。该网络引入基于体素的索引技术以加快查找邻域点的过程, 并设计特征融合策略以

确保模型具有高精度和低延迟。

在基于卷积的方法中, 文献[31]实现非均匀采样; 文献[32]通过半径邻域搜索解决了非均匀条件下KNN不具备鲁棒性的问题, 且提出的算法使得GPU内存占用较少、速度较快, 并可在不定个数的点云上直接训练; 文献[33]适用于现有的点云深度学习网络; 文献[34]利用逐点MLP的鲁棒性, 使模型在真实数据集上具有良好的性能。

3.2 基于图的方法

2018年, Zhang等针对三维点云数据分类, 融合局部图卷积和2种类型的图下采样(池化), 提出图卷积

神经网络 (graph convolutional neural networks, Graph-CNN), 称为 PointGCN^[35]. 该网络结合了卷积层、池化层和完全连接层, 且卷积层与池化层是特定于图设置的. 该方法实现两种形式的池化操作聚合信息: 一用于计算所有输出点的全局统计信息, 二作用于点云的局部, 导致多分辨率池化. 为避免计算图粗化或聚类, 通过对彼此最分散的一组点进行子采样实现多分辨率池化.

2019年, Wang等提出边缘卷积 EdgeConv^[18], 可嵌入到多个已有的神经网络框架中. 该方法先提取中心点与邻近点间的边缘特征, 然后在边缘特征上完成卷积操作, 即将中心点的特征与两点的特征差串联后再输入到 MLP 中, 因此边缘特征融合了点间的局部信息与点的全局信息. 在逐层边缘卷积时, 先在特征空间中构造图, 然后在网络的每一层之后动态更新.

2021年, 宋巍等结合动态图卷积和空间注意力提出点云分类分割方法 (dynamic graph convolution spatial attention neural networks, DGCSA)^[15]. 基于动态图卷积

构造 K 近邻图并提取边特征; 基于空间注意力模块学习局部特征. 受 GAPNet 启发, Lu 等结合通道注意力模块与动态图卷积神经网络, 实现网络模型构建^[36]. 该网络基于通道注意力模块优化动态图结构, 基于通道注意力机制产生节点间的通道注意力映射, 利用空间注意力焦点补偿动态图中的通道注意力.

在基于图的方法中, 文献 [35] 通过构建图中编码的局部结构实现三维点云数据分类; 文献 [18] 应用边缘卷积实现分类, 但并未学习每个中心点对的信息; 文献 [15] 提出的空间注意力模块应用图卷积实现分类性能的提高; 文献 [36] 考虑通道注意力和空间注意力的互补性, 并基于图结构的适用特征提取点云的局部邻域信息.

3.3 算法分析与比较

本节分析与对比典型的基于卷积、基于图的分类算法. 关于形状分类算法的研究工作如表 4 所示, 比较内容包括类别、方法、年份、评价指标、贡献.

表 4 分类算法对比

类别	方法	年份	评价指标	贡献	弊端
基于卷积	Hermosilla等 ^[31]	2018	OA	基于蒙特卡罗卷积, 并引入泊松磁盘抽样构造点层次结构, 实现非均匀点云分类	该网络必须依赖核密度估计获得概率密度函数
	KPConv ^[32]	2019	OA	在每个卷积位置产生不同的偏移, 这意味着可以根据输入点云的不同区域调整其内核形状, 且内核点数量不受限制, 较为灵活	可变形卷积适用于更大型数据集或如目标检测、激光雷达流量计算等任务中有待解决
	SAC ^[33]	2020	OA	提出的结构感知卷积方法可集成于先进的点云深度学习网络	SAC中的核与输入点云直接匹配, 因此具有任意方向的几何结构很难用有限核表示
	MBCNN ^[34]	2021	OA	提出基于体素的索引以加快邻域搜索时间, 设计基于点的球CNN直接处理点云, 并利用融合策略获取全面特征	该方法能否适用于更大型数据集的处理任务中有待验证
基于图	PointGCN ^[35]	2018	OA, mAcc	基于图卷积运算以及2种专门设计的点云数据池化层, 在不同的感受野学习总结每个点云的潜在特征	最远点采样的使用导致内存占用较高、计算效率较低
	DGCNN ^[18]	2019	OA, mAcc	提出新的神经网络模块EdgeConv, 可微, 并能嵌入已有的网络架构	EdgeConv未考虑相邻点间的向量方向, 导致损失部分局部几何信息
	DGCSA ^[15]	2021	OA, mAcc	设计空间注意力模块解决局部邻域聚合过程中产生的信息丢失问题, 利用注意力机制学习局部特征, 提高分类精度	分割任务实验结果中出现部分类别分割精度较低的现象
	Lu等 ^[36]	2021	OA, mAcc	提出将通道注意力模块嵌入动态图卷积神经网络, 以增强节点之间的表示, 基于密集连接以提高各层利用率	该方法无法应用于实时环境检测和大型真实场景点云的分类测试中

在基于卷积的方法中, Hermosilla等^[31]、KPConv、SAC 主要研究在连续空间定义卷积核, Hermosilla等^[31]将卷积当做蒙特卡罗估计的过程, 并使用泊松磁盘抽样构建点云, KPConv 利用相关性函数计算出每个矩阵的系数, 在此基础上, 核函数对所有矩阵加权求和得到权重矩阵实现特征的更新, SAC 将卷积核看作由三维

点构成的几何模板; MBCNN 是在规则网格上定义卷积核, 利用共享参数的过滤器 (kernel), 通过计算中心像素点与相邻像素点的加权和以构成特征映射实现特征的提取. 在基于图的方法中, PointGCN 将卷积定义为频谱滤波, 借助图谱的理论实现拓扑图上的卷积操作, 在训练阶段, 将输入信号转换为潜在特征映射的图

滤波器系数以区分对象类,利用重新缩放的拉普拉斯算子以提高网络的稳定性和性能;而 DGCNN、DGCSA、Lu 等^[36] 直接在每个结点的连接关系上定义卷积操作,类似于传统的卷积神经网络中的卷积,可直接在图像的像素点上进行卷积, DGCSA 与 Lu 等^[36] 均结合 DGCNN 的优势进行改进,在计算效率与精度方面均优于 DGCNN。

形状分类算法的性能比较如表 5 所示,比较内容包括类别、方法、总体精度、均值精度。通过对比, Lu 等^[36] 的总体精度和均值精度都明显高于其他分类算法,主要原因在于利用 DGCNN 的优势并结合密集连接提高网络性能; KPConv 可根据输入点云的不同区域调整其内核形状,且内核点数量不受限制,在总体精度方面优于其他基于卷积的方法和部分基于图的方法。综合分析,基于卷积的网络可以在非均匀三维点云上实现优异的性能。

表 5 不同分类算法性能对比 (%)

类别	方法	ModelNet10		ModelNet40	
		OA	mAcc	OA	mAcc
基于卷积	Hermosilla等 ^[31]	—	—	均匀: 90.90	—
		—	—	非均匀: 90.10	—
	KPConv (rigid) ^[32]	—	—	92.90	—
	KPConv (deform) ^[32]	—	—	92.70	—
	MBCNN ^[34]	—	—	91.90	—
基于图	PointGCN ^[35]	91.91	91.57	89.51	86.05
	DGCNN ^[18]	—	—	92.90	90.20
	DGCSA ^[15]	—	—	90.40	90.60
	Lu等 ^[36]	—	—	93.47	90.83

4 目标检测

目标检测指找出指定区域中所有感兴趣的物体,包含物体定位和物体分类 2 个子任务。依据数据输入方式,分为基于多视图、基于体素、基于点的方法。

4.1 基于多视图的方法

基于多视图的方法将三维对象投影到多个视图中提取相应视图特征,再将其融合。

2016 年, Chen 等选取点云候选三维包围盒,并投影到二维图像中完成检测^[37],避免了多尺度搜索,并基于非最大值抑制构造最终的对象候选框集合。基于三维包围盒与二维包围盒的映射关系,利用二维空间中的特征描述三维包围盒。

2018 年, Qi 等提出 2D-3D 的目标检测方法^[38],基

于 CNN 在 RGB 图像中生成二维目标区域,并拉伸为三维视锥,从深度数据中得到点云,为保证预测的对象中心坐标不会发生偏移, Frumstum PointNet 预测三维包围盒确定对象真实中心。在三维包围盒任务上,该模型对距离合理且未遮挡的物体输出的准确率表现优异。Liang 等提出端到端的可学习三维物体检测器,基于连续卷积、以不同分辨率融合图像和激光雷达特征图,实现精确的三维定位^[39]。Ku 等提出聚集视图对象检测 (aggregate view object detection, AVOD) 网络,实现了无人驾驶场景下三维物体的实时检测^[40]。基于鸟瞰视图 (bird's eye view, BEV) 映射和 RGB 图像提取特征并融合,经过区域候选框网络得到无向区域候选框,将候选框对应的类别和第 1 阶段回归基于第 2 阶段探测器网络得到的三维包围盒产生精确有向的三维包围盒,实现物体检测和识别。该体系结构可以实时运行,且具有较低的内存开销。

2019 年, Liang 等提出多任务多传感器检测模型,用于二维和三维目标检测、地面估计和深度补全^[41]。该方法基于逐点和感兴趣区域 (region of interesting, ROI) 特征融合的多传感器体系结构,利用深度补全,实现密集点特征融合。同时,多任务学习提供额外的几何线索,实现更好的表示学习和更密集的特征融合。

2020 年, Yan 等提出深度学习网络,该网络结合二维数据的纹理信息和三维数据的几何信息实现目标检测^[42]。基于 Hough 投票技术区分前景点和背景点,引入反向映射层结合 RGB 纹理特征和点云几何特征,并基于特征编码网络学习点云特征。

在基于多视图的方法中,文献 [38] 提出基于上下文模型和语义学习、生成具有高召回率的特定类三维对象候选框方法;文献 [39] 提出的目标检测器在 BEV 中推理,并通过学习将图像特征投影到 BEV 空间;文献 [40] 利用显式方向向量回归解决从包围盒推断出的模糊方向估计;文献 [41] 的模型是简化的 2 阶段检测器,第 1 阶段是单次激发检测器,输出少量高质量三维检测,第 2 阶段应用 ROI 特征融合,实现更精确的二维和三维包围盒回归;文献 [42] 可以在复杂场景中保持准确性,同时提高操作效率。

4.2 基于体素的方法

基于体素的方法实质是点云体素化后利用二维或三维卷积实现目标检测。

2018 年, Zhou 等提出 VoxelNet,直接处理基于点

云的三维目标检测^[43]。通过设计体素特征编码 (voxel feature encoding, VFE) 层, 结合逐点特征与局部聚集特征, 实现体素内部的点间交互。2020年, Cui 等提出两阶段 PVF-Net 框架, 用于在自动驾驶场景中定位目标的三维边界框^[44]。第1阶段, 利用三维稀疏 CNN 和 PointNet 获取基于体素和基于点的候选框, 基于双重候选框融合模块集成两种候选框。第2阶段, 基于 RoI 深度融合模块提取并聚合基于体素的特征和逐点特征, 以丰富特征的有效性和多样性。

2019年, Chen 等提出了通用、快速的2阶段三维目标检测框架^[45]。该方法首先体素化点云, 并将其送入体素区域候选框网络 (region proposal network, RPN) 生成少量初始预测; 再通过融合坐标和来自体素网络的上下文特征, 生成预测的框特征; 最后将其送入 Refine-Net 完成细化。然而, 由于体素中姿态细节的信息丢失, 检测精度受到严重限制。为解决此问题, 2021年, Li 等提出点到体素特征学习方法, 通过点语义和局部空间特征对点云进行体素化^[46]。该方法保留了用于构建高召回率视图 RPN 的体素特征, 并提供了用于保存对象精确姿势信息的点特征。

2021年, Li 等提出改进的点体素区域卷积神经网络 (improved point-voxel region convolutional neural network, IPV-RCNN), 用于三维目标检测^[47]。该方法设计改进的基于数据增强的训练方法以提高三维目标的训练效率和检测精度, 并采用基于锚的方法预测三维边界框和类。

在基于体素的方法中, 文献 [43] 克服了手工提取特征方法的瓶颈, 提出基于点云三维检测端到端可训练的深度体系结构体素网; 文献 [44] 提出独特的两阶段三维目标检测框架; 文献 [45] 提出的目标检测框架同时利用点云的体素表示和原始密集输入; 文献 [46] 基于关注角聚合模块学习透视不变的特征, 提高重新匹配阶段的检测精度; 文献 [47] 设计了改进的三维目标检测算法, 提出的数据增强训练策略用于结合三维稀疏卷积的上采样与特征融合。

4.3 基于点的方法

基于点的方法是以点云作为输入数据, 解决了体素化带来的计算量大和内存占用高的问题。

2019年, Shi 等提出 PointRCNN, 基于三维包围盒获得分割标签, 以自下而上的方式从原始点云中直接生成三维候选框^[48]。该网络包括自下而上的三维提案生成阶段和规范包围盒细化阶段。前者避免在三维空

间中使用大量预定义的三维边界框, 并显著限制搜索空间。后者将每个提案的集合点转换为标准坐标, 以更好地学习局部空间特征, 并结合与全局语义特征, 实现精确的边界框细化和置信度预测。

2020年, Sun 等提出双注意力图卷积网络 (dual attention and graph convolution, DAGC)^[49]。通过引入双重注意力模块区分与任务相关的特征。同时, 引入残差图卷积网络 (residual graph convolution network, ResGCN) 模块聚合点的多级相邻点的局部特征, 提高效率, 利用动态图的相邻特征生成可靠的全局描述符。

2021年, Feng 等提出三维关系模块, 在三维边界框候选对象之间建立三维目标关系图, 用于特征增强, 以实现精确的目标检测^[50]。基于点注意力池化方法, 为三维目标方案提取统一的外观特征。结合外观特征与位置特征, 建立三维目标关系图。Shi 等受 PointRCNN 的启发, 提出基于点云的三维目标检测框架, 即部分感知和聚合神经网络, 称为 Part-A² 网络^[51]。该网络分为两阶段: 第1阶段学习目标内位置标签和前景标签以估计准确的目标内部件位置; 第2阶段捕获目标部件的几何信息, 以准确评分并优化其位置。

在基于点的方法中, 文献 [48] 分割点云为前景对象和背景, 生成少量高质量的三维提案; 文献 [49] 充分利用点关系和特征重要性提高模型的特征提取能力; 文献 [50] 提出的点注意力池化方法同时建模内部点的语义、空间和方向关系, 从三维候选框中获得信息; 文献 [51] 引入可微感兴趣区域的点云区域池化操作, 消除现有有点云区域池化操作中的模糊性。

4.4 算法分析与对比

本节对典型的基于多视图、基于体素、基于点的目标检测算法进行分析与比较。

关于目标检测算法研究工作的分析与比较结果如表6所示, 比较内容包括类别、方法、年份、贡献。在基于多视图的方法中, MV3D 和 Frustum PointNet 主要解决 2D-3D 目标检测问题, 但均非常依赖于从第1步生成的二维检测结果, 其中, MV3D 在小目标物体检测上表现不佳, 为解决此问题, AVOD 在特征提取阶段融合浅层和深层特征, 产生高分辨率特征映射, 并缩小参数维度; Liang 等^[39] 和 MMS 重点学习跨模态特征表示, 实现不同模式 (图像与激光雷达点云) 的信息融合; 为解决由摄像头传感器受光线影响丢失深度信息的问题, Yan 等^[42] 设计反向映射层和聚集层将二维纹理信息与三维几何信息结合, 以获取更多有用信息。在基于

体素的方法中, VoxelNet 提出单阶段检测法, 虽然可以学习具有各种几何形状的物体的有效区分表示, 并取得较好结果, 但它需要权衡准确率与处理速度; 为解决体素化造成的部分信息丢失的问题, PVF-Net 充分利用体素密集结构 (voxel dense structure, VDS) 和点稀疏结构 (point sparse structure, PSS) 的优势, 并在两种结构之间进行折衷; 针对 RCNN 出现的训练时间慢、步

骤复杂、内存占用高等问题, FP R-CNN、P2V-RCNN、IPV-RCNN 均提出改进两阶段目标检测算法. 在基于点的方法中, 由于点云池化策略的模糊性 (不同的对象可能最终汇集到同一组点), PointRCNN 会丢失对几何信息编码的能力; 为解决此问题, Part-A² 提出可区分感兴趣区域的点云池化操作以消除模糊性; DAGC 和 Feng 等^[50] 均提出两个新的模块以提高类别精度.

表6 目标检测算法对比

类别	方法	年份	贡献	弊端
基于多视图	MV3D ^[37]	2016	利用分割、上下文信息、大小和先验位置执行准确的三维目标检测, 并提出最小化能量的方法	该方法对三维小障碍物检测的应用较困难
	Frustum PointNet ^[38]	2018	利用成熟的二维目标检测器和先进的三维深度学习进行目标定位, 即使在强遮挡或点非常稀疏的情况下也能精确估计三维包围盒	该方法在稀疏点云中尺寸预测不准确, 且由于光照、遮挡等影响, 会导致二维探测器错过物体
	Yan等 ^[42]	2020	提出结合二维图像数据的纹理特征和三维数据的几何特征的深度学习网络结构, 并基于反向映射层映射二维数据到三维空间	网络结构较为复杂, 且计算效率和复杂场景中的检测精度不高
基于体素	VoxelNet ^[43]	2018	将原始点云作为输入, 并将其转换为体素, 将体素内的点转换为表征形状信息的向量	准确率和处理速度之间没有得到较好的权衡
	PVF-Net ^[44]	2020	结合基于体素和逐点数据结构的优点设计双重候选框融合模块, 集成基于体素和基于点的候选框	该方法是否可以检测到行人和自行车等类别有待证明
	FP R-CNN ^[45]	2019	第1阶段以体素作为输入, 应用卷积运算获得少量初始预测, 第2阶段基于原始点云和提取的卷积特征对其进行进一步细化	使用选择性搜索提取区域候选框导致目标检测时间较慢
	IPV-RCNN ^[47]	2021	基于全新主干获取细粒度特征, 并采用K-means聚类计算目标的大小和数量	无法将其应用到机器人中
基于点	PointRCNN ^[48]	2019	第1阶段网络以自底向上的方式直接从点云生成三维提案, 第2阶段网络结合语义特征和局部空间特征, 在规范坐标系中细化提案	由于点云池化策略的模糊性, 该方法丢失了对几何信息编码的能力
	DAGC ^[49]	2020	双注意力模块和ResGCN模块充分利用了点关系和特征的重要性	引用的EdgeConv模块并未考虑相邻点间的向量方向, 导致损失部分局部信息
	Part-A ^[251]	2021	提出两种三维候选框生成策略以处理不同的场景. 无锚策略具有更高的内存效率, 而基于锚的策略具有更高的目标召回率	从稀疏点云中很难识别行人的方向, 导致检测到的行人性能稍差

目标检测性能比较如表7所示, “E”表示“Easy”, “M”表示“Moderate”, “H”表示“Hard”. 通过对比, IPV-RCNN 的类别精度明显高于其他基于体素、基于多视图的方法和基于点的方法, 主要原因在于提出的在线数据处理方法提高了训练效率, 并降低计算复杂度. 综合分析, 基于点的方法的类别精度均达到了较高水平且相对稳定, 主要是因为体素化和多视图方法会造成部分重要信息丢失, 三维 CNN 的存储和计算效率都很低, 而基于点的方法直接处理点云, 避免了有用信息丢失的问题.

5 总结

本文分析与比较了近年来深度学习在处理大规模

三维点云的点云分割、形状分类、目标检测等3个方面的研究进展, 以及常用的数据集和相关评价指标, 并针对算法精度、效率、点云规模等方面, 分析比较各类方法的优势与不足.

综上所述, 基于深度学习算法的大规模三维点云处理今后的发展趋势如下.

(1) 点云分割

点云规模: PointNet^[22]、PointNet++^[23] 等算法通过点云分块、对象部分之间上下文关系等策略处理点云导致计算复杂度高、内存占用较大, 因此实现直接处理百万级的大规模点云是今后的重要研究课题. 可以考虑结合超点图、给出与点-体素联合等思路, 能较好

地降低计算复杂度和内存占用。

计算效率: 采样方法直接影响计算效率, SGNP^[27]采用最远点抽样法导致内存占用较高、耗时巨大, 而

RandLA-Net^[25]采用随机抽样法降低内存和时间消耗。因此, 可以考虑结合 RandLA-Net 基于随机抽样法降低空间复杂度、给出 Transformer 等思路, 提高计算效率。

表7 KITTI 数据集上的不同目标检测算法性能对比 (%)

类别	方法	Car (IoU=0.7)			Pedestrians (IoU=0.5)			Cyclists (IoU=0.5)		
		E	M	H	E	M	H	E	M	H
基于多视图	MV3D ^[37]	92.33	88.66	78.96	80.35	66.68	63.44	76.04	66.36	58.87
	Frustum PointNet ^[38]	81.20	70.39	62.19	51.21	44.89	40.23	71.96	56.77	50.39
	Liang等 ^[39]	82.54	66.22	64.04	—	—	—	—	—	—
	Yan等 ^[42]	84.88	80.77	77.44	—	—	—	—	—	—
基于体素	VoxelNet ^[43]	81.97	65.46	62.85	57.86	53.42	48.87	67.17	47.65	45.11
	PVF-Net ^[44]	87.80	76.79	72.68	51.27	46.59	40.49	72.62	60.30	53.33
	P2V-RCNN ^[46]	—	—	—	50.91	43.19	40.81	78.62	63.13	56.81
	IPV-RCNN ^[47]	96.28	92.26	84.39	—	—	—	82.64	68.27	61.32
基于点	PointRCNN ^[48]	85.94	75.76	68.32	49.43	41.78	38.63	73.93	59.60	53.59
	Feng等 ^[50]	87.21	77.94	73.70	53.07	44.60	42.30	75.91	61.12	54.30
	Part-A ² -free ^[51]	90.23	88.05	85.85	70.73	64.13	57.45	88.18	73.35	70.75
	Part-A ² -anchor ^[51]	90.42	88.61	87.81	70.37	63.84	57.48	88.31	73.07	70.20

(2) 形状分类

算法精度和健壮性: 局部特征直接影响算法精度, PointNet^[22]无法捕获局部特征导致精度不高、DGCNN^[18]虽然解决了此问题, 但由于 EdgeConv 未考虑相邻点间的向量方向, 依然损失了部分局部特征, 直接影响算法精度。因此, 可以考虑基于邻域查询机制、融合二维图像特征的三维局部几何特征将有助于提高算法精度。

(3) 目标检测

信息融合: 二维 RGB 图像中蕴含丰富的深度信息, 而点云数据具有独特的几何信息, 将图像与点云数据中的信息融合, 可以让网络学习到更多有用特征。所以, 目标检测性能更好, 可以考虑结合 AVOD 算法提取图像特征、MMS 算法学习图像与点间的信息融合等思路实现。

数据稀疏问题: 由于近程雷达点云传感器在扫描远距离目标或遮挡目标时, 采集的数据非常稀疏, 从而影响检测算法, 如 Frustum PointNet^[38]对遮挡目标的输出准确率较低、Part-A²^[51]从稀疏点云中很难辨别出行人的方向。因此, 可以考虑结合 IPV-RCNN 中基于数据增强的训练方法与 AVOD 算法实现检测精度和准确率的提高。

参考文献

1 张政旭, 庞为光, 谢文静, 等. 面向实时应用的深度学习研究综述. 软件学报, 2020, 31(9): 2654–2677. [doi: 10.13328/

j.cnki.jos.005946]

- Redmon J, Divvala S, Girshick R, *et al.* You only look once: Unified, real-time object detection. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016. 779–788.
- 彭天强, 栗芳. 哈希编码结合空间金字塔的图像分类. 中国图象图形学报, 2016, 21(9): 1138–1146. [doi: 10.11834/jig.20160903]
- Guo YL, Sohel F, Bennamoun M, *et al.* Rotational projection statistics for 3D local surface description and object recognition. International Journal of Computer Vision, 2013, 105(1): 63–86. [doi: 10.1007/s11263-013-0627-y]
- Wu ZR, Song SR, Khosla A, *et al.* 3D ShapeNets: A deep representation for volumetric shapes. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston: IEEE, 2015. 1912–1920.
- Dai A, Chang AX, Savva M, *et al.* ScanNet: Richly-annotated 3D reconstructions of indoor scenes. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017. 2432–2443.
- Armeni I, Sener O, Zamir AR, *et al.* 3D semantic parsing of large-scale indoor spaces. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016. 1534–1543.
- Behley J, Garbade M, Milioto A, *et al.* SemanticKITTI: A dataset for semantic scene understanding of LiDAR sequences. Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul: IEEE, 2019.

- 9296–9306.
- 9 Maddern W, Pascoe G, Linegar C, *et al.* 1 year, 1000 km: The Oxford RobotCar dataset. *The International Journal of Robotics Research*, 2017, 36(1): 3–15. [doi: [10.1177/0278364916679498](https://doi.org/10.1177/0278364916679498)]
- 10 Tompson J, Stein M, LeCun Y, *et al.* Real-time continuous pose recovery of human hands using convolutional networks. *ACM Transactions on Graphics*, 2014, 33(5): 1–10.
- 11 Zhang YJ, Meng XW. SPCR: Course resource service platform for computer supported teaching and online learning. *Proceedings of the 16th International Conference on Computer Science & Education (ICCSE)*. Lancaster: IEEE, 2021. 66–71.
- 12 He T, Shen CH, van den Hengel A. DyCo3D: Robust instance segmentation of 3D point clouds through dynamic convolution. *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Nashville: IEEE, 2021. 354–363.
- 13 张硕, 叶勤, 史婧, 等. 改进 RangeNet++损失函数的车载点云小目标语义分割方法. *计算机辅助设计与图形学学报*, 2021, 33(5): 704–711.
- 14 Guo MH, Cai JX, Liu ZN, *et al.* PCT: Point cloud transformer. *Computational Visual Media*, 2021, 7(2): 187–199. [doi: [10.1007/s41095-021-0229-5](https://doi.org/10.1007/s41095-021-0229-5)]
- 15 宋巍, 蔡万源, 何盛琪, 等. 结合动态图卷积和空间注意力的点云分类与分割. *中国图象图形学报*, 2021, 26(11): 2691–2702.
- 16 Fooladgar F, Kasaei S. A survey on indoor RGB-D semantic segmentation: From hand-crafted features to deep convolutional neural networks. *Multimedia Tools and Applications*, 2020, 79(7): 4499–4524. [doi: [10.1007/s11042-019-7684-3](https://doi.org/10.1007/s11042-019-7684-3)]
- 17 Li JX, Chen BM, Lee GH. SO-Net: Self-organizing network for point cloud analysis. *Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Salt Lake City: IEEE, 2018. 9397–9406.
- 18 Wang Y, Sun YB, Liu ZW, *et al.* Dynamic graph CNN for learning on point clouds. *ACM Transactions on Graphics*, 2019, 1(1): 1–13.
- 19 Wu WX, Qi ZA, Li FX. PointConv: Deep convolutional networks on 3D point clouds. *Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach: IEEE, 2019. 9613–9622.
- 20 Wang ZJ, Lu F. VoxSegNet: Volumetric CNNs for semantic part segmentation of 3D shapes. *IEEE Transactions on Visualization and Computer Graphics*, 2020, 26(9): 2919–2930. [doi: [10.1109/TVCG.2019.2896310](https://doi.org/10.1109/TVCG.2019.2896310)]
- 21 Chen C, Fragonara LZ, Tsourdos A. GAPointNet: Graph attention based point neural network for exploiting local feature of point cloud. *Neurocomputing*, 2021, 438: 122–132. [doi: [10.1016/j.neucom.2021.01.095](https://doi.org/10.1016/j.neucom.2021.01.095)]
- 22 Charles RQ, Su H, Mo KC, *et al.* PointNet: Deep learning on point sets for 3D classification and segmentation. *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu: IEEE, 2017. 77–85.
- 23 Qi CR, Yi L, Su H, *et al.* PointNet++: Deep hierarchical feature learning on point sets in a metric space. *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach: Curran Associates Inc., 2017. 5100–5109.
- 24 Zhao HS, Jiang L, Fu CW, *et al.* PointWeb: Enhancing local neighborhood features for point cloud processing. *Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach: IEEE, 2019. 5560–5568.
- 25 Hu QY, Yang B, Xie LH, *et al.* RandLA-Net: Efficient semantic segmentation of large-scale point clouds. *Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle: IEEE, 2020. 11105–11114.
- 26 Li Y, Ma LF, Zhong ZL, *et al.* TGNet: Geometric graph CNN on 3-D point cloud segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, 2020, 58(5): 3588–3600. [doi: [10.1109/TGRS.2019.2958517](https://doi.org/10.1109/TGRS.2019.2958517)]
- 27 Wang WY, Yu R, Huang QG, *et al.* SGPN: Similarity group proposal network for 3D point cloud instance segmentation. *Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Salt Lake City: IEEE, 2018. 2569–2578.
- 28 Wang XL, Liu S, Shen XY, *et al.* Associatively segmenting instances and semantics in point clouds. *Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach: IEEE, 2019. 4091–4100.
- 29 Pham QH, Nguyen T, Hua BS, *et al.* JSIS3D: Joint semantic-instance segmentation of 3D point clouds with multi-task pointwise networks and multi-value conditional random fields. *Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach: IEEE, 2019. 8819–8828.
- 30 Yang B, Wang JN, Clark R, *et al.* Learning object bounding boxes for 3D instance segmentation on point clouds.

- Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2019. 6737–6746.
- 31 Hermosilla P, Ritschel T, Vázquez PP, *et al.* Monte Carlo convolution for learning on non-uniformly sampled point clouds. *ACM Transactions on Graphics*, 2018, 37(6): 1–12.
- 32 Thomas H, Qi CR, Deschaud JE, *et al.* KPConv: Flexible and deformable convolution for point clouds. *Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. Seoul: IEEE, 2019. 6410–6419.
- 33 Wang L, Liu YX, Zhang SM, *et al.* Structure-aware convolution for 3D point cloud classification and segmentation. *Remote Sensing*, 2020, 12(4): 634–651. [doi: [10.3390/rs12040634](https://doi.org/10.3390/rs12040634)]
- 34 Tang ZF, Hao FS, Zhang QS, *et al.* Complementary multi-branch CNNs towards real-world 3D point classification. *Proceedings of IEEE International Conference on Real-time Computing and Robotics (RCAR)*. Xining: IEEE, 2021. 869–874.
- 35 Zhang YX, Rabbat M. A graph-CNN for 3D point cloud classification. *Proceedings of 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Calgary: IEEE, 2018. 6279–6283.
- 36 Lu XL, Liu BD, Liu WF, *et al.* Linked attention-based dynamic graph convolution module for point cloud classification. *Proceedings of IEEE International Conference on Image Processing (ICIP)*. Anchorage: IEEE, 2021. 3153–3157.
- 37 Chen XZ, Kundu K, Zhang ZY, *et al.* Monocular 3D object detection for autonomous driving. *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas: IEEE, 2016. 2147–2156.
- 38 Qi CR, Liu W, Wu CX, *et al.* Frustum PointNets for 3D object detection from RGB-D data. *Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Salt Lake City: IEEE, 2018. 918–927.
- 39 Liang M, Yang B, Wang SL, *et al.* Deep continuous fusion for multi-sensor 3D object detection. *Proceedings of 15th European Conference on Computer Vision (ECCV)*. Munich: Springer, 2018. 663–678.
- 40 Ku J, Mozifian M, Lee J, *et al.* Joint 3D proposal generation and object detection from view aggregation. *Proceedings of 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Madrid: IEEE, 2018. 1–8.
- 41 Liang M, Yang B, Chen Y, *et al.* Multi-task multi-sensor fusion for 3D object detection. *Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach: IEEE, 2019. 7337–7345.
- 42 Yan M, Li ZT, Yu XY, *et al.* An end-to-end deep learning network for 3D object detection from RGB-D data based on Hough voting. *IEEE Access*, 2020, 8: 138810–138822. [doi: [10.1109/ACCESS.2020.3012695](https://doi.org/10.1109/ACCESS.2020.3012695)]
- 43 Zhou Y, Tuzel O. VoxelNet: End-to-end learning for point cloud based 3D object detection. *Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Salt Lake City: IEEE, 2018. 4490–4499.
- 44 Cui ZH, Zhang ZH. PVF-NET: Point & voxel fusion 3D object detection framework for point cloud. *Proceedings of 17th Conference on Computer and Robot Vision (CRV)*. Ottawa: IEEE, 2020. 125–133.
- 45 Chen YL, Liu S, Shen XY, *et al.* Fast point R-CNN. *Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. Seoul: IEEE, 2019. 9774–9783.
- 46 Li JL, Sun Y, Luo SJ, *et al.* P2V-RCNN: Point to voxel feature learning for 3D object detection from point clouds. *IEEE Access*, 2021, 9: 98249–98260. [doi: [10.1109/ACCESS.2021.3094562](https://doi.org/10.1109/ACCESS.2021.3094562)]
- 47 Li YJ, Yang S, Zheng YC, *et al.* Improved point-voxel region convolutional neural network: 3D object detectors for autonomous driving. *IEEE Transactions on Intelligent Transportation Systems*, 2022, 23(7): 9311–9317. [doi: [10.1109/TITS.2021.3071790](https://doi.org/10.1109/TITS.2021.3071790)]
- 48 Shi SS, Wang XG, Li HS. PointRCNN: 3D object proposal generation and detection from point cloud. *Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach: IEEE, 2019. 770–779.
- 49 Sun Q, Liu HY, He J, *et al.* DAGC: Employing dual attention and graph convolution for point cloud based place recognition. *Proceedings of 2020 International Conference on Multimedia Retrieval (ICMR '20)*. Dublin: ACM, 2020. 224–232.
- 50 Feng MT, Gilani SZ, Wang YN, *et al.* Relation graph network for 3D object detection in point clouds. *IEEE Transactions on Image Processing*, 2021, 30: 92–107. [doi: [10.1109/TIP.2020.3031371](https://doi.org/10.1109/TIP.2020.3031371)]
- 51 Shi SS, Wang Z, Shi JP, *et al.* From points to parts: 3D object detection from point cloud with part-aware and part-aggregation network. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(8): 2647–2664.

(校对责编: 孙君艳)