

DOI: 10.19650/j.cnki.cjsi.J2311134

# 基于深度学习的三维点云分析方法研究进展\*

陈慧娴, 吴一全, 张耀

(南京航空航天大学电子信息工程学院 南京 211106)

**摘要:**点云是目前自动驾驶、机器人、遥感、增强现实(AR)、虚拟现实(VR)、电力、建筑等领域最常用的三维数据处理形式,深度学习方法能够处理大型数据,且可自主提取特征,因此点云深度学习方法已逐渐成为研究热点。本文综述了近十年来基于深度学习的三维点云分析方法的研究进展。首先给出了三维点云深度学习的相关概念;然后针对点云目标检测与跟踪、分类分割、配准和匹配以及拼接这4种任务,分别阐述了相应的深度学习方法的原理,分析并比较了各自的优缺点;随后整理了18种点云数据集和4种点云分析任务的性能评价指标,并给出了性能对比结果;最后总结了点云分析方法目前存在的问题,对进一步的研究工作进行了展望。

**关键词:**三维点云分析;深度学习;目标检测;分类分割;点云配准;点云拼接

**中图分类号:** TP391.4 TH89 **文献标识码:** A **国家标准学科分类代码:** 520.20

## Research progress of 3D point cloud analysis methods based on deep learning

Chen Huixian, Wu Yiquan, Zhang Yao

(College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China)

**Abstract:** Point cloud is the most commonly used form of 3D data processing in the fields of autonomous driving, robotics, remote sensing, augmented reality (AR), virtual reality (VR), electric power, architecture, etc. Deep learning methods can not only handle large-scale data, but also extract features independently. Therefore, point cloud deep learning methods have gradually become a research hotspot. This article reviews the research progress of 3D point cloud analysis methods based on deep learning in the past decade. Firstly, the relevant concepts of deep learning for 3D point cloud are presented. Then, for the four tasks of point cloud object detection and tracking, classification and segmentation, registration and matching, and stitching, the principles of the corresponding deep learning methods are elaborated. Their advantages and disadvantages are analyzed and compared. Next, eighteen kinds of point cloud datasets and performance evaluation indexes for four types of point cloud analysis tasks are introduced. The performance comparison results are given. Finally, the existing problems of point cloud analysis methods are pointed out, and the further research work is prospected.

**Keywords:** 3D point cloud analysis; deep learning; object detection; classification and segmentation; point cloud registration; point cloud stitching

## 0 引 言

随着三维传感器技术的逐步发展,三维数据采集设备的种类和数量越来越多,三维数据的获取变得更加精准容易。三维点云是三维数据最重要的表达方式之一,它包含物体的几何、形状和尺度信息<sup>[1]</sup>,可以直接提供代

表物体表面形状的三维空间数据。三维点云分析其任务主要包括点云目标检测与跟踪、分类分割、配准和匹配以及拼接,在自动驾驶<sup>[2]</sup>、机器人感知<sup>[3]</sup>、遥感测绘<sup>[4]</sup>、增强现实<sup>[5]</sup>(augmented reality, AR)、虚拟现实<sup>[6]</sup>(virtual reality, VR)、文物修复<sup>[7]</sup>、工业农业生产<sup>[8]</sup>、医疗<sup>[9]</sup>、电力<sup>[10]</sup>、建筑<sup>[11]</sup>等领域有广泛的应用。因此,研究三维点云分析方法具有重要的现实意义。

收稿日期:2023-03-04 Received Date: 2023-03-04

\* 基金项目:国家自然科学基金(61573183)项目资助

目前主要采用传统机器学习方法和深度学习方法进行三维点云分析。传统机器学习方法在处理规模较小的数据时表现较好,但由于三维点云数据维度较高,参数空间会呈指数式增大,导致输入点的附近难以找到训练集中的数据点,分析三维点云变得极为困难。机器学习中大多数的特征(如像素值、纹理、形状、位置等)需要专家根据数据类型人工标识,提取特征的准确性决定了机器学习方法的性能。而深度学习具有足够的计算和训练数据能力,能自动筛选海量点云数据并学习点云的高级特征,同时可以发现三维空间结构的本质规律,大大降低了获取点云信息的成本。不同于传统机器学习分块、逐步解决再重新组合的解决方式,深度学习是一次性、端到端的解决方式,直接输入数据进行运算得到结果。虽然深度学习在训练过程中要花费大量时间学习参数,对进行矩阵运算的硬件设备要求较高,但深度学习的测试时间远小于机器学习,而且通过训练得到的模型能进行更精确的预测,性能更好。因此,本文讨论基于深度学习的三维点云分析方法。由于三维点云局部密度不等、分布稀疏,这使得利用深度学习直接分析点云数据存在一定的困难,如何高效地进行三维点云分析是近年来研究人员关注的热点问题。

当前相关综述文献情况如下:文献[12-14]分别介绍了点云目标检测、点云分割以及点云配准单个任务的深度学习方法,虽然有详细的对比分析,但比较片面。目前系统性归纳基于深度学习的三维点云分析方法的文章还比较少,文献[15-17]分别阐述了点云形状分类、目标检测与追踪以及分割任务,点云分类、分割和检测任务,点云模型重建与变换、分类分割、检测跟踪与姿态估计任务的深度学习方法的研究进展,但涉及的方法不够系统全面。本文在前人的基础上,对目前国内外常用的三维点云分析的深度学习方法进行了梳理和总结,是对上述综述文献的补充和扩展。

## 1 三维点云深度学习的基本概念

### 1.1 三维点云及其表示方法

三维点云(3D point cloud),顾名思义,是由  $X$ 、 $Y$  和  $Z$  三维坐标按照规则网格排列、表示物体外部曲面特征的采样点的集合,是表示点集的新型数据结构。除了包含采样点的色彩、亮度和坐标信息外,还提供回波强度、扫描角度与方向等信息。点云和图像数据的区别如表 1 所示。与二维图像数据相比,三维点云数据直观灵活,具有强大的表达能力,对光照和色彩变化不太敏感。此外,点云数据还包含丰富的三维空间信息,如相邻点之间的关联性以及空间结构等,因此点云更适用于计算机视觉分析任务。

表 1 点云和图像数据的区别

Table 1 The difference between point cloud and image data

类型	点云	图像
存储形式	$N \times 3$ 的矩阵,一行代表一个点,点数量为 $N$	$H \times W \times 3$ 的张量, $H$ 为高度, $W$ 为宽度
是否有序	无序 交换两行点云不变	有序 交换像素图像改变
数据索引	搜索所有的点	根据下标索引
影响因素	目标的反射特性	光照和色彩变化

点云的表示方法有以下 4 种,如图 1 所示。

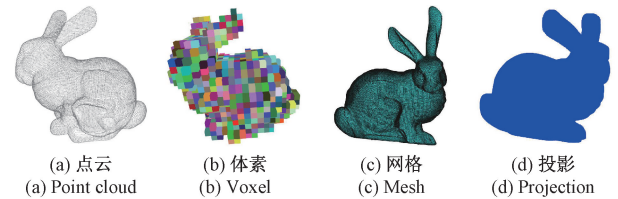


图 1 4 种点云的表示方法

Fig. 1 Four types of point cloud representation

1) 原始点云:一般采用  $N \times N \times N$  的三维矩阵表示,也可编码  $X, Y, Z$  3 个通道的栅格数据或以深度图的方式来表示。

2) 三维体素法:三维体素法是对点云进行体素化,将其转换为空间上规则排布的栅格,这使得图像领域的框架可以应用于三维领域。

3) 网格表示法:在原点云基础上借助具有局部连接关系的点构建网格,利用分布函数近似表示点云。

4) 二维投影法:需要对原始点云进行转换,从多个角度将点云数据投影到二维平面,再对得到的二维图像进行处理,点云数据丢失不可避免。

目前,常用的点云分析任务有 4 种:点云目标检测与跟踪、分类分割、点云配准和匹配以及点云拼接。至于点云处理方法,如点云去噪、滤波、压缩、超分辨率、修复、补全、重建等已另文综述。

1) 目标检测与跟踪:目标检测需要将点云中所有的目标物体找出,并在点云中确定物体的类别和定位。目标跟踪以定位为基础,对物体位置进行实时跟踪,得到完整的运动轨迹,在视频中应用较多。三维场景流估计可以通过理解动态场景下的 3D 运动来获取点云中点的动态信息,主要应用于自动驾驶及人机交互领域。

2) 分类分割:分割是将点云数据划分为不同区域,使同一区域内的点云特征(纹理、几何等)相似,不需要监督的先验知识。分类是判断三维点云属于哪一个类别,并为每个点或每一组分配一个语义标记,故点云分类也被称为点标记或者点云语义分割,一般需要监督学习

的方法来实现。目前的点云分类分割可细分为4种,包括语义、实例、零件以及全景分割。

3) 配准和匹配:通过位置变换可以使2个点云完全重合,即在同一坐标系下能转换成具有同一坐标表示的点云。常用的位置变换包含旋转、平移等。

4) 拼接:是指通过旋转平移变换,使各视角坐标系下的点云得到与目标点云相匹配的位姿,并融合成一个完整的点云,在机器人、三维重建领域应用较广。

## 1.2 点云分析的深度学习思想发展

点云分析方法经历了较长的发展阶段,核心问题是如何借助深度学习从大规模点云数据中提取有效信息,进而达到理想的分析精度。早期的处理方法将投影与深度学习方法中的二维卷积神经网络(convolutional neural network, CNN)结合,利用多视角投影将三维数据转换成二维图像,通过二维网络分析并将结果映射回三维空间。虽然二维分析方法比较成熟且获取了点云信息,但是多视角投影方法会造成点云空间信息的损失,数据转换也使点云分析计算成本增加。针对这些问题,一些研究者提出利用体素化点云,并通过稀疏卷积的方法舍弃空白体素,实现了三维卷积,计算量减少但是分辨率较低。之前的投影和体素方法仍旧存在点云空间信息的丢失,PointNet<sup>[18]</sup>利用多层感知机提取特征,通过最大池化操作保证点云的置换不变性。直接分析原始点云的方法简单,但忽略了点之间的相关性。为此,许多PointNet的优化网络也被提出,如融合点之间的关联信息的图结构网络。上述方法虽然学习到了点云的有效信息,但是点云分析精度还不够高,部分学者将二维和三维网络相结合来处理多模态数据,利用深度学习方法高效学习点云与体素或者图像的特征。目前,点云分析任务的性能有待提升,还需要结合实际应用需求深入探索深度学习方法以实现复杂点云任务的有效分析。

## 1.3 对深度学习效果的影响

点云数据采集时会混入噪声,点云自身的特性也给三维点云分析任务带来了不小的挑战。

1) 噪声和异常值。三维传感器在采集点云数据时必然会受到噪声的影响出现点云扰动,场景杂乱或扫描的模型被遮挡会导致异常值的出现,两者都会直接影响深度学习分析点云的精度。

2) 点云的不规则性。受采集设备和采集区域的影响,点的采样是不均匀的,点云数据分布密集和分散不均。而且点之间的距离不固定,点的排列也不规则,这使得直接使用二维深度学习框架处理三维点云很困难。

3) 点云的无序性。点云是扫描物体得到的一组点集,点存储顺序的变化并不会影响点云所表示的场景。

深度学习适合对有序的、结构化的数据进行分析,而点云则是杂乱无章的,没有结构的。因此,卷积神经网络直接处理点云数据难度较大。

4) 点云的旋转性。经过旋转或平移变化后的点云,其空间位置会发生改变,坐标也会改变,这可能导致无法在网络中正确识别点云。

## 2 点云目标检测与跟踪

点云目标检测与跟踪在自动驾驶、机器人感知等领域有着重要的应用。随着无人驾驶技术的深入研究,菜鸟等物流公司采用无人车在大学校园配送快递,目标检测是无人驾驶感知模块的核心,不仅需要准确地检测行人及车辆等动态目标的位置和尺寸,还要及时检测出路边障碍物和交通标志的类型,实时规划可行驶的安全区域。本节归纳了目标检测、目标跟踪和三维场景流的深度学习方法。

### 2.1 目标检测

根据输入网络的数据形式可以将点云目标检测方法分为以下3大类:基于结构化点云、基于原始点云和基于混合点云。

#### 1) 基于结构化点云的方法

基于结构化点云的方法可以划分为基于投影的方法和基于体素的方法2种。

##### (1) 基于投影的方法

从多个角度将点云数据投影到二维平面,将其离散化为二维图像,通过二维图像的处理方法进行目标检测。常用的投影视图有:前视图、鸟瞰图以及范围图。

大部分基于投影的方法采用鸟瞰图作为二维检测网络的输入。Zeng等<sup>[19]</sup>将离散雷达点云投影为二维网格以满足CNN的输入要求,借助区域预池化卷积技术,检测效率大大提升。文献[20]设计了一种无提案的单级探测器,利用残差块在主干网络中提取图像特征,在头网络中通过2个分支网络分别进行目标识别和定位任务,实现了高精度和实时性平衡的三维目标检测。为了减少投影带来的点云信息损失,Beltrán等<sup>[21]</sup>提出了对LiDAR点云进行密度归一化处理的BirdNet模型,通过Faster R-CNN获取二维图像中目标的位置、类别和尺寸,再结合高度信息实现三维目标检测。考虑到BirdNet<sup>[21]</sup>在计算目标高度时存在一定的误差,该团队又提出了BirdNet+<sup>[22]</sup>,使用两阶段检测架构对定向框进行回归,获取三维目标边界框的全部特征信息,避免了后续对高度信息的处理。Zhou等<sup>[23]</sup>提出将点云投影到圆柱表面,生成保留丰富信息的前视特征图,引入提案生成网络获取感兴趣的目标,通过学习逐点特征以估计最终边界框参数,检测的实时效率得到显著提升。由于目前多视图投

影方法大多需要对齐且固定的视图,Hamdi 等<sup>[24]</sup>提出了多视图转换网络,通过可微分渲染器进行训练以预测多视图网络的最佳视点,提升了目标检测的准确性。

## (2) 基于体素的方法

将点云数据转换为规则间隔的体素网格输入到网络中,每个体素单元包含点云的坐标信息和反射强度等信息,利用三维卷积获取高阶特征并学习。由于点云是稀疏分布的,因此体素网格也是稀疏的,基于体素的方法具有复杂的计算冗余性。

Li 等<sup>[25]</sup>首次将全卷积技术应用到三维检测任务中,将点云转换为二维点图,在此基础上又提出了将点云离散为包含长度、宽度、高度和通道信息的四维张量<sup>[26]</sup>,将二维卷积推广到三维领域中。为了降低计算成本,Vote3Deep<sup>[27]</sup>基于投票机制构建卷积层,利用了输入数据的稀疏性,在实现较高检测精度的同时检测速率略有降低。Zhou 等<sup>[28]</sup>提出了 VoxelNet,通过堆叠多个体素特征编码(voxel feature encoding, VFE)层来学习复杂的逐点特征和局部特征,但是该方法受到体素稀疏性和三维卷积操作的限制,检测速度很慢。

鉴于体素占用了大量的网络训练和推理时间,Yan 等<sup>[29]</sup>提出基于稀疏卷积的 3D 目标检测(sparsely embedded convolutional detection, SECOND)网络,应用稀疏卷积处理雷达点云数据,计算量和内存占用减少。由于点云的稀疏性导致搜索空间很大,文献[30]提出了 CenterPoint 架构,利用关键点检测器检测目标中心,采用体素编码网络卷积来获取检测目标的信息。为了减少传统三维卷积网络的计算冗余性,文献[31]设计了一种将体素划分与图表示方法结合的三维目标检测网络,通过图卷积模块有效学习不规则点云的特征信息。考虑到 SECOND 方法<sup>[29]</sup>提取的点云特征信息有限,文献[32]提出了一种稀疏三维卷积的方法 Reinforced SECOND,引入了堆叠三重注意力机制辅助体素特征编码,利用残差稀疏卷积网络提取更丰富的特征。

基于投影的目标检测方法本质是一种降维方法,将点云数据转化为二维图像,将三维目标检测问题转换到二维图像领域。基于体素的目标检测方法通过将二维卷积扩展到三维卷积,稀疏卷积方法提高了计算效率,实现了三维目标检测。无论是采用投影方法还是体素方法,都需要将原始点云数据转换为规则均匀的结构化点云。虽然能够利用二维检测方法获取较好的检测性能,也成功实现了三维卷积操作,但是点云特征信息和空间信息的损失以及高额计算量是不可避免的。

## 2) 基于原始点云的方法

为了保留更多的原始点云信息,直接将点云数据输入网络的方法应运而生。可分为基于 PointNet、基于注意力机制、基于图神经网络和其他方法 4 种。

## (1) 基于 PointNet 的方法

大部分三维目标检测方法基于 PointNet 的网络架构加以改进。PointPillars<sup>[33]</sup>以点云作为输入,通过一种新型编码器来学习垂直列的特征,使用二维卷积架构进行检测。为了充分利用训练数据的三维信息,PointRCNN<sup>[34]</sup>对点云进行语义分割并预测标签,再通过规范的边界框进行细化。为了减少计算成本,STD (sparse-to-dense 3D object detector)<sup>[35]</sup>基于球形锚点生成准确的候选区域,通过 PointsPool 层获取密集的区域特征表示,实现高质量的预测。由于三维目标的中心可能不在物体表面,Qi 等<sup>[36]</sup>提出了 VoteNet,将传统的霍夫投票与深度学习方法结合,以端到端的可微网络架构实现三维检测。考虑到 VoteNet 缺乏颜色信息,Qi 等<sup>[36]</sup>又提出了专用于 RGB-D (red-green-blue-depth) 场景的 ImVoteNet<sup>[37]</sup>,将二维的语义、纹理特征融合到三维特征中并辅助三维投票。由于先前的方法在最终边界框回归前丢弃了许多重要的前景点,Zhang 等<sup>[38]</sup>提出了 IA-SSD,利用两种基于学习的实例感知下采样策略来分层选取感兴趣目标的前景点,在 KITTI 数据集上实现了每秒 83 帧的检测速度。

## (2) 基于注意力机制的方法

一些方法将能够自适应地增强目标关键特征信息的注意力机制引入到网络中。针对噪声环境下的目标检测较差的问题,三重注意力网络(triple attention network, TANet)<sup>[39]</sup>借助三重注意和粗到细回归模块强化了对象的关键信息,增强了定位的精确度,算法稳健性增强。大多数 3D 目标检测方法只考虑单一对象,忽略了对象之间的上下文关系,文献[40]采用自注意力机制和多尺度特征融合模块来获取多层次上下文信息,检测精度得到有效提高。

## (3) 基于图神经网络的方法

还有一些方法通过构建合适的图神经网络来提取特征。文献[41]开创性地提出利用图表示的图卷积网络检测车辆,通过 R-GCN 和 C-GCN 模块提取区域特征并聚合上下文信息,实现了精确的三维目标检测。目前直接输入点云的网络的计算成本较大,Point-GNN<sup>[42]</sup>提出连接固定半径内的点来构建图,通过自动校准机制和评分机制实现准确的顶点的类别和边界框检测,同时在单次拍摄中可以快速检测多个对象。

## (4) 其他方法

鉴于三维目标检测任务需要大量的标记数据,三维数据的采集和标记耗时较长,一些方法尝试在深度学习方法中融入弱监督和无监督思想。文献[43]提出了无监督预训练网络 PointContrast,使用全卷积几何特征和无监督指标对大规模数据集进行预训练,在室内和室外、合成和真实数据集上都取得了更高的检测和定位精度。

Qin 等<sup>[44]</sup>提出了弱监督的三维检测方法 VS3D,引入无监督模块利用标准化的点云密度来生成潜在的3D框,再将其投影到图像上,采用预先训练好的图像检测网络结果进行监督学习。Xu 等<sup>[45]</sup>针对室内场景的三维目标检测,提出了回归现实(back to reality, BR)的方法,使用合成的3D形状构建原始虚拟场景,通过增强弱标签得到全标注标签,以虚拟标签来补充和细化真实标签,再对网络进行额外的监督训练。

直接处理点云数据不需要投影和体素化等额外处理过程,避免了信息的丢失。基于 PointNet 的方法对 PointNet 的网络架构加以改进,但是对于点云局部特征信息的捕获不足;基于注意力机制的方法使原始点云的特征信息得到了充分利用,但还是会受到稀疏点云数据的干扰;基于图神经网络的方法使用了点之间的信息,能更好地感知场景信息,但是图结构增加了时间成本,构建适合的图结构也是一个难点;弱监督和无监督学习可以通过少量标注数据实现较高水平的三维检测性能,极大地降低了计算成本。目前,仍需要克服点云数据的稀疏性和不规则性,这也是基于原始点云检测方法的研究关键点。

### 3) 基于混合点云的方法

标准的三维点云数据虽然包含目标物体的坐标及深度信息,但是缺乏高分辨率的纹理信息;二维图像数据可以提供丰富的颜色和纹理信息,但是缺少深度和空间信息;体素数据本身并不含有空间中位置的数据,通过有序存储反映空间结构,当分辨率选择过大时,内存占用几乎呈三次方增长。混合点云数据融合能够充分利用点云、图像或体素数据的特征信息,有助于增强三维点云目标检测的网络性能。

#### (1) 基于点云-体素的方法

基于点云的方法直接从原始点云中学习特征进行检测,但是运算量较大;基于体素的方法计算效率较高,但丢失了空间信息,降低了检测框的定位精度。而基于点云-体素的方法结合了基于点云和基于体素的特征优势,提升了三维检测性能。Fast Point R-CNN<sup>[46]</sup>通过二维和三维卷积分别获取体素和点云特征,使用注意力机制优化检测效果。为了学习更多有差异的点云特征,PV-RCNN<sup>[47]</sup>将多尺度体素特征编码与关键点相结合,发挥了体素的高效学习和点的灵活感受野的优势,提升了目标位置检测的准确性。为了提升三维动态目标检测的精度,文献[48]提出了两阶段多次多尺度特征融合的检测框架,第1阶段提取多尺度体素特征和点云关键点特征并融合,生成预选框;第2阶段吸收参考点周围的信息再次进行特征融合,并在检测模块中输入最终特征以优化预选框。

#### (2) 基于点云-图像的方法

基于点云-图像的方法主要有特征融合与视锥两种。

特征融合方法主要思想是进行基于区域的特征融合,并行处理点云和图像。多视图三维网络(multi-view 3D object detection network, MV3D)<sup>[49]</sup>以多模态数据作为输入,采用多视图对点云进行编码并预测3D边界框。聚合视图对象检测网络(aggregate view object detection network, AVOD)<sup>[50]</sup>对 MV3D<sup>[49]</sup>进行了改进,从雷达点云和图像中生成高分辨率特征图,通过区域建议网络(region proposal network, RPN)架构进行多模态特征融合。考虑到大多数检测方法难以获取多尺度上下文信息,文献[51]提出了空间通道注意网络(spatial-channel attention network, SCANet),通过扩展空间上采样模块丰富高级特征信息,采用多层次融合实现视图特征的交互。鉴于先前的点云-图像融合方法的精度一般,Xie 等<sup>[52]</sup>提出了三维多传感器多任务网络 PI-RCNN (pointcloud-image RCNN),使用分割子网络从图像中提取语义信息,通过对三维点执行逐点连续卷积操作直接融合多传感器特征和三维点,生成并优化目标三维候选框。由于网络计算成本较大,RoIFusion<sup>[53]</sup>基于轻量级网络架构,利用融合关键点生成(fused keypoints generation, FKG)层来估计点云和图像上的一组关键点并生成潜在的感兴趣区域,通过二维和三维池化层分别提取几何特征和纹理特征并融合。为了提升点云目标检测的精度,文献[54]采用 Point-GNN 网络获取点云目标的三维框和三维中心,再融合图像目标外形来确定三维目标的最终位置及外形。

基于视锥的检测方法使用图像信息辅助点云生成三维边界框。核心是产生视锥区域,采用二维目标检测器生成二维候选区域,再将其映射回三维点云中得到视锥候选区域。Frustum PointNets<sup>[55]</sup>是最早提出利用2D目标检测器生成3D视锥区域的方法,通过获取每个视锥区域的特征得到精确的3D边界框。为了提高网络架构的通用性,PointFusion<sup>[56]</sup>应用融合网络融合图像和点云特征信息,通过图像对应的视锥区域内的点云信息预测三维边界框。针对点云的稀疏性,尺度不变和特征权重网络(scale invariant and feature reweighting network, SIFRNet)<sup>[57]</sup>使用前视图和视锥点云信息进行目标检测,具有尺度不变性。由于先前的视锥方法前景点过少,会出现分割错误,文献[58]提出了 Frustum ConvNet,针对每个二维区域生成一系列视锥体,提取视锥体特征并转换为特征图,通过全卷积网络(fully convolutional network, FCN)预测3D边界框。考虑到传统目标检测方法需要大量的标记数据,文献[59]提出了视锥体感知几何推理方法(frustum-aware geometric reasoning, FGR),利用二维边界框标记与稀疏点云生成三维伪标签进行检测训练。为了提高检测的实时性,文献[60]提出了多模态信息融合的三维检测方法,根据点云与图像的投影关系生成2D矩

形框所对应的 3D 视锥区域,并对区域内的点云聚类并拟合三维边界框。

基于点云-体素的方法在获取丰富点云特征信息的同时,兼顾了运算效率。基于点云-图像的方法能够融合点云和图像的特征信息,获取三维对象可能存在的位置。从理论上来说,混合点云的方法可以为深度神经网络提供更丰富、更全面的点云信息,目标检测性能应该有明显提高。而在实验过程中,混合点云的方法难以充分融合不同类型数据的特征,目标检测效果暂不如单模态点云数据的方法。但是,随着深度学习技术的不断发展,混合点云的检测方法有望成为重要的研究方向。

## 2.2 目标跟踪

点云目标跟踪需要根据给定的对象位置预测其接下来的运动轨迹。文献[61]同时将图像和雷达点云数据作为输入,采用卷积神经网络进行检测和匹配,生成精确的运动轨迹。考虑到恶劣天气和光照条件会干扰跟踪任务,文献[62]提出利用雷达点云进行 3D 目标检测,将 Siamese 网络与自动编码器结合,使用语义信息对网络进行正则化,进而提升跟踪的性能。为了实现实时的点云跟踪,Simon 等<sup>[63]</sup>提出了 Complexer-YOLO,融合了二维视觉语义信息和点云数据,通过尺度旋转平移分数(scale-rotation-translation score, SRTs)检测目标姿态,充分利用了时间信息,精确度和鲁棒性得到提高。由于之前的跟踪算法对转弯和急停等极端运动敏感,文献[64]提出了端到端网络 PointTrackNet,仅需要输入两个相邻的原始点云,使用数据关联模块合并点特征,生成预测的边界框和运动轨迹,但是该方法难以处理高度遮挡的目标。鉴于现有方法在稀疏的小规模点云场景的跟踪效果较差,文献[65]提出附加神经网络的单目标跟踪方法,在训练时分割前景并对中心坐标进行偏移回归,更好地学习并提取目标点云的典型特征,实现高效、实时的单目标跟踪。

## 2.3 三维场景流估计

三维场景流估计是一种新颖的点云分析思路,借助动态场景中的矢量场来表示点的运动状态。为了实现大规模场景点云的高效分析,分级排列自形网格流网(hierarchical permutohedral lattice flownet, HPLFlowNet)<sup>[66]</sup>基于双边卷积层原理分析非结构化点云,归一化不同密度的点云,计算成本大大降低。Li 等<sup>[67]</sup>针对场景流数据注释稀少的问题,将自监督学习方法结合到场景流估计中,利用点匹配生成伪标签,通过局部一致性进一步细化,训练时不需要任何人工标记,但生成伪标签耗时较多。为了提高无序点云特征捕获的有效性,点体素递归全对场变换(point-voxel recurrent all-pairs field transforms, PV-RAFT)<sup>[68]</sup>构建了点体素相关场,融合

了点与体素的相关性并获取全对字段的特征信息,通过截断和细化模块对迭代更新的场景流进行平滑估计。考虑到采样时可能存在区域遮挡,文献[69]设计了遮挡引导场景流网络(occlusion guided scene flow network, OGSF-Net),将场景流和遮挡图联合起来进行学习和估计,网络的遮挡检测性能大大提高。鉴于先前的方法未考虑点云中点的关联性,文献[70]提出了邻域一致性传播更新方法(neighborhood consistency propagation update method, NCPUM)方法对困难点匹配进行优化,通过置信度图将匹配准确的点的场景流传播到局部特征信息不足的点上,整体场景流估计的准确度得到提升。

## 2.4 点云目标检测与跟踪方法小结

点云目标检测方法的对比分析如表 2 所示。基于投影和体素的目标检测方法需要数据转换,存在信息缺失;基于原始点云的方法要求解决点云特性带来的一系列问题;基于图的检测方法关键在于构建适用的图结构;视锥检测方法受限于 2D 目标检测网络。点云目标跟踪技术能够充分利用点云的几何信息,已经逐渐从单目标发展至多目标。但是多目标之间相互遮挡,且特征相似,这给点云多目标跟踪方法构成了不小的挑战。目前,三维场景流估计的研究相对较少,是目前点云目标检测与跟踪方法的一个研究热点。

## 3 点云分类分割

目前点云分类分割任务主要有 4 种:语义分割(semantic segmentation)、实例分割(instance segmentation)、零件分割(part segmentation)和全景分割(panoptic segmentation)。点云语义分割是指对点云中的每一个点划定相应的分类,按不同分类为每一个点分配一个语义标签,但不能区分同一分类中的点。而点云实例分割是在语义分割的基础上,对同一个类别中的点进行区分。零件分割需要区分属于不同部件的点,全景分割则需要将前景与背景分离并避免混叠预测。下面将从语义分割、实例分割、零件分割和全景分割 4 个部分介绍深度学习的点云分类分割方法。

### 3.1 语义分割

根据点云的不同表现形式,可分为基于结构化点云和基于原始点云两种方法。

#### 1) 基于结构化点云的方法

结构化点云需要对非结构化点云进行转化,目前主要有多视图和体素 2 种方式。

##### (1) 基于多视图的方法

将非结构化的点云投影为 2D 图像。文献[71]用二维卷积网络训练三维物体从多个投影视角得到的二维图

表2 点云目标检测任务主要方法对比

Table 2 Comparison of major methods for point cloud object detection task

类别	文献	具体方法	贡献	局限性
基于结构化点云	[20]	无提案的单级探测器	通过两个分支网络分别进行识别和定位	难以检测垂直堆叠的多个目标
	[23]	将点云投影到圆柱表面	提升了检测的实时效率	不适用于严重遮挡场景
	[28]	端到端可训练的网络框架	直接处理稀疏点云	检测速度较慢
	[29]	改进的稀疏卷积网络	角度损失回归和数据增强新方法	行人和骑车者检测性能较差
基于原始点云	[33]	PointPillars 编码器	使用二维卷积实现端到端学习,高精度	运行时间变长
	[38]	实例感知下采样策略	前景点本质上比背景点更重要,高效	容易受到类不平衡分布的干扰
	[39]	注意力机制	三重注意模块和粗到细回归模块	仅提升了噪声很强场景的性能
	[40]	自注意与多尺度特征融合	利用多级上下文信息	处理复杂场景的效果一般
	[41]	基于图表示的图卷积网络	引入图表示方法细化候选区域框	车辆和植被遮挡会导致误报
	[42]	图神经网络单元	自动配准机制降低了误差	算法运行速度有待优化
	[43]	无监督预训练	全卷积几何特征,进一步扩展数据收集	小规模数据集会出现过拟合
其他方法	[45]	弱监督	训练仅需要不到5%的标注量	在小物体检测方面仍有困难
	[46]	两阶段的检测网络	同时利用体素表示和原始点云,高效	会丢失部分准确的位置信息
基于混合点云	[48]	多次多尺度特征融合	体素和点云关键点进行多次特征融合	优化雷达和相机间的校正关系
	[49]	多视角投影	多模态融合,将雷达点云与图像作为输入	实际场景的速度较慢
	[50]	特征融合区域提议网络	对极端天气也有泛化性,计算成本较低	仅对车辆前方物体敏感
	[54]	改进 Point-GNN	点云与图像数据融合确定三维目标框位置	仅考虑了部分车辆的外形尺寸
	[55]	基于二维区域生成视锥区域	在强遮挡下也能精确估计3D边界框	夜间使用受限
	[56]	同时利用图像和点云信息	引入融合网络,具有普适性	输入点数过少时性能迅速下降
	[59]	视锥体感知几何推理	利用上下文信息自适应调整阈值	需要找到最佳阈值

像,再经池化层聚合特征得到分类分割结果,但存在复杂点和曲面问题。针对该问题, SnapNet<sup>[72]</sup>通过生成 RGB (red-green-blue) 视图和几何特征的深度视图作为语义分割网络的输入,使用 FCN 进行标记,再将语义标记图像投影回原始点云,适用于多种类型的点云分割。为了提升 SnapNet<sup>[72]</sup>的分割能力, Guerry 等<sup>[73]</sup>又设计了多视图卷积神经网络 SnapNet-R,可以独立分割三维场景生成的二维图像并分配标签,再投影回对应的点云中。鉴于上述方法得到的投影图像特征较稀疏, SqueezeSeg<sup>[74]</sup>使用球面投影将稀疏的 LiDAR 点云转化为密集的二维网格,在 CNN 中引入了条件随机场(conditional random field, CRF)再进行语义分割,经 CRF 细化后通过传统聚类方式得到语义标签。由于 SqueezeSeg<sup>[74]</sup>对失调噪声敏感,文献[75]提出了改进的 SqueezeSegV2 模型,借助上下文聚合模块(context aggregation module, CAM)聚合更多的上下文信息,进而减轻失调噪声的干扰,提高分割精度。考虑到 SqueezeSegV2 模型<sup>[75]</sup>没有充分融合点云信息,文献[76]通过前视投影得到点云深度图,并结合点云特征构建多特征图像,使得网络可以处理高级语义信息,提升了行人和自行车的语义分割精度。

## (2) 基于体素的方法

将点云转换为 3D 网格表示,再通过神经网络进行语义分割。VoxNet<sup>[77]</sup>是最早构建的三维体素卷积网络,充分利用点云信息有效地处理点云数据,但是稀疏的点云数据使网格维度增加,计算时间变长。为了降低体素方法的计算量,研究人员提出了一系列改进网络。针对点云的稀疏性, FPNN (field probing based neural network)<sup>[78]</sup>使用场探测滤波器捕捉点云特征,网络的计算量大大减少。文献[79]应用 3D 全卷积网络提供体素概率,采用三线性插值将其转移回原始点云,再提供全连接条件随机场(fully connected conditional random fields, FC-CRF)获得 3D 分割标签。PointGrid<sup>[80]</sup>则基于三维卷积网络,借助点和网格的混合表示来学习局部几何形状的细节信息,分割任务占用的内存很小。此外,一些方法借助八叉树和 Kd-tree 结构来降低计算量和内存占用。OctNet<sup>[81]</sup>基于八叉树结构根据数据疏密程度分割三维空间,对高分辨率输入也有好的点云标记结果。文献[82]提出了基于 Kd-tree 空间划分结构的 3DContextNet,能够逐步学习局部和全局上下文信息并聚合特征。这 2 种方法的语义分割效果不错,但是依赖体素边缘信息。

基于多视图的分割方法容易受投影角度的影响,存在遮挡和误差问题,适用于小型或特定的场景,实用性强。体素化分割方法的空间复杂度较高,需要占用较多内存和运算量,存在量化伪影问题,实用性较弱。这 2 种结构化点云的分割方法都没有充分利用点云属性,为了减少三维点云细节信息的损失,部分研究人员提出新思路,直接处理原始点云。

## 2) 基于原始点云的方法

基于原始点云的方法直接使用三维点的位置信息,不会造成空间信息丢失,随着 PointNet<sup>[18]</sup>的提出和改进网络的不断出现逐渐成为主流的点云分类分割方法。但是 PointNet 没有考虑每个点的局部特征和点之间的相关性,研究人员针对这些问题提出了许多改进方法,包括基于特征学习、卷积、无监督和旋转性方法 4 类。

### (1) 基于特征学习的方法

针对 PointNet 局部特征提取不足的问题做出了改进。PointNet++<sup>[83]</sup>借助分层抽取特征的思路,通过迭代最远点采样法对局部区域进行划分并获取局部特征,重复使用 PointNet 逐层提取点云深层特征。由于尺度变换会影响点云特征捕获,文献[84]借鉴了二维图像中尺度不变特征变换(scale invariant feature transform, SIFT)的优点,通过方向编码单元在关键方向上提取最近点的特征,堆叠多尺度下的方向编码单元来实现尺度不变以提高点云分割精度,但是计算量增加。形状上下文网络(ShapeContextNet, SCN)<sup>[85]</sup>则将局部形状上下文和深度神经网络结合,通过分层模块传播局部零件信息来捕获局部和全局信息。考虑到上述方法未利用局部点之间的联系,PointWeb<sup>[86]</sup>通过自适应特征调整(adaptive feature adjustment, AFA)模块来学习每个点对其他点的影响并传递信息,应用密集连接的点对学习并传递局部上下文信息。由于特征增加导致了内存占用问题, RandLA-Net<sup>[87]</sup>采用随机采样法,引入保留更多几何信息和局部结构的局部特征聚合模块,以实现高效的大规模三维点云语义分割。为了从大规模点云中学习更有效的特征,文献[88]设计了 SCF-Net,通过学习每个点的  $z$  轴旋转不变表示,根据几何距离与特征距离自动学习局部上下文特征,再结合每个点的全局上下文特征综合改善网络语义分割效果。

### (2) 基于卷积的方法

主要有 2 种思路,第 1 种是针对现有 CNN 进行优化。PointCNN<sup>[89]</sup>将  $\chi$ -变换矩阵与点云数据点乘得到规范排列的点云,再用分层卷积捕获局部特征。由于离散卷积的适用性易受数据结构限制,文献[90]提出了参数连续卷积算子,借助跨越整个连续向量空间的参数化核函数学习数据结构特征。鉴于 CNN 对不同形状点云的泛化性较差,文献[91]设计了权重位于核点决定的欧式

空间内的核点卷积(kernel point convolution, KPConv),可根据点云分割对象的几何形状自适应变化。为了优化 CNN 的计算效率,文献[92]提出了离散 CNN 的泛化,用连续内核替换离散内核处理非结构化点云;而 Pointwise<sup>[93]</sup>采用逐点卷积算子学习并输出每个点的特征,网络简单有效;ShellNet<sup>[94]</sup>则通过定义同心球壳划分区域上内壳到外壳的卷积顺序,统计点信息并提取特征以提升卷积效率。考虑到点云密度对特征信息的干扰,文献[95]提出了密度重加权卷积 PointConv,借助多层感知机网络和核密度估计学习权重函数,实现了平移不变和置换不变的卷积运算;文献[96]则设计了方向诱导卷积算子(direction-induced convolution, DIConv)对点云进行滤波处理,将密集的局部点稀疏地投影到标准有序的方向集空间,并保留点之间可区分的几何特征,以良好区分密集点云区域内的点特征。为了学习更丰富的点云细粒度局部特征,文献[97]构建了基于自注意力机制的网络(space self-attention PointNet++, SSA-PointNet++),综合采样点中心和邻域自注意力机制学习采样点邻域拓扑结构,利用注意力池化模块传递有效的细粒度信息。

另一种改进思路是将图结构和卷积运算结合。文献[98]建议在局部图使用谱图卷积,将结构信息并入点的邻域中,通过递归聚类谱坐标在图节点聚集特征。正则图卷积神经网络(regularized graph convolutional neural network, RGCNN)<sup>[99]</sup>则将点的特征视为图形信号,根据每个点的坐标和法线构建图形并进行卷积和特征学习。为了改善图结构的边界分割效果, DGCNN<sup>[100]</sup>基于边缘卷积(EdgeConv)模型,通过  $k$  邻近方式提取中心点和邻域点之间的边特征,然后对边特征卷积;图注意力卷积网络(graph attention convolution network, GACNet)<sup>[101]</sup>则结合注意力机制来计算图结构内中心点和各邻接点的边缘权重。由于图卷积需要消耗大量内存,分层深度图卷积神经网络(hierarchical depthwise graph convolutional neural network, HDGCN)<sup>[102]</sup>结合 DGConv 块和层次结构来提取图结构中有效的局部和全局特征。考虑到大型点云应用图结构提取特征比较困难,文献[103]引入了可分离的球形卷积核,通过范围搜索和最远点采样构建图卷积网络来学习典型特征。针对点云分割的鲁棒性问题,三维图卷积网络(3D graph convolution networks, 3D-GCN)<sup>[104]</sup>引入了可在训练中学习形状和权重的新内核,借助图最大池化机制提取并总结不同尺度的点云特征,数据移位和缩放的影响有效降低;旋转不变量与多尺度特征图卷积神经网络(rotation-invariant and multi-scale feature graph convolutional neural network, RMGnet)<sup>[105]</sup>则将 3D 形状特征与多尺度特征图卷积神经网络结合,通过多分辨率分析实现旋转不变的分割。



### (3) 基于无监督的方法

优化目标是减少三维数据的标注。SO-Net<sup>[106]</sup>利用自组织映射(self-organizing map, SOM)拟合点云的空间布局,将单个点和SOM节点分层提取特征,通过预先训练提升网络的分割性能。为了综合点云的局部和全局特征信息,文献[107]设计了基于点的无监督域自适应网络(domain adaptation network for point, PointDAN),用自适应节点模块实现局部对齐,借助对抗训练策略来学习全局特征对齐,以端到端的方式获取特征并进行语义分割;文献[108]则构建了双向全局-局部推理(global-local reasoning, GLR)的无监督网络,将不同级别的局部表示和全局表示映射到共享特征空间,再从数据中获取需要的语义信息;文献[109]结合对抗思想提出了基于端到端机制的三维对抗自编码器(3D adversarial autoencoder, 3DAAE)网络,能够在学习空间特征的同时生成三维形状;而文献[110]设计了一个多尺度区域关系特征分割网络(multiscale regional relation feature segmentation network, MS-RRFSegNet),应用区域关系特征推理模块(regional relation feature reasoning module, RFRM)自动学习特征关联性。

### (4) 基于旋转性的方法

当点云进行旋转变换时,分类分割任务的性能会急剧下降,尽可能降低点云旋转的干扰是分类分割成功且准确的首要条件,因此提出了基于旋转性的改进方法。部分研究人员通过球形体素卷积方法<sup>[111]</sup>实现点云的旋转不变性。为了提取点云的旋转不变特征,点态旋转不变网络(pointwise rotation invariant network, PRIN)<sup>[112]</sup>通过密度感知自适应采样模块来处理球坐标畸变点问题。针对球体结构的三维卷积难题,文献[113]设计了一种基于端到端的球形分形卷积神经网络,利用规则二十面体晶格及其分形的结构来近似球体,以便将原始点自适应地投影到新结构上并卷积。

由于离散球体无法保证严格的点云旋转不变,研究人员考虑采用点投影方法<sup>[114-115]</sup>来学习旋转不变性。SRINet<sup>[114]</sup>引入了点投影来重构原始点云的坐标,通过图形聚合和关键点检测来感知三维形状结构;基于点云的多头注意力卷积层网络<sup>[115]</sup>通过点投影的坐标变换得到严格旋转不变的点云表示,网络的特征提取能力得到增强。不同于上述点投影方法,一些方法设计了旋转不变模块<sup>[116-119]</sup>。文献[116]设计了一种可实现旋转不变性的点云卷积算子RICov,借助旋转不变特征定义卷积,能够学习距离和角度等潜在的旋转不变几何特征;而ClusterNet<sup>[117]</sup>引入了一种严格旋转不变(rigorous rotation-invariant, RRI)的点云表示形式,通过无监督层次聚类学习点云的底层几何结构;文献[118]则提出了用于3D旋转估计的RotPredictor模块,能以无监督的方式将点云从

随机视点旋转到规范视点,分类准确性大幅提高;考虑到ClusterNet<sup>[117]</sup>的局部特征并非旋转不变的,文献[119]设计了有序格拉姆矩阵(sorted Gram matrix, SGM),该矩阵对任意旋转都能保持不变,可以用作卷积网络的特征。

基于特征学习的方法通过构建新的网络模块来获取点云的局部信息,将局部特征与全局特征结合,充分学习点云的空间结构,进而提升语义分割的精度。基于卷积的方法将二维卷积网络扩展到三维空间,通过设计新卷积算子优化CNN的网络结构,或者利用图卷积网络进行点云语义分割,这两种方法都需要解决三维卷积操作带来的计算内存增大问题。无监督方法主要借助自编码器实现类型的跨越式生成,通过少数带标注的数据就能实现特征学习,语义分割性能进一步增强。基于旋转性的方法通过借助球形体素卷积、投影方法以及旋转不变模块可以增强旋转变换的鲁棒性,进而提升了旋转点云分割的准确性。与PointNet结构相比,上述4种方法保留了更多的点云原始信息,进一步提高了语义分割的准确性。

## 3.2 实例分割

点云实例分割不仅要分出点云的类别,还需要将其分为不同的对象,受到研究人员的广泛关注,主要方法包括基于候选区域和免候选区域2种。

### 1) 基于候选区域

顾名思义,基于候选区域的方法需要先获取感兴趣的候选区域,再在区域内进一步预测点云实例标签。

生成形状建议网络(generative shape proposal network, GSPN)<sup>[120]</sup>能通过噪声的观察结果重建形状,生成高质量的候选区域,利用基于区域的PointNet实例分割框架(region-based PointNet, R-PointNet)对非结构化数据进行灵活处理。考虑到点云的颜色信息也包含丰富特征,3D-SIS网络<sup>[121]</sup>通过神经网络学习RGB和点云特征,将二维卷积网络提取到的二维特征反向投影到对应的三维网格上再进行实例分割。为了识别并分割单个对象,PanopticFusion<sup>[122]</sup>通过空间散列体积分图表示区分单个对象并执行密集语义标记,还可以恢复物体的类别标签和网格标记。针对上述实例分割方法运算量大的问题,3D-BoNet<sup>[123]</sup>基于逐点多层感知架构(multi-layer perception, MLP),对所有实例个体进行3D边界框回归,然后在每个边界框中探索点的特征以掩盖背景,并预测每个实例的掩码。由于3D-BoNet<sup>[123]</sup>在边界框回归时并未考虑密度变化对特征提取的影响,IP-BoNet<sup>[124]</sup>通过各点特征相加实现多层特征融合,在特征提取中引入注意力机制实现密度权重的自适应分配,使网络充分学习点云数据及特征信息。针对大规模LiDAR点云,文献[125]设计了一种基线实例分割方法,使用密集特征编码技术对远处的小物体进行定位和分割,结合自我注意学习原始

点的特征表示,通过密集连接的骨干网预测实例标签。为了减少实例分割的后处理操作,3D-MPA<sup>[126]</sup>以对象为中心从稀疏体积骨干中学习特征,然后用图卷积网络细化,使相邻候选区域之间进行高阶交互,采用聚合特征的方法代替非最大抑制生成实例标签。

该方法能得到很好的实例标签,但是需要多阶段的训练,带来了额外的计算负担。

## 2) 基于免候选区域

基于免候选区域的方法不需要候选区域检测模块。相似组建议网络(similarity group proposal network, SGPN)<sup>[127]</sup>利用 PointNet 或 PointNet++ 提取点云的特征,然后在这些点云特征上计算相似矩阵和对应的置信度矩阵得到候选区域,在语义标记的基础上得到实例标签。为了提高分割方法对实例标签的关注度,文献[128]提出了用于密集三维体素的实例分割方法,对与单个对象有关的信息进行聚类以有效区分体素的实例标签。考虑到文献[128]中体素会导致量化误差,文献[129]应用基于注意力的 k 最近邻算法聚合邻域信息,将三维 CNN 作为骨干网络,借助结构感知损失函数同时考虑几何信息和嵌入信息并监督 3D 实例的嵌入。部分方法针对点的实例和语义标记之间的关联性将二者结合以提升分割性能,实例和语义关联分割(associatively segmenting instances and semantics, ASIS)<sup>[130]</sup>应用基线同时进行点云语义与实例分割,语义感知实例分割,实例的语义特征融合得到更准确的语义分割,两种分割任务合作实现精度的相互提升;文献[131]设计了一种多任务逐点网络,在预测点云语义标签的同时能够将点云转换为高维向量并聚类到实例中,利用多值条件随机场联合优化语义和实例标签。此外,一些方法将传统聚类和滤波方法与深度学习结合来优化实例分割效果。文献[132]设计了 OccuSeg 网络,借助多任务学习和尺度感知生成可变的占用信号和嵌入表示,将特征与空间嵌入结合起来指导聚类以进行实例分割;而 PointGroup 网络<sup>[133]</sup>利用空隙空间对每个点进行分组,并通过基于原始点和偏移点坐标的聚类方式分割;DyCo3D<sup>[134]</sup>则基于动态卷积进行点云实例分割,通过构建轻量级子网络对特定类别的上下文进行编码,使过滤器能够区分不同实例。

该方法虽然不需要选取候选区域,计算量减少,但是对象边界的检测模糊。在实例分割中结合语义分割结果可以提升分割性能,但是目前相关的研究还比较少。

## 3.3 零件分割

了解分割对象的各个零件结构是点云分类分割的基础。针对零件的多尺度分析问题,SyncSpecCNN<sup>[135]</sup>使用频谱卷积分割不规则的零件形状。由于体素数据存在信

息丢失,文献[136]通过空间密集提取模块提取细粒度零件特征,结合注意力机制融合特征。考虑到上述方法对相同语义标记零件的分割精度不高,Wang 等<sup>[137]</sup>提出了三维形状全卷积网络(shape fully convolutional networks, SFCN),通过 3 个几何特征进行训练,结合多标签图分割方法优化分割结果;而 ResSAnet<sup>[138]</sup>则利用基于几何基元的置换不变卷积来学习局部几何信息,通过融合多层次特征来提升零件分割的准确性。为了增强算法对噪声和采样密度的鲁棒性,PointSpherical<sup>[139]</sup>对每个采样点创建一个球形感受野,采用 1×1 CNN 提取初始特征,所有特征信息经池化汇总后再进行零件分割。

点云零件分割主要还是基于卷积神经网络来实现的,一方面要解决噪声和采样密度的干扰问题,另一方面要有效区分具有相同语义标签的零件部分。

## 3.4 全景分割

点云全景分割需要将点云中表示前景的点和表示背景的点区分开,前景点中一般包括人、车等目标对象,背景点中通常是天空和树木。目前大部分全景分割方法是通过结合语义分割和实例分割处理前景对象与背景,对背景区域进行语义标记的同时区分不同的对象。全景分割方法有 2 种:自上而下以及自下而上。

针对雷达点云的稀疏性、遮挡以及重投影问题,Sirohi 等<sup>[140]</sup>提出了自上而下的高效激光雷达泛光学分割(efficient LiDAR panoptic segmentation, EfficientLPS)架构,构建了获取上下文特征的尺度不变的语义头和基于 Mask R-CNN 变体的实例头,通过全景融合模块细化对象边界。为了提高点云全景分割的速度,大部分点云全景分割方法采用自下而上架构<sup>[141-145]</sup>。Gasperini 等<sup>[141]</sup>利用基于学习的聚类方法提取实例 ID,在后续分割时可以滤除这些 ID,减少了计算开销;与 Gasperini 等<sup>[141]</sup>的实例聚类方法不同,Panoptic-PolarNet<sup>[142]</sup>通过极地鸟瞰图的中心回归处理实例分割问题,利用无注释的边界框训练实例头避免了预测重叠问题;由于自动驾驶任务对实时性和计算效率要求较高,Li 等<sup>[143]</sup>提出了基于激光雷达的无提案分割方法 SMAC-Seg,使用稀疏多向注意力聚类动态聚合二维投影对象,引入质感感知排斥损失监督网络,减少每个对象集群与其相邻对象的混淆;Duerr 等<sup>[144]</sup>则设计了多视图全景分割框架,通过聚合范围视图和鸟瞰图特征并重复传播到基于点的网络,结合邻域注意模块实现邻域的快速搜索,分割实时性进一步提升;为了在大规模点云上执行全景分割任务,Yang<sup>[145]</sup>提出了一种分形多门混合专家全景分割网络,编码器中的门结构可以有效过滤点云和体素特征,解码器中的多门专家网络能够获得更有表现力的数据特征,但是针对小样本的分割性能有待改善。

自上而下的全景分割方法借助 Mask R-CNN 变体以两阶段方式进行全景分割,自下而上的方法学习语义标签并将其与类别未知的实例分割融合,难点在于语义和实例头部的重叠现象以及对象之间的区分。

### 3.5 点云分类分割方法小结

点云语义分割和实例分割的深度学习大多直接处理原始点云数据,研究的方法较多且性能优异,将实例

分割与语义分割结合可以进一步优化分割效果。零件分割方法的核心是 CNN,但是分割具有相同语义标签的点云零件的难度很大。现阶段的全景分割还是一种新分割任务,尚未得到很好的研究。

点云语义分割和点云实例分割任务的方法对比分别如表 3 和 4 所示。

表 3 点云语义分割任务的方法对比

Table 3 Comparison of methods for point cloud semantic segmentation task

类别	文献	具体方法	贡献	局限性
基于结构化点云	[71]	二维卷积网络,投影	利用二维卷积网络学习三维点云	最大池化会丢失大量信息
	[74]	在 CNN 中引入条件随机场	分割快速、稳定且精确	对失调噪声比较敏感
	[77]	体素表示的三维 CNN 网络	充分利用点云信息	体素网格的内存增加
	[79]	3D 全卷积神经网络	实现了三维空间的全卷积网络	难以区分颜色、位置相近的物体
基于特征学习	[81]	基于八叉树结构的网络	以低内存消耗处理密集输入	依赖体素边缘信息
	[83]	分层抽取特征	解决了非均匀点采样的问题	缺乏点与点之间的关联性
	[84]	结合了二维的尺度不变特征变换	分割精度得到提高	受点云无序性和稀疏性影响
	[88]	SCF 模块	具有通用性	存在部分错误分类
	[90]	参数连续卷积算子	可以学习关系可计算的数据结构	对纹理相似的物体会错误分类
	[91]	核点卷积	可以改变卷积核形状	整体网络的复杂性增加
基于原始点云	[101]	注意力机制与图卷积结合	提高对象边界的分割效果	难以区分颜色和形状相似的地形
	[104]	可学习内核	网络具有移位和尺度不变性	算法需要设置合理的邻域数目
	[105]	多尺度特征图卷积神经网络	具有高分割精度和旋转不变性	缺少真实数据的实验评估
	[107]	自适应节点模块	利用了局部几何信息	训练样本不平衡会影响模型性能
	[110]	多尺度区域关系特征分割网络	对人工特征的选择和设计依赖降低	算法对不稳定的上下文信息敏感
基于旋转性	[112]	密度感知自适应采样模块	能够感知密度	无法感知曲率变化
	[115]	多头注意力卷积层神经网络	变换容易,结构简单	存在高维度问题
	[116]	RIConv 算子	在不同旋转下具有高度一致性	卷积设计中缺乏低级几何特征
	[117]	基于点投影层次聚类的神经网络	有效地提取并整合点云特征	局部特征不是旋转不变的
	[118]	RotPredictor 模块	具有很强的稳健性	算法并非完全是旋转等变的

## 4 点云点云配准和匹配

点云配准需要找到能够使不同坐标系下的点云坐标正确匹配的位置变换关系。最经典的两种传统点云配准方法是迭代最近点(iterative closest point, ICP)和正态分布变换(normal distribution transform, NDT),处理特征明显的点云效果不错,但对噪声和遮挡敏感。最近,点云配准成功结合了深度学习方法,提高了配准精度和鲁棒性,是点云配准研究的新思路。深度学习的点云配准方法有 3 种:基于特征、端到端和无监督。

### 4.1 基于特征的方法

基于特征的配准方法使用网络估计两个输入点云的特征,然后迭代运行对应和变换估计来求得最终解。

一些研究人员重点利用局部特征描述符进行点云配准。3DMatch 方法<sup>[146]</sup>学习局部空间块的描述符建立 3D 数据的对应关系,采用自监督方法在现有 RGB-D 场景重建中训练相应的标签并学习特征;为了在网络训练中获取精确标注,3DFeat-Net<sup>[147]</sup>利用弱监督学习来捕获局部点云配准的特征描述符,并对 GPS/INS 标记的三维点云使用对齐和注意力机制来学习特征点匹配;针对匹配精度提升问题,点对特征网络(point pair feature network, PPFNet)<sup>[148]</sup>应用于邻域点分布的局部点对特征(point pair features, PPF),

表 4 点云实例分割任务的方法对比

Table 4 Comparison of methods for point cloud instance segmentation task

类别	文献	具体方法	贡献	局限性
基于 候选区域	[120]	生成形状建议网络	实例分割性能提高	后处理时间长
	[122]	空间散列体积图表示	实现了事物层面的场景理解	忽略了少于 100 个顶点的预测事物
	[123]	基于逐点多层感知机的网络	不需要进行后续处理,分割效率提高	未考虑点云密度变化
	[125]	密集特征编码技术	可以预测大规模雷达点云的实例标签	小物体的分割精度不高
基于 免候选区域	[127]	相似组建议网络	网络简单灵活	相似矩阵需要占用大量内存
	[129]	三维卷积神经网络	具有较好的扩展性	会误判几何形状相似的物体类别
	[130]	语义分割和实例分割联合	分割精度提升	距离很近的同类对象难以分割开
	[133]	通过目标间的空隙空间对点进行分组	具有通用性	存在实例分组的语义不准确性问题

并将其输入网络进行深度特征学习,但是旋转鲁棒性较差;为此,作者又提出了 PPF-FoldNet<sup>[149]</sup>,借助类似 FoldingNet 的端到端自动编码器训练四维 PPF 集合并自动重建 PPF;由于大多数 3D 特征描述符对旋转敏感,3DSmoothNet<sup>[150]</sup>借助体素化的平滑密度值表示输入数据,应用完全卷积层结合来学习局部特征,通过局部参考框架对齐来实现旋转不变性。考虑到先前的方法都采用特征描述符进行匹配,D3Feat<sup>[151]</sup>则应用 3D 特征检测器寻找匹配关键点,利用基于 KPConv 的 FCN 对三维点云进行局部特征检测,采用密度不变关键点选择策略获取三维点云重复点,通过自监督检测器损失更好地指导特征匹配。针对错误匹配的干扰问题,迭代距离感知相似矩阵卷积网络 (iterative distance-aware similarity matrix convolution network, IDAM)<sup>[152]</sup>将几何特征和距离特征合并,通过卷积层学习点匹配,使用两阶段点消除方法排除错误点对并降低了计算复杂度。

上述深度学习方法大多存在局部特征提取不足的问题,为了提高配准的精度,研究人员又提出了一些新方法。PACNet<sup>[153]</sup>引入卷积层 PACConv,根据每个点及其相邻点的位置关系学习权重系数和点云局部信息,整合关键点的局部相关性和全局信息;为了增强网络在未知场景的配准性能,Horache 等<sup>[154]</sup>设计了多尺度深度神经网络多尺度稀疏体素卷积 (multi-scale sparse voxel convolution, MS-SVConv),在不同尺度的点云上通过 3D 稀疏体素卷积网络计算特征,然后通过全连接层融合特征,在两个场景之间快速进行 3D 配准。针对点云部分重叠的情况,OMNet<sup>[155]</sup>通过学习重叠掩码去除非重叠区域,将局部到局部的匹配转换为相同形状的匹配。考虑到异常值干扰,文献[156]引入深度图匹配,在设置对应关系时通过节点特征和图结构有效去除异常值。

基于特征的方法结合了神经网络优势,在网络优化中预估姿态变换,点云配准灵活且计算复杂度降低。

## 4.2 基于端到端的方法

端到端的配准方法可分为 3 个部分:对应预测模块、权重预测模块和奇异值分解 (singular value decomposition, SVD) 模块,主要思想是将两个待配准的点云送入神经网络,使用端到端框架来估计这两个点云之间的变换矩阵。目前,涉及端到端的配准方法可以大致分为以下两大类。

第 1 类方法对经典的点云配准方法进行了改进,将传统数学理论与神经网络结合。深度最近点 (deep closest point, DCP)<sup>[157]</sup>是基于深度学习的单步点云配准方法,利用 DGCNN 学习特征,使用注意力机制学习待配准点云之间的对应关系和特征信息,通过 SVD 模块预测旋转并实现点云匹配。针对局部到局部点云配准问题,部分注册网络 (partial registration network, PRNet)<sup>[158]</sup>通过添加关键点检测寻找局部视图间的共同点来扩展 DCP 方法,通过深度网络解决部分对应问题;而 OPRNet<sup>[159]</sup>则将 DCP 网络的 softmax 替换为最佳传输 (optimal transport, OT) 层,利用两个点云的相似性度量来计算传输成本,通过 Sinkhorn 算法解决传输成本问题进而获得匹配关系。为了实现稳健、快速且准确的点云配准,深度全局注册 (deep global registration, DGR) 方法<sup>[160]</sup>应用网络预测权重,通过 6 维卷积网络为匹配点提供置信度估计,然后采用加权 SVD 求解相对位姿,再通过梯度优化器细化对齐。考虑到 DCP 方法<sup>[157]</sup>对未知类别点云的配准效果不佳,深度好友相似性学习 (deep best buddy similarity learning, DeepBBS) 方法<sup>[161]</sup>引入了最佳伙伴对来计算两个点云之间的 3D 旋转和平移,迭代计算直至收敛,最终实现准确的配准。由于点云配准对初始变换敏感,RPM-Net<sup>[162]</sup>使用可微 Sinkhorn 层和退火算法从混合特征中得到点对应的软分配矩阵,通过二次网络预估最优退火参数进而提升配准性能。

第 2 类方法则是利用全局特征差异进行配准。深层虚拟对应点 (deep virtual corresponding points,

DeepVCP)<sup>[163]</sup>是第1个采用端到端架构的点云配准框架,它先使用PointNet++从待配准的两个点云中获取每个点的语义特征并生成关键点,再通过局部相似性和全局几何约束的损失函数提高准确度。PointNetLK方法<sup>[164]</sup>结合PointNet与Lucas & Kanade算法进行配准,将点云映射成多维特征向量,再借助LK算法进行配准。由于PointNetLK<sup>[164]</sup>对噪声敏感,点云配准网络(point cloud registration network, PCRNet)<sup>[165]</sup>则使用PointNet从点云中提取全局特征,连接2个全局特征并采用5层全连接网络返回相对位姿,无需计算点云间的对应关系。然而上述方法并未考虑到噪声和异常值以及重复点云对配准精度的影响。特征度量配准(feature-metric registration, FMR)方法<sup>[166]</sup>通过最小化特征空间上的投影误差来估计两个输入点云之间的变换矩阵,确保了在噪声干扰下点之间对应关系的准确性;文献[167]则设计了具有旋转不变性和代表性的特征描述符,来提取鲁棒性更强的局部特征,捕获足够的信息实现准确的配准;而文献[168]针对稀疏和重复点云提出了多视点配准方法,直接学习以全局一致的方式匹配所有视图,利用置信估计块来预测成对配准参数的置信度,借助边缘权重的迭代更新来优化全局配准;PointDSC<sup>[169]</sup>则基于SM算法去除异常对应,采用空间一致性引导的非局部模块进行特征嵌入,通过可微分的谱匹配模块去除异常值。此外,为了解决混合特征提取导致的维度差异和语义差距问题,多特征引导网络(multi-features guidance network, MFGNet)<sup>[170]</sup>通过形状特征和空间坐标单独指导对应搜索,借助对应可信度计算模块减少了错误匹配的影响,进而实现可靠的点云配准。

端到端的方法通过训练神经网络直接解决三维点云的配准问题。SVD算法具有可微性,可以与神经网络结合,高效学习的同时也提高了匹配速度。

### 4.3 基于无监督的方法

无监督的点云配准方法不需要人工标注或类标签。无监督稳定兴趣点检测器(unsupervised stable interest point detector, USIP)方法<sup>[171]</sup>首先对输入点云进行随机旋转平移变换,通过特征提议网络(feature proposal network, FPN)学习输入点云中稳定的关键点,利用概率倒角损失和点对点损失提高关键点定位的精度以及点云匹配的准确度;文献[172]则设计了显著点分析(salient points analysis, SPA)方法,借助局部表面特性在两个点云中选择一组对应的显著点,学习每个点的特征并建立点对应关系,但对旋转角度大的点云配准效果较差;针对该问题,R-PointHop<sup>[173]</sup>利用每个点的最近邻确定点的局部参考帧(local reference frame, LRF),LRF可以使点的层次特征在大幅度旋转和平移时保持不变性,并在空间中建立可靠的点对应;为了保留点云的结构信息,VK-Net<sup>[174]</sup>从原始点云中学习关键点并建立对应关系,通过最小二乘拟合算法计算初始刚性变换,再用ICP细化配准结果。

与监督学习相比,无监督学习减少了点云配准的训练时间,缩小了模型尺寸,但是对旋转幅度大、噪声干扰的点云的配准效果还需要优化。

### 4.4 点云配准和匹配方法小结

基于深度学习的点云配准方法目前只能将2组点云粗略的转换到同一坐标系下,还不能直接得到满意的效果,同时噪声、异常值以及采样处理也会降低点云数据的精度。虽然深度学习方法在搜索对应关系方面具有一定的优势,但还需要进行更多更深入的研究。点云配准方法的具体对比如表5所示。

表5 点云配准任务主要方法对比

Table 5 Comparison of main methods for point cloud registration task

类别	文献	具体方法	贡献	局限性
基于特征	[146]	基于三维CNN的局部特征提取方法	在重构的局部特征匹配表现良好	需要大量训练数据,手动标记耗时且易出错
	[147]	弱监督学习和注意力机制	不需要手动标记点云匹配点集	忽略了点之间的空间关联,不适合嘈杂点云
	[148]	基于邻域点分布的局部点对特征	对稀疏点云具有鲁棒性	存在二次内存占用的问题,对旋转比较敏感
	[155]	基于全局特征的迭代网络	对噪声具有鲁棒性	未在重叠点中估计匹配候选点
基于端到端	[157]	单步点云配准方法	不需要迭代	以软方式预测对应关系,精确度不够高
	[162]	可微Sinkhorn层和退火算法	解决了异常点敏感的问题	每次迭代都要重新计算参数,非常耗时
	[163]	基于学习的关键点检测方法	对不准确的初始姿势有鲁棒性	依赖于高质量的初始化
	[164]	PointNet与LK算法结合	完全可微,易集成	占用大量存储空间,对噪声敏感
	[167]	具有旋转不变性的特征描述符	噪声鲁棒性增强	时间长,人为设置的参数较多
基于无监督	[171]	无监督点云关键点检测器	网络在旋转下也有良好的性能	重复性指标较低,还可以继续优化
	[172]	使用部分显著点配准两个点云	训练成本少,模型小	无法对齐两个彼此相关但旋转角度大的点云

## 5 点云拼接

在点云数据的采集过程中,由于受到三维传感器视角的限制,只能获取某个角度的点云数据而且存在对象遮挡问题,因此需要三维点云拼接技术来获取点云的完整信息。传统的点云拼接方法着重强调拼接初值的选取,故而点云拼接的深度学习方法也针对初值选取进行了改进,包含多视图和特征 2 类。

### 5.1 基于多视图的方法

基于多视图的点云拼接方法的核心是对多视角下的点云数据进行坐标变换。文献[175]采用你只需看一次(you only look once, YOLO)网络识别待拼接目标的边界框,借助位姿估计将坐标系统一到单目相机下以便进行初始拼接,再应用传统 ICP 算法优化拼接结果。考虑到传统视图拼接会出现误差累积,文献[176]

在结构光三维扫描过程中构建标记点多视图网络,通过光束法平差优化控制点的坐标信息,再在局部坐标系下与点云数据拼接。

### 5.2 基于特征的方法

基于特征的方法采用神经网络捕获点云的特征。为了更好地拼接文物碎片的断裂面,文献[177]利用 CNN 学习待拼接断裂面的全局和局部特征,根据关键点的局部特征获取 2 个待拼接碎片的对应关系,拼接点对的有效性和准确度大大提升。

### 5.3 点云拼接方法小结

基于多视图的点云拼接方法应用网络辅助多视图坐标的变换,基于特征的方法充分利用了神经网络的学习能力。拼接的准确度和效率有一定的提升,拼接误差相对其他传统方法有所降低。但是目前点云拼接方法仍以机器学习方法为主,点云拼接的网络模型还值得深入探索。点云拼接方法的对比分析如表 6 所示。

表 6 点云拼接任务主要方法对比

Table 6 Comparison of main methods for point cloud stitching task

类别	文献	具体方法	贡献	局限性
基于多视图	[175]	YOLO 网络,位姿估计	多视角拼接效率提高,适用场景更多	对前期训练信息的依赖更强
	[176]	构建标志点多视图网络	避免了传统拼接的误差累计问题	时间成本增加
基于特征	[177]	卷积神经网络	解决了碎片全局特征差异大但相互匹配的问题	未考虑碎片本身缺失的拼接情况

## 6 数据集及评价指标

### 6.1 数据集

数据集的好坏是检验点云分析方法性能的首要条件。本节整理了常用的 18 种数据集,给出了每种数据集适用的点云分析任务以及数据集网址,如表 7 所示。

### 6.2 评估指标

点云分析任务有很多种类,每一类任务包含多种方法。为了对这些方法的性能进行客观公正的评估,本节整理了常用的 4 类评价指标。

#### 1) 多目标跟踪度量

衡量多目标跟踪任务的性能指标包括多目标跟踪准确度(multiple object tracking accuracy, MOTA)和多目标跟踪精确度(multiple object tracking precision, MOTP),计算公式分别如下:

$$MOTA = 1 - \frac{\sum_i (m_i + fp_i + mme_i)}{\sum_i g_i} \quad (1)$$

$$MOTP = \frac{\sum_{i,t} d_i^t}{\sum_i c_i} \quad (2)$$

式中: $m_i$  为  $t$  帧时的漏检数量, $fp_i$  为  $t$  帧时的误检数量, $mme_i$  为  $t$  帧时的误匹配数量, $c_i$  是当前帧成功匹配的数量, $d_i^t$  表示目标与预测值的平均度量距离。

#### 2) 全景分割度量

全景分割度量包含全景质量、识别质量以及分割质量 3 方面指标。全景质量(panoptic quality, PQ)是衡量点云全景分割性能的指标;识别质量(recognition quality, RQ)反映全景分割中每个实例对象识别的准确性;分割质量(segmentation quality, SQ)表示匹配后预测部分与标注部分的平均交并比,公式分别为:

$$PQ = SQ \times RQ \quad (3)$$

$$SQ = \frac{\sum_{(p,g) \in TP} IoU(p,g)}{|TP|} \quad (4)$$

$$RQ = \frac{|TP|}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|} \quad (5)$$

式中: $IoU(p, g)$  是预测的分割  $p$  和真实值  $g$  的交并比, $TP$  是  $IoU > 0.5$  的分割结果, $FP$  为预测为正实际为负的分割结果, $FN$  为预测为负实际为正的分割结果。

表7 点云分析任务常用数据集

Table 7 Datasets commonly used in point cloud analysis task

数据集	时间	类别	点云任务类型	应用场景	网址
Oakland3D <sup>[178]</sup>	2009	5	分类分割	城市街道	<a href="http://www.cs.cmu.edu/~vmr/datasets/oakland_3D/cvpr09/doc/">http://www.cs.cmu.edu/~vmr/datasets/oakland_3D/cvpr09/doc/</a>
RGB-D Object <sup>[179]</sup>	2011	51	分类分割、拼接	室内场景	<a href="https://rgbd-dataset.cs.washington.edu/">https://rgbd-dataset.cs.washington.edu/</a>
KITTI <sup>[180]</sup>	2012	3	目标检测与跟踪、 分类分割、配准	自动驾驶	<a href="http://www.cvlibs.net/datasets/kitti">http://www.cvlibs.net/datasets/kitti</a>
ETH <sup>[181]</sup>	2012	8	目标检测与跟踪、 分类分割、配准	户外场景	<a href="http://projects.asl.ethz.ch/datasets">http://projects.asl.ethz.ch/datasets</a>
SUN3D <sup>[182]</sup>	2013	-	分类分割	室内场景	<a href="http://sun3d.cs.princeton.edu/">http://sun3d.cs.princeton.edu/</a>
ModelNet <sup>[183]</sup>	2015	-	分类分割、配准	多种	<a href="http://modelnet.cs.princeton.edu/">http://modelnet.cs.princeton.edu/</a>
ShapeNet <sup>[184]</sup>	2015	55	分类分割、配准	物体零件	<a href="https://shapenet.org/">https://shapenet.org/</a>
ShapeNet Part <sup>[185]</sup>	2016	16	零件分割	物体零件	<a href="https://cs.stanford.edu/~ericyi/project_page/part_annotation/">https://cs.stanford.edu/~ericyi/project_page/part_annotation/</a>
S3DIS <sup>[186]</sup>	2016	13	分类分割、配准	室内场景	<a href="http://buildingparser.stanford.edu/dataset.html">http://buildingparser.stanford.edu/dataset.html</a>
vKITTI <sup>[187]</sup>	2016	13	目标检测与跟踪、 分类分割	城市街道	<a href="https://europe.naverlabs.com/Research/Computer-Vision/Proxy-Virtual-Worlds/">https://europe.naverlabs.com/Research/Computer-Vision/Proxy-Virtual-Worlds/</a>
3DMatch <sup>[146]</sup>	2017	-	配准、拼接	室内场景	<a href="http://3Dmatch.cs.princeton.edu/#rgbd-reconstruction-datasets">http://3Dmatch.cs.princeton.edu/#rgbd-reconstruction-datasets</a>
Semantic3D <sup>[188]</sup>	2017	8	分类分割	城市街道	<a href="http://semantic3d.net/">http://semantic3d.net/</a>
ScanNet <sup>[189]</sup>	2017	21	分类分割	室内场景	<a href="http://www.scan-net.org/">http://www.scan-net.org/</a>
SemanticKITTI <sup>[190]</sup>	2019	25	分类分割	自动驾驶	<a href="http://semantic-kitti.org/">http://semantic-kitti.org/</a>
PartNet <sup>[191]</sup>	2019	24	分类分割	多种	<a href="http://kevinkaixu.net/projects/partnet.html">http://kevinkaixu.net/projects/partnet.html</a>
WHU-TLS <sup>[192]</sup>	2020	-	分类分割、配准	多种	<a href="http://3s.wlu.edu.cn/ybs/en/benchmark.htm">http://3s.wlu.edu.cn/ybs/en/benchmark.htm</a>
nuScenes <sup>[193]</sup>	2020	23	目标检测与跟踪、 分类分割	自动驾驶	<a href="https://nuscenes.org/">https://nuscenes.org/</a>
Waymo <sup>[194]</sup>	2020	4	目标检测与跟踪	自动驾驶	<a href="https://waymo.com/open/">https://waymo.com/open/</a>

### 3) 特征匹配度量

特征匹配度量包含内点比例 (inlier ratio, IR) 和特征匹配召回率 (feature match recall, FMR) 两个指标。

内点比例表示配准后正确点云对关系占所有点云对关系的比例,用于衡量点云对的配准性能,计算公式如下:

$$IR = \frac{1}{|\mathbf{M}_{i,j}^*|_{(p,q) \in \mathbf{M}_{i,j}^*}} \sum 1[\|\mathbf{R}^* \mathbf{p} + \mathbf{t}^* - \mathbf{q}\| < \tau_1] \quad (6)$$

式中:  $1[\ ]$  表示括号内为真时取 1, 反之则取 0;  $\mathbf{R}^*$  是真实旋转值,  $\mathbf{t}^*$  是真实平移值,  $\tau_1$  是最小距离误差阈值。

特征匹配召回率表示高置信度任务在总配准任务中的比例,用于衡量整个数据集中所有点云对的配准性能,公式为:

$$FMR = \frac{1}{K} \sum_{i=1}^K 1[R_{i_i} > \tau_2] \quad (7)$$

式中:  $K$  是需要配准的点云对数量,  $R_{i_i}$  表示第  $i$  个点云对的 IR,  $\tau_2$  表示 IR 的最小阈值。

### 4) 精确度和误差参数

精确度是点云分析性能的关键指标,其中最常用的两个指标是平均交并比 (mean intersection over union,

mIoU) 和总体精度 (overall accuracy, OA)。 $mIoU$  表示真实值和预测值集合的交集与并集的比,公式为:

$$mIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij} + \sum_{j=0}^k p_{ji} - p_{ii}} \quad (8)$$

式中: 加上背景共有  $k+1$  个分类,  $p_{ij}$  表示真实值为  $i$  类, 预测值为  $j$  类的点的数目。

总体精度是模型预测所有测试集中正确点数和整体点数之比,公式为:

$$OA = \frac{\sum_{i=0}^k p_{ii}}{\sum_{i=0}^k \sum_{n=0}^k p_{in}} \quad (9)$$

除了上述参数外,误差评估参数也能反映算法性能。均方根误差 (root mean square error, RMSE) 与平均绝对误差 (mean absolute error, MAE) 的公式分别如下:

$$RMSE = \sqrt{\frac{1}{|\mathbf{M}_{i,j}^*|_{(p,q) \in \mathbf{M}_{i,j}^*}} \sum \|\mathbf{R} \mathbf{p} + \mathbf{t} - \mathbf{q}\|^2} \quad (10)$$

$$MAE = \frac{1}{|M_{i,j}^*|} \sum_{(p,q) \in M_{i,j}^*} |Rp + t - q| \quad (11)$$

式中:  $R$  和  $t$  是估计的刚性变换,  $(p, q) \in M_{i,j}^*$  表示真实的对应关系。

相对平移误差 (relative translation error, RTE) 表示平移估计值与真实值的差值, 相对旋转误差 (relative rotation error, RRE) 表示旋转估计值与真实值的差值, 公式分别如下:

$$RTE = \|t - t^*\|_2 \quad (12)$$

$$RRE = \arccos\left(\frac{\text{tr}(R^T R^*) - 1}{2}\right) \quad (13)$$

式中:  $\text{tr}(\quad)$  表示矩阵的迹。

配准召回率 (registration recall, RR) 是指误差小于给定阈值的点云对占所有点云对的比重, 公式为:

$$RR = \frac{1}{C} \sum_{i=1}^c 1[RMSE < \tau_3] \quad (14)$$

倒角距离 (chamfer distance, CD) 可以计算生成点云数据和标记点云数据之间的平均最短点距离<sup>[14]</sup>, 公式

如下:

$$CD = \frac{1}{|X|} \sum_{x \in X} \min_{y \in Y_c} \|x - y\|^2 + \frac{1}{|Y|} \sum_{y \in Y} \min_{x \in X_c} \|x - y\|^2 \quad (15)$$

式中:  $CD$  越小表明点云对配准越好, 重建效果越好;  $C$  是数据集内用于配准的点云对的数量,  $\tau_3$  是点云对的均方根误差阈值。

### 6.3 性能对比

考虑到点云拼接的深度学习的方法很少且数据集不同, 难以进行实验对比。因此, 本文仅整理了点云目标检测与跟踪、点云分类分割以及点云配准和匹配的深度学习方法的相关实验数据, 并进行了对比分析。

KITTI 数据集下点云目标检测深度学习方法的性能对比如表 8 所示, 指标为平均精度, 单位为 [%], “-” 表示该方法没有给出相应结果。与原始点云和混合点云相比, 结构化点云在汽车、行人和自行车 3 种类别的总体检测精度最低, 其中投影方法比体素方法的效果更差; 基于原始点云的检测方法在汽车的困难类别中最低也能达

表 8 点云目标检测任务在 KITTI 数据集上的性能对比

Table 8 Performance comparison of point cloud object detection task on KITTI dataset

类别	方法	汽车			行人			自行车			
		简单	中等	困难	简单	中等	困难	简单	中等	困难	
结构化点云	投影	RT3D <sup>[19]</sup>	23.74	19.14	18.86	-	-	-	-	-	-
		BirdNet <sup>[21]</sup>	40.99	27.26	25.32	22.04	17.08	15.82	43.98	30.25	27.21
		BirdNet+ <sup>[22]</sup>	70.14	51.85	50.03	37.99	31.46	29.46	67.38	47.72	42.89
	体素	VoxelNet <sup>[28]</sup>	77.47	65.11	57.73	39.48	33.69	31.51	61.22	48.36	44.37
		SECOND <sup>[29]</sup>	83.34	72.55	65.82	48.96	38.78	34.91	71.33	52.08	45.83
原始点云	PointNet	PointPillars <sup>[33]</sup>	82.58	74.31	68.99	51.45	41.92	38.89	77.10	58.65	51.92
		PointRCNN <sup>[34]</sup>	86.96	75.64	70.70	47.98	39.37	36.01	74.96	58.82	52.53
		STD <sup>[35]</sup>	87.95	79.71	75.09	42.47	53.29	38.35	78.69	61.59	55.30
	注意力机制	IA-SSD (single) <sup>[38]</sup>	88.87	80.32	75.10	47.90	41.03	37.98	82.36	66.25	59.70
	TANet <sup>[39]</sup>	84.39	75.94	68.82	53.72	44.34	40.49	75.70	59.44	52.53	
神经网络	PointRGCN <sup>[41]</sup>	85.97	75.73	70.60	-	-	-	-	-	-	
	Point-GNN <sup>[42]</sup>	88.33	79.47	72.29	51.92	43.77	40.14	78.60	63.48	57.08	
点云-体素	Fast point R-CNN <sup>[46]</sup>	85.29	77.40	70.24	-	-	-	-	-	-	
	PV-RCNN <sup>[47]</sup>	90.25	81.43	76.82	52.17	43.29	40.29	52.17	43.29	40.29	
	MV3D <sup>[49]</sup>	74.97	63.63	54.00	-	-	-	-	-	-	
混合点云	AVOD <sup>[50]</sup>	76.39	66.47	60.23	36.10	27.86	25.76	57.19	42.08	38.29	
	SCANet <sup>[51]</sup>	79.22	67.13	60.65	-	-	-	-	-	-	
	点云-图像	PI-RCNN <sup>[52]</sup>	84.37	74.82	70.03	-	-	-	-	-	-
	Frustum PointNets <sup>[55]</sup>	82.19	69.79	60.59	50.53	42.15	38.08	72.27	56.12	49.01	
	PointFusion <sup>[56]</sup>	77.92	63.00	53.27	33.36	28.04	23.38	49.34	29.42	26.98	
Frustum ConvNet <sup>[58]</sup>	87.36	76.39	66.69	52.16	43.38	38.80	81.98	65.07	56.54		



到 68.82% 的准确度,在 3 种点云形式中效果最佳;混合点云方法除 PV-RCNN<sup>[47]</sup> 外,检测性能并没有大幅提升,主要原因有以下 4 点:1) 混合点云数据网络的参数增加,模型易过拟合;2) 不同数据的特征信息融合困难,信息未必得到合理利用;3) 不同的数据类型都有各自对应的网络结构,优化网络的设计难度比较大;4) 三维目标检测的搜索空间过大,在时间及计算成本的限制下难以找到优异模型结构和参数。

不同点云语义分割方法在典型数据集下的性能对比结果如表 9 所示。基于体素的语义分割方法整体优于多视图方法,基于原始点云的方法基本优于结构化点云的方法。基于特征学习和卷积方法在多个数据集上明显提高了点云语义分割的准确性,无监督方法在 ModelNet40 数据集上达到了最高的 OA 精度(93.4%),尽管基于旋转性的方法并未有效改善分割精度,实验结果证明了旋转性辅助点云语义分割的可行性。

表 9 不同点云语义分割方法的性能对比

Table 9 Performance comparison of different point cloud semantic segmentation methods

类别	方法	ModelNet40		ShapeNet		S3DIS		Semantic3D		KITTI
		OA/%	OA/%	mIoU	OA/%	mIoU	OA/%	mIoU	mIoU	
结构化点云	多视图	MVCNN <sup>[71]</sup>	90.1	-	-	-	-	-	-	-
		SnapNet <sup>[72]</sup>	-	-	-	-	-	88.6	59.1	-
	体素	SEGCloud <sup>[79]</sup>	-	-	-	-	48.92	-	60.3	36.78
		PointGrid <sup>[80]</sup>	92.0	86.1	86.4	-	-	-	-	-
		3DContextNet <sup>[82]</sup>	-	-	-	84.90	55.60	-	-	-
特征学习	PointNet <sup>[18]</sup>	89.2	-	83.7	78.62	47.71	-	-	14.60	
	PointNet++ <sup>[83]</sup>	91.9	-	85.1	80.10	54.50	-	63.1	20.10	
	PointWeb <sup>[86]</sup>	92.3	-	-	86.97	60.28	-	-	-	
	RandLA-Net <sup>[87]</sup>	-	-	-	-	-	94.8	77.4	53.90	
	SCF-Net <sup>[88]</sup>	-	-	-	88.40	71.60	94.7	77.6	-	
原始点云	PointCNN <sup>[89]</sup>	92.5	-	84.6	-	65.39	-	-	-	
	KPconv <sup>[91]</sup>	92.9	-	86.4	-	70.60	-	74.6	-	
	ShellNet <sup>[94]</sup>	-	-	-	87.10	66.80	-	69.4	-	
	卷积	PointConv <sup>[95]</sup>	92.5	-	85.7	-	-	-	-	-
		RGCNN <sup>[99]</sup>	90.5	-	84.3	-	-	-	-	-
		DGCNN <sup>[100]</sup>	92.2	-	85.1	-	-	-	-	-
		GACNet <sup>[101]</sup>	-	-	-	87.79	62.85	91.9	70.8	-
无监督	SO-Net <sup>[106]</sup>	93.4	-	84.6	-	-	-	-	-	
	MS-RRFSegNet <sup>[110]</sup>	-	-	-	-	-	93.0	72.5	-	
旋转性	文献 <sup>[116]</sup>	86.5	-	75.5	-	-	-	-	-	

点云全景分割任务在 SemanticKITTI 和 nuScenes 数据集下的性能对比如表 10 所示。其中, S 和 N 分别代表 SemanticKITTI 和 nuScenes 数据集; PQ 是全景质量, PQ<sup>†</sup> 是改进的全景质量, SQ 是分割质量, RQ 是识别质量, 所有分数单位都为 [%]; St 表示 Stuff(道路和植被), Th 表示 Thing(行人和汽车)。目前深度学习方法的 PQ 最高只有 67.7, mIoU 基本维持在 60~70 之间, 虽然点云全景分割难度较大, 但可提升空间很大。

3DMatch 和 ModelNet40 数据集下点云配准深度学习方法的性能对比如表 11 所示, “-” 表示该方法没有给出

相应结果。DGR 方法利用网络预测点对置信度, RR 指标最高; 而 OMNet 主要针对部分重叠的点云配准, RR 指标最低; RPM-Net 利用软匹配方法搜索最佳对应关系, 鲁棒性较强, 但表 11 中其他端到端的配准方法对旋转敏感。

## 7 展 望

近些年来, 得益于计算设备和传感器技术的不断提高, 基于深度学习的三维点云分析方法发展迅速, 取得了

表 10 不同点云全景分割方法的性能对比

Table 10 Performance comparison of different point cloud panoptic segmentation methods

方法	数据集	$PQ$	$PQ^\dagger$	$RQ$	$SQ$	$PQ^{Th}$	$RQ^{Th}$	$SQ^{Th}$	$PQ^{St}$	$RQ^{St}$	$SQ^{St}$	$mIoU$
EfficientLPS <sup>[140]</sup>	S	57.4	63.2	68.7	83.0	53.1	60.5	87.8	60.5	74.6	79.5	61.4
	N	59.2	62.8	70.7	82.9	51.8	62.7	80.6	71.5	84.1	84.3	69.4
Panoster <sup>[141]</sup>	S	52.7	59.9	64.1	80.7	49.4	58.5	83.3	55.1	68.2	78.8	59.9
Panoptic-PolarNet <sup>[142]</sup>	S	54.1	60.7	65.0	81.4	53.3	60.6	87.2	54.8	68.1	77.2	59.5
	N	67.7	71.0	78.1	86.0	65.2	74.0	87.2	71.9	84.9	83.9	69.3
RangeBird <sup>[144]</sup>	S	58.9	64.8	69.4	84.0	58.2	64.9	88.9	59.5	72.7	80.5	66.8
	N	61.2	64.7	72.5	83.8	54.3	64.6	82.5	72.5	84.7	84.8	74.0
FMP-Net <sup>[145]</sup>	N	40.7	46.0	49.1	76.0	26.7	32.2	72.7	64.1	77.3	81.8	62.3

表 11 不同点云配准方法的性能对比

Table 11 Performance comparison of different point cloud registration methods

类别	方法	3DMatch	ModelNet40	
		$RR/\%$	$RRE/(\circ)$	$RTE/m$
特征	3DMatch <sup>[146]</sup>	67.0	-	-
	PPFNet <sup>[148]</sup>	71.0	-	-
	3DSmoothNet <sup>[150]</sup>	78.4	-	-
	D3Feat <sup>[151]</sup>	81.6	-	-
	OMNet <sup>[155]</sup>	35.9	2.947	0.032
端到端	DCP <sup>[157]</sup>	-	11.975	0.171
	DGR <sup>[160]</sup>	85.3	-	-
	RPM-Net <sup>[162]</sup>	-	1.712	0.018
	PointNetLK <sup>[164]</sup>	-	29.725	0.290

许多优异成果,但仍有一定的局限性,可以考虑从以下几个方面深入研究。

1) 点云数据融合。由于点云的无序性和不规则性,传统的多视图和体素化点云分析方法存在转换信息丢失、计算量大和内存占用的问题。而基于原始点云的方法不需要任何转换,是目前重要的研究方向。但是直接处理点云的方法无法充分利用点云的空间结构信息,还需要搜索邻域信息,会增加额外的计算量。点云数据融合是一种新颖的研究方向,将体素或多视图的纹理特征与原始点云的三维几何特征融合,使二维与三维信息的优势互补,对点云分析研究有重要意义。

2) 点云旋转变换。在数据采集过程中,设备的放置角度会造成场景内物体的遮挡。点云进行旋转变换后虽然形状不变,但坐标发生改变。遮挡和旋转都会给点云分析任务带来极大的干扰。能否通过对点云进行旋转来降低物体之间的遮挡程度,进而实现更好的点云分析性能是一个值得思考的新方向。

3) 轻量化的网络模型。自动驾驶等领域对点云分析的实时性要求较高,大多数方法提高了点云分析的精度,但是模型变得更复杂,效率降低,难以得到实时的处理结果。为了降低计算成本,可以尝试构建轻量化的网络模型,实现精度和实时性共存。

4) 搜索密集点云区域。三维点云空间分布稀疏不均,稀疏空间中有效的特征信息也很少,直接处理点云时很难判断有价值的点云信息所在的区域。在点云分析前搜索密集点云区域并高效地提取点云信息是一个很有价值的研究方向。

5) 多种点云分析技术结合。在点云实际分析过程中,每种任务并非完全独立进行的。点云拼接需要配准技术支撑,全景分割技术结合了语义和实例分割。场景流估计可以从动态场景获取丰富的特征信息,将场景流估计与目标检测、跟踪技术结合有望提升点云分析的精度。多种技术结合是未来点云分析的一种趋势,综合了各种技术的优势,能够实现更高效的性能,同时减少了深度学习方法的计算负担。

6) 无监督学习。无监督学习方法对数据标记要求很低,可以利用大量无标记的点云数据进行处理。将无监督学习与深度学习方法结合,能够减少数据标记的时间,同时避免错误标记数据的影响。目前无监督学习的应用相对较少,在点云分析速度提升方面有很大的发展空间。

7) 点云分析精度。目前,仅采用基于深度学习的点云分析方法难以精确分割复杂物体的边界,检测被遮挡的困难目标的准确率较低,对部分重叠点云的配准精度也未得到明显提升。可以考虑将传统方法与深度学习方法结合,如传统特征描述符和检测符、滤波、采样等辅助三维点云分析任务,也可以融合点云自身的几何与属性等信息提升分析精度。

8) 数据集及其评价指标。数据集的质量关乎模型训练和预测效果。大多数点云分析方法仅能改善某个

方面的缺陷,在特定的数据集上取得良好效果,对新获取的数据并不适用。因此,如何构建一个丰富、有效且适用性强的数据集是点云分析首先要解决的问题。目前各种点云分析任务的评价标准尚未统一,为了对比各类点云分析方法的性能,还需要进一步探索统一的数据集评价指标。

9) 硬件实现和应用。目前的点云分析任务大多依赖于单一传感器,获取的数据信息不够全面,多传感器融合以获取全方位的点云信息是今后硬件实现的一个重要方向。点云分析任务在自动驾驶、机器人、遥感、AR/VR等领域得到了广泛应用,此外在文物修复、工业农业安全生产、医疗以及电力巡检等方面也得到了推广。随着硬件技术的不断发展,未来有望实现更多服务于人们的日常生活的大众化应用。

## 8 结 论

在自动驾驶、机器人和遥感等领域,点云分析任务的应用日益增多。相比于传统的机器学习方法,采用深度学习分析方法分析点云的效率得到了提高,算法性能得到改善。本文从点云目标检测与跟踪、点云分类分割、点云配准和匹配以及点云拼接 4 种任务全面回顾了基于深度学习的三维点云分析方法的国内外研究进展。介绍了 18 种常用数据集和网址,给出了点云分析的评价指标和性能对比结果。针对目前点云分析存在的问题,对未来进一步的研究工作进行了展望。通过本文的总结和分析,期望能为三维点云分析研究提供参考。

### 参考文献

[ 1 ] GUO Y, BENNAMOUN M, SOHEL F, et al. 3D object recognition in cluttered scenes with local surface features: A survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(11): 2270-2287.

[ 2 ] 郑少武,李巍华,胡坚耀. 基于激光点云与图像信息融合的交通环境车辆检测[J]. 仪器仪表学报, 2019, 40(12): 143-151.

ZHENG SH W, LI W H, HU J Y. Vehicle detection in the traffic environment based on the fusion of laser point cloud and image information [J]. Chinese Journal of Scientific Instrument, 2019, 40(12): 143-151.

[ 3 ] 余洪山,付强,孙健,等. 面向室内移动机器人的改进 3D-NDT 点云配准算法[J]. 仪器仪表学报, 2019, 40(9): 151-161.

YU H SH, FU Q, SUN J, et al. Improved 3D-NDT point cloud registration algorithm for indoor mobile robot[J]. Chinese Journal of Scientific Instrument, 2019, 40(9): 151-161.

[ 4 ] ESPINOSA N, LENZ A, GROSS W, et al. Towards fast 3D reconstruction of urban areas from aerial nadir images for a near real-time remote sensing system[C]. IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, 2018: 6472-6475.

[ 5 ] LIU W, LAI B, WANG C, et al. Learning to match 2D images and 3D LiDAR point clouds for outdoor augmented reality[C]. 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), 2020: 654-655.

[ 6 ] WIRTH F, QUEHL J, OTA J, et al. PointAtMe: Efficient 3D point cloud labeling in virtual reality[C]. 2019 IEEE Intelligent Vehicles Symposium, 2019: 1693-1698.

[ 7 ] 褚彤. 基于多尺度点云和表面纹理特征的兵马俑修复方法研究与应用[D]. 西安:西北大学, 2021.

CHU T. Research and application on repairing method of terracotta warriors amy based on multi-scale point clouds and surface texture features [D]. Xi' an: Northwest University, 2021.

[ 8 ] 司垒,谭超,朱嘉皓,等. 基于 X 射线图像和激光点云的煤矸识别方法[J]. 仪器仪表学报, 2022, 43(9): 193-205.

SI L, TAN CH, ZHU J H, et al. A coal-gangue recognition method based on X-ray image and laser point cloud [J]. Chinese Journal of Scientific Instrument, 2022, 43(9): 193-205.

[ 9 ] HOSOKI D, LU H, KIM H, et al. Detection of facial symmetric plane based on registration of 3D point cloud[C]. 2019 19th International Conference on Control, Automation and Systems ( ICCAS ), 2019: 1037-1041.

[ 10 ] 陈驰,麦晓明,宋爽,等. 机载激光点云数据中电力线自动提取方法[J]. 武汉大学学报(信息科学版), 2015, 40(12): 1600-1605.

CHEN CH, MAI X M, SONG SH, et al. Automatic power lines extraction method from airborne LiDAR point cloud[J]. Geomatics and Information Science of Wuhan University, 2015, 40(12): 1600-1605.

[ 11 ] 闫利,魏峰. 利用密集匹配点云的建筑单体提取算法研究[J]. 中国激光, 2018, 45(7): 270-277.

YAN L, WEI F. Single part of building extraction from dense matching point cloud [J]. Chinese Journal of Lasers, 2018, 45(7): 270-277.

- [12] 王亚东,田永林,李国强,等. 基于卷积神经网络的三维目标检测研究综述[J]. 模式识别与人工智能, 2021,34(12):1103-1119.  
WANG Y D, TIAN Y L, LI G Q, et al. 3D object detection based on convolutional neural networks: A Survey [J]. Pattern Recognition and Artificial Intelligence, 2021, 34(12): 1103-1119.
- [13] 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.  
LU J, JIA X R, ZHOU J, et al. A review of deep learning based on 3D point cloud segmentation [J]. Control and Decision, 2023, 38(3): 595-611.
- [14] 秦红星,刘镇涛,谭博元. 深度学习刚性点云配准前沿进展[J]. 中国图象图形学报,2022,27(2):329-348.  
QIN H X, LIU ZH T, TAN B Y. Review on deep learning rigid point cloud registration [J]. Journal of Image and Graphics, 2022, 27(2): 329-348.
- [15] GUO Y, WANG H, HU Q, et al. Deep learning for 3D point clouds: A survey [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 43(12): 4338-4364.
- [16] BELLO S A, YU S, WANG C, et al. Deep learning on 3D point clouds[J]. Remote Sensing, 2020, 12(11): 1729.
- [17] 李娇娇,孙红岩,董雨,等. 基于深度学习的3维点云处理综述[J]. 计算机研究与发展,2022,59(5):1160-1179.  
LI J J, SUN H Y, DONG Y, et al. Survey of 3-Dimensional point cloud processing based on deep learning [J]. Journal of Computer Research and Development, 2022, 59(5): 1160-1179.
- [18] QI C R, 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, 2017: 77-85.
- [19] ZENG Y, HU Y, LIU S, et al. RT3D: Real-time 3D vehicle detection in LiDAR point cloud for autonomous driving [J]. IEEE Robotics and Automation Letters, 2018, 3(4): 3434-3440.
- [20] YANG B, LUO W, URTASUN R. PIXOR: Real-time 3D object detection from point clouds[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7652-7660.
- [21] BELTRÁN J, GUINDEL C, MORENO F, et al. BirdNet: A 3D object detection framework from LiDAR information[C]. 2018 21st International Conference on Intelligent Transportation Systems, 2018: 3517-3523.
- [22] BARRERA A, GUINDEL C, BELTRÁN J, et al. BirdNet+: End-to-end 3D object detection in LiDAR bird's eye view [C]. 2020 IEEE 23rd International Conference on Intelligent Transportation Systems, 2020: 1-6.
- [23] ZHOU J, TAN X, SHAO Z, et al. FVNet: 3D front-view proposal generation for real-time object detection from point clouds[C]. 2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), 2019: 1-8.
- [24] HAMDÍ A, GIANCOLA S, GHANEM B. MVTN: Multi-view transformation network for 3D shape recognition [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, 2021: 1-11.
- [25] LI B, ZHANG T, XIA T. Vehicle detection from 3D LiDAR using fully convolutional network [J]. ArXiv Preprint, 2016, ArXiv: 1608.07916.
- [26] LI B. 3D fully convolutional network for vehicle detection in point cloud [C]. 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2017: 1513-1518.
- [27] ENGELCKE M, RAO D, WANG D Z, et al. Vote3Deep: Fast object detection in 3D point clouds using efficient convolutional neural networks[C]. 2017 IEEE International Conference on Robotics and Automation (ICRA), 2017: 1355-1361.
- [28] 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, 2018: 4490-4499.
- [29] YAN Y, MAO Y, LI B. SECOND: Sparsely embedded convolutional detection [J]. Sensors, 2018, 18(10): 3337.
- [30] YIN T, ZHOU X, KRAHENBUHL P. Center-based 3D object detection and tracking [C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 11784-11793.
- [31] 赵毅强,艾西丁·艾克白尔,陈瑞,等. 基于体素化图卷积网络的三维点云目标检测方法[J]. 红外与激光工程,2021,50(10):281-289.  
ZHAO Y Q, ARXIDIN A, CHEN R, et al. 3D point cloud object detection method in view of voxel based on

- graph convolution network [ J ]. Infrared and Laser Engineering, 2021, 50(10) : 281-289.
- [ 32 ] 李梢,李垚辰,张玉龙,等.采用稀疏 3D 卷积的单阶段点云三维目标检测方法[ J ].西安交通大学学报, 2022,56(9) :112-122.  
LI Q, LI Y CH, ZHANG Y L, et al. A single-stage point cloud 3D object detection method using sparse 3D convolution[ J ]. Journal of Xi'an Jiaotong University, 2022, 56(9) : 112-122.
- [ 33 ] LANG A H, VORA S, CAESAR H, et al. PointPillars: Fast encoders for object detection from point clouds[ C ]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 12697-12705.
- [ 34 ] SHI S, WANG X, LI H. PointRCNN: 3D object proposal generation and detection from point cloud[ C ]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 770-779.
- [ 35 ] YANG Z, SUN Y, LIU S, et al. STD: Sparse-to-dense 3D object detector for point cloud[ C ]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 1951-1960.
- [ 36 ] QI C R, LITANY O, HE K, et al. Deep hough voting for 3D object detection in point clouds[ C ]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 9277-9286.
- [ 37 ] QI C R, CHEN X, LITANY O, et al. ImVoteNet: Boosting 3D object detection in point clouds with image votes[ C ]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 4404-4413.
- [ 38 ] ZHANG Y, HU Q, XU G, et al. Not all points are equal: Learning highly efficient point-based detectors for 3D LiDAR point clouds [ C ]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 18953-18962.
- [ 39 ] LIU Z, ZHAO X, HUANG T, et al. TANet: Robust 3D object detection from point clouds with triple attention[ C ]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7) : 11677-11684.
- [ 40 ] XIE Q, LAI Y K, WU J, et al. MLCVNet: Multi-level context votenet for 3D object detection[ C ]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10447-10456.
- [ 41 ] ZARZAR J, GIANCOLA S, GHANEM B. PointRGCN: Graph convolution networks for 3D vehicles detection refinement[ J ]. ArXiv Preprint, 2019, ArXiv: 1911.12236.
- [ 42 ] SHI W, RAJKUMAR R. Point-GNN: Graph neural network for 3D object detection in a point cloud[ C ]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 1711-1719.
- [ 43 ] XIE S, GU J, GUO D, et al. PointContrast: Unsupervised pre-training for 3D point cloud understanding [ C ]. Proceedings of the European Conference on Computer Vision, 2020: 574-591.
- [ 44 ] QIN Z, WANG J, LU Y. Weakly supervised 3D object detection from point clouds[ C ]. Proceedings of the 28th ACM International Conference on Multimedia, 2020: 4144-4152.
- [ 45 ] XU X, WANG Y, ZHENG Y, et al. Back to reality: Weakly-supervised 3D object detection with shape-guided label enhancement[ C ]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 8438-8447.
- [ 46 ] CHEN Y, LIU S, SHEN X, et al. Fast point R-CNN[ C ]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 9775-9784.
- [ 47 ] SHI S, GUO C, JIANG L, et al. PV-RCNN: Point-voxel feature set abstraction for 3D object detection[ C ]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10529-10538.
- [ 48 ] 周锋,陶重犇,张祖峰,等.体素点云融合的三维动态目标检测算法[ J ].计算机辅助设计与图形学学报, 2022,34(6) :901-912.  
ZHOU F, TAO CH B, ZHANG Z F, et al. 3D dynamic target detection algorithm based on voxel point cloud fusion [ J ]. Journal of Computer-Aided Design & Computer Graphics, 2022, 34(6) : 901-912.
- [ 49 ] CHEN X, MA H, WAN J, et al. Multi-view 3D object detection network for autonomous driving [ C ]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1907-1915.
- [ 50 ] KU J, MOZIFIAN M, LEE J, et al. Joint 3D proposal generation and object detection from view aggregation[ C ]. 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2018: 1-8.
- [ 51 ] LU H, CHEN X, ZHANG G, et al. SCANet: Spatial-channel attention network for 3D object detection[ C ]. ICASSP 2019—2019 IEEE International Conference on Acoustics, Speech and Signal Processing ( ICASSP ), 2019: 1992-1996.
- [ 52 ] XIE L, XIANG C, YU Z, et al. PI-RCNN: An efficient

- multi-sensor 3D object detector with point-based attentive cont-conv fusion module[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12460-12467.
- [53] CHEN C, FRAGONARA L Z, TSOURDOS A. RoIFusion: 3D object detection from LiDAR and vision[J]. IEEE Access, 2021, 9: 51710-51721.
- [54] 冯明驰,高小倩,汪静姝,等. 基于立体视觉与激光雷达的车辆目标外形位置融合算法研究[J]. 仪器仪表学报,2021,42(10):210-220.
- FENG M CH, GAO X Q, WANG J SH, et al. Research on the fusion algorithm of vehicle object shape-position based on stereo vision and lidar[J]. Chinese Journal of Scientific Instrument, 2021, 42(10): 210-220.
- [55] 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, 2018: 918-927.
- [56] XU D, ANGUELOV D, JAIN A. PointFusion: Deep sensor fusion for 3D bounding box estimation [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 244-253.
- [57] ZHAO X, LIU Z, HU R, et al. 3D object detection using scale invariant and feature reweighting networks[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1): 9267-9274.
- [58] WANG Z, JIA K. Frustum convnet: Sliding frustums to aggregate local point-wise features for amodal 3D object detection[C]. 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2019: 1742-1749.
- [59] WEI Y, SU S, LU J, et al. FGR: Frustum-aware geometric reasoning for weakly supervised 3D vehicle detection[C]. 2021 IEEE International Conference on Robotics and Automation (ICRA), 2021: 4348-4354.
- [60] 黄远宪,李必军,黄琦,等. 融合相机与激光雷达的目标检测、跟踪与预测[J/OL]. 武汉大学学报(信息科学版):1-8[2023-05-04] <https://kns.cnki.net/kcms/detail/42.1676.TN.20220712.1009.001.html>.
- HUANG Y X, LI B J, HUANG Q, et al. Camera-LiDAR fusion for object detection, tracking and prediction[J/OL]. Geomatics and Information Science of Wuhan University:1-8 [2023-05-04], <https://kns.cnki.net/kcms/detail/42.1676.TN.20220712.1009.001.html>.
- [61] FROSSARD D, URTASUN R. End-to-end learning of multi-sensor 3D tracking by detection[C]. 2018 IEEE International Conference on Robotics and Automation (ICRA), 2018: 635-642.
- [62] GIANCOLA S, ZARZAR J, GHANEM B. Leveraging shape completion for 3D siamese tracking[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 1359-1368.
- [63] SIMON M, AMENDE K, KRAUS A, et al. Complexer-YOLO: Real-time 3D object detection and tracking on semantic point clouds[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019: 1-10.
- [64] WANG S, SUN Y, LIU C, et al. PointTrackNet: An end-to-end network for 3-D object detection and tracking from point clouds[J]. IEEE Robotics and Automation Letters, 2020, 5(2): 3206-3212.
- [65] 周笑宇,王玲,马燕新,等. 融合附加神经网络的激光雷达点云单目标跟踪[J]. 中国激光,2021,48(21): 158-170.
- ZHOU X Y, WANG L, MA Y X, et al. Single object tracking of LiDAR point cloud combined with auxiliary deep neural network[J]. Chinese Journal of Lasers, 2021, 48(21): 158-170.
- [66] GU X, WANG Y, WU C, et al. HPLFlowNet: Hierarchical permutohedral lattice flownet for scene flow estimation on large-scale point clouds[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 3254-3263.
- [67] LI R, LIN G, XIE L. Self-point-flow: Self-supervised scene flow estimation from point clouds with optimal transport and random walk [C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 15577-15586.
- [68] WEI Y, WANG Z, RAO Y, et al. PV-RAFT: Point-voxel correlation fields for scene flow estimation of point clouds[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 6954-6963.
- [69] OUYANG B, RAVIV D. Occlusion guided scene flow estimation on 3D point clouds[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 2805-2814.
- [70] 郑晗,王宁,马新柱,等. 基于邻域一致性的点云场景流传播更新方法[J]. 计算机研究与发展,2023, 60(2):426-434.
- ZHENG H, WANG N, MA X ZH, et al. Point cloud scene flow propagation update method based on

- neighborhood consistency [ J ]. Journal of Computer Research and Development, 2023, 60(2): 426-434.
- [71] SU H, MAJI S, KALOGERAKIS E, et al. Multi-view convolutional neural networks for 3D shape recognition[ C ]. Proceedings of the IEEE International Conference on Computer Vision, 2015: 945-953.
- [72] 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.
- [73] GUERRY J, BOULCH A, LE SAUX B, et al. SnapNet-R: Consistent 3D multi-view semantic labeling for robotics [ C ]. Proceedings of the IEEE International Conference on Computer Vision Workshops, 2017: 669-678.
- [74] WU B, WAN A, YUE X, et al. SqueezeSeg: Convolutional neural nets with recurrent CRF for real-time road-object segmentation from 3D LiDAR point cloud[ C ]. 2018 IEEE International Conference on Robotics and Automation (ICRA), 2018: 1887-1893.
- [75] 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), 2019: 4376-4382.
- [76] 张爱武,刘路路,张希珍.道路三维点多特征卷积神经网络语义分割方法[ J ]. 中国激光, 2020, 47(4): 269-277.
- ZHANG AI W, LIU L L, ZHANG X ZH. Multi-feature 3D road point cloud semantic segmentation method based on convolutional neural network[ J ]. Chinese Journal of Lasers, 2020, 47(4): 269-277.
- [77] 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, 2015: 922-928.
- [78] LI Y, PIRK S, SU H, et al. FPNN: Field probing neural networks for 3D data[ C ]. Proceedings of the 30th International Conference on Neural Information Processing Systems, 2016: 307-315.
- [79] TCHAPMI L, CHOY C, ARMENI I, et al. SEGCloud: Semantic segmentation of 3D point clouds [ C ]. 2017 International Conference on 3D Vision, 2017: 537-547.
- [80] LE T, DUAN Y. PointGrid: A deep network for 3D shape understanding [ C ]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 9204-9214.
- [81] RIEGLER G, OSMAN U A, GEIGER A. OctNet: Learning deep 3D representations at high resolutions[ C ]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 3577-3586.
- [82] ZENG W, GEVERS T. 3DContextNet: K-d tree guided hierarchical learning of point clouds using local and global contextual cues[ C ]. Proceedings of the European Conference on Computer Vision Workshops, 2018: 314-330.
- [83] QI C R, YI L, SU H, et al. PointNet ++: Deep hierarchical feature learning on point sets in a metric space [ C ]. Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017: 5105-5114.
- [84] JIANG M, WU Y, ZHAO T, et al. PointSIFT: A sift-like network module for 3D point cloud semantic segmentation [ J ]. ArXiv Preprint, 2018, ArXiv: 1807.00652.
- [85] XIE S, LIU S, CHEN Z, et al. Attentional Shape Context Net for point cloud recognition[ C ]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4606-4615.
- [86] ZHAO H, JIANG L, FU C W, et al. PointWeb: Enhancing local neighborhood features for point cloud processing [ C ]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 5565-5573.
- [87] HU Q, YANG B, XIE L, et al. RandLA-Net: Efficient semantic segmentation of large-scale point clouds[ C ]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11105-11114.
- [88] FAN S, DONG Q, ZHU F, et al. SCF-Net: Learning spatial contextual features for large-scale point cloud segmentation [ C ]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 14504-14513.
- [89] LI Y, BU R, SUN M, et al. PointCNN: Convolution on X-transformed points [ C ]. Proceedings of the 32nd International Conference on Neural Information Processing Systems, 2018: 828-838.
- [90] WANG S, SUO S, MA W C, et al. Deep parametric continuous convolutional neural networks [ C ]. Proceedings of the IEEE Conference on Computer Vision

- and Pattern Recognition, 2018; 2589-2597.
- [91] THOMAS H, QI C R, DESCHAUD J E, et al. KPconv: Flexible and deformable convolution for point clouds[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019; 6411-6420.
- [92] BOULCH A. ConvPoint: Continuous convolutions for point cloud processing [J]. Computers & Graphics, 2020, 88; 24-34.
- [93] HUA B S, TRAN M K, YEUNG S K. Pointwise convolutional neural networks [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018; 984-993.
- [94] ZHANG Z, HUA B S, YEUNG S K. ShellNet: Efficient point cloud convolutional neural networks using concentric shells statistics [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019; 1607-1616.
- [95] WU W, QI Z, FUXIN L. PointConv: Deep convolutional networks on 3D point clouds [C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019; 9621-9630.
- [96] FANG Y, XU C, ZHOU C, et al. Direction-induced convolution for point cloud analysis [J]. Multimedia Systems, 2021; 1-12.
- [97] 吴军, 崔玥, 赵雪梅, 等. SSA-PointNet++: 空间自注意力机制下的 3D 点云语义分割网络 [J]. 计算机辅助设计与图形学学报, 2022, 34(3): 437-448.  
WU J, CUI Y, ZHAO X M, et al. SSA-PointNet++: A space self-attention CNN for the semantic segmentation of 3D point cloud [J]. Journal of Computer-Aided Design & Computer Graphics, 2022, 34(3): 437-448.
- [98] WANG C, SAMARI B, SIDDIQI K. Local spectral graph convolution for point set feature learning [C]. Proceedings of the European Conference on Computer Vision, 2018; 52-66.
- [99] TE G, HU W, ZHENG A, et al. RGCNN: Regularized graph CNN for point cloud segmentation [C]. Proceedings of the 26th ACM International Conference on Multimedia, 2018; 746-754.
- [100] WANG Y, SUN Y, LIU Z, et al. Dynamic graph CNN for learning on point clouds [J]. ACM Transactions on Graphics (ToG), 2019, 38(5): 1-12.
- [101] WANG L, HUANG Y, HOU Y, et al. Graph attention convolution for point cloud semantic segmentation [C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019; 10296-10305.
- [102] LIANG Z, YANG M, DENG L, et al. Hierarchical depthwise graph convolutional neural network for 3D semantic segmentation of point clouds [C]. 2019 International Conference on Robotics and Automation (ICRA), IEEE, 2019; 8152-8158.
- [103] LEI H, AKHTAR N, MIAN A. Spherical kernel for efficient graph convolution on 3D point clouds [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 43(10): 3664-3680.
- [104] LIN Z H, HUANG S Y, WANG Y C F. Convolution in the cloud: Learning deformable kernels in 3D graph convolution networks for point cloud analysis [C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020; 1800-1809.
- [105] FURUYA T, HANG X, OHBUCHI R, et al. Convolution on rotation-invariant and multi-scale feature graph for 3D point set segmentation [J]. IEEE Access, 2020, 8; 140250-140260.
- [106] LI J, CHEN B M, LEE G H. SO-Net: Self-organizing network for point cloud analysis [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018; 9397-9406.
- [107] QIN C, YOU H, WANG L, et al. PointDAN: A multi-scale 3D domain adaption network for point cloud representation [C]. Proceedings of the 33rd International Conference on Neural Information Processing Systems, 2019; 7192-7203.
- [108] RAO Y, LU J, ZHOU J. Global-local bidirectional reasoning for unsupervised representation learning of 3D point clouds [C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020; 5376-5385.
- [109] ZAMORSKI M, ZIĘBA M, KLUKOWSKI P, et al. Adversarial autoencoders for compact representations of 3D point clouds [J]. Computer Vision and Image Understanding, 2020, 193; 102921.
- [110] LUO H, CHEN C, FANG L, et al. MS-RRFsegNet: Multiscale regional relation feature segmentation network for semantic segmentation of urban scene point clouds [J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(12): 8301-8315.
- [111] COHEN T S, GEIGER M, KÖHLER J, et al. Spherical CNNs [J]. ArXiv Preprint, 2018, ArXiv: 1801.10130.



- [112] YOU Y, LOU Y, LIU Q, et al. Pointwise rotation-invariant network with adaptive sampling and 3D spherical voxel convolution[J]. ArXiv Preprint, 2018, ArXiv: 1811.09361.
- [113] RAO Y, LU J, ZHOU J. Spherical fractal convolutional neural networks for point cloud recognition[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 452-460.
- [114] SUN X, LIAN Z, XIAO J. SRINet: Learning strictly rotation-invariant representations for point cloud classification and segmentation[C]. Proceedings of the 27th ACM International Conference on Multimedia, 2019: 980-988.
- [115] TAO Z, ZHU Y, WEI T, et al. Multi-head attentional point cloud classification and segmentation using strictly rotation-invariant representations [J]. IEEE Access, 2021, 9: 71133-71144.
- [116] ZHANG Z, HUA B S, ROSEN D W, et al. Rotation invariant convolutions for 3D point clouds deep learning[C]. 2019 International Conference on 3D Vision, 2019: 204-213.
- [117] CHEN C, LI G, XU R, et al. ClusterNet: Deep hierarchical cluster network with rigorously rotation-invariant representation for point cloud analysis [C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 4994-5002.
- [118] FANG J, ZHOU D, SONG X, et al. RotPredictor: Unsupervised canonical viewpoint learning for point cloud classification[C]. 2020 International Conference on 3D Vision, 2020: 987-996.
- [119] XU J, TANG X, ZHU Y, et al. SGMNet: Learning rotation-invariant point cloud representations via sorted gram matrix [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 10468-10477.
- [120] YI L, ZHAO W, WANG H, et al. GSPN: Generative shape proposal network for 3D instance segmentation in point cloud [C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 3947-3956.
- [121] HOU J, DAI A, NIEBNER M. 3D-SIS: 3D semantic instance segmentation of RGB-D scans [C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 4416-4425.
- [122] NARITA G, SENO T, ISHIKAWA T, et al. PanopticFusion: Online volumetric semantic mapping at the level of stuff and things [C]. 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2019: 4205-4212.
- [123] YANG B, WANG J, CLARK R, et al. Learning object bounding boxes for 3D instance segmentation on point clouds [C]. Proceedings of the 33rd International Conference on Neural Information Processing Systems, 2019: 6740-6749.
- [124] 李东威. 基于深度学习的三维实例分割技术研究[D]. 长沙: 湖南大学, 2021.
- LI D W. Research of 3D instance segmentation technology on deep learning [D]. Changsha: Hunan University, 2021.
- [125] ZHANG F, GUAN C, FANG J, et al. Instance segmentation of LiDAR point clouds [C]. 2020 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2020: 9448-9455.
- [126] ENGELMANN F, BOKELOH M, FATHI A, et al. 3D-MAP: Multi-proposal aggregation for 3D semantic instance segmentation [C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 9031-9040.
- [127] WANG W, YU R, HUANG Q, et al. SGPn: Similarity group proposal network for 3D point cloud instance segmentation [C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 2569-2578.
- [128] LAHOUD J, GHANEM B, POLLEFEYS M, et al. 3D instance segmentation via multi-task metric learning [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 9256-9266.
- [129] 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.
- [130] WANG X, LIU S, SHEN X, et al. Associatively segmenting instances and semantics in point clouds [C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 4091-4100.
- [131] PHAM Q H, NGUYEN D 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]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Reco-

- gnition, 2019: 8819-8828.
- [132] HAN L, ZHENG T, XU L, et al. OccuSeg: Occupancy-aware 3D instance segmentation[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 2940-2949.
- [133] JIANG L, ZHAO H, SHI S, et al. PointGroup: Dual-set point grouping for 3D instance segmentation[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 4867-4876.
- [134] HE T, SHEN C, VAN DEN HENGEL A. DyCo3D: Robust instance segmentation of 3D point clouds through dynamic convolution[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 354-363.
- [135] YI L, SU H, GUO X, et al. SyncSpecCNN: Synchronized spectral CNN for 3D shape segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2282-2290.
- [136] WANG Z, LU F. VoxSegNet: Volumetric CNNs for semantic part segmentation of 3D shapes[J]. IEEE Transactions on Visualization and Computer Graphics, 2019, 26(9): 2919-2930.
- [137] WANG P, GAN Y, SHUI P, et al. 3D shape segmentation via shape fully convolutional networks[J]. Computers & Graphics, 2018, 76: 182-192.
- [138] ZHU X, ZHANG Z, RUAN J, et al. ResSANet: Learning geometric information for point cloud processing[J]. Sensors, 2021, 21(9): 3227.
- [139] LIN H, FAN B, LIU Y, et al. PointSpherical: Deep shape context for point cloud learning in spherical coordinates[C]. 2020 25th International Conference on Pattern Recognition (ICPR), IEEE, 2021: 10266-10273.
- [140] SIROHI K, MOHAN R, BÜSCHER D, et al. Efficient LPS: Efficient LiDAR panoptic segmentation[J]. IEEE Transactions on Robotics, 2022, 38(3): 1894-1914.
- [141] GASPERINI S, MAHANI M A N, Marcos-Ramiro A, et al. Panoster: End-to-end panoptic segmentation of LiDAR point clouds[J]. IEEE Robotics and Automation Letters, 2021, 6(2): 3216-3223.
- [142] ZHOU Z, ZHANG Y, FOROOSH H. Panoptic-PolarNet: Proposal-free LiDAR point cloud panoptic segmentation[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13194-13203.
- [143] LI E, RAZANI R, XU Y, et al. SMAC-Seg: LiDAR panoptic segmentation via sparse multi-directional attention clustering[C]. 2022 International Conference on Robotics and Automation (ICRA), 2022: 9207-9213.
- [144] DUERR F, WEIGEL H, BEYERER J. RangeBird: Multi view panoptic segmentation of 3D point clouds with neighborhood attention[C]. 2022 International Conference on Robotics and Automation (ICRA), 2022: 11131-11137.
- [145] YANG C. FMP-Net: Fractal multi-gate mixture-of-experts panoramic segmentation for point cloud[C]. 2022 IEEE International Conference on Multimedia and Expo (ICME), 2022: 1-6.
- [146] ZENG A, SONG S, NIEBNER M, et al. 3DMatch: Learning local geometric descriptors from RGB-D reconstructions[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1802-1811.
- [147] YE W Z J, LEE G H. 3DFeat-Net: Weakly supervised local 3D features for point cloud registration[C]. Proceedings of the European Conference on Computer Vision, 2018: 607-623.
- [148] DENG H, BIRDAL T, ILIC S. PPFNet: Global context aware local features for robust 3D point matching[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 195-205.
- [149] DENG H, BIRDAL T, ILIC S. PPF-FoldNet: Unsupervised learning of rotation invariant 3D local descriptors[C]. Proceedings of the European Conference on Computer Vision, 2018: 602-618.
- [150] GOJCIC Z, ZHOU C, WEGNER J D, et al. The perfect match: 3D point cloud matching with smoothed densities[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 5545-5554.
- [151] BAI X, LUO Z, ZHOU L, et al. D3Feat: Joint learning of dense detection and description of 3D local features[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 6359-6367.
- [152] LI J, ZHANG C, XU Z, et al. Iterative distance-aware similarity matrix convolution with mutual-supervised point elimination for efficient point cloud registration[C]. Proceedings of the European Conference on Computer Vision, 2020: 378-394.
- [153] MEI Q, WANG F, TONG C, et al. PACNet: A high-

- precision point cloud registration network based on deep learning [C]. 2021 13th International Conference on Wireless Communications and Signal Processing (WCSP), IEEE, 2021: 1-5.
- [154] HORACHE S, DESCHAUD J E, GOULETTE F. 3D point cloud registration with multi-scale architecture and unsupervised transfer learning [C]. 2021 International Conference on 3D Vision, 2021, 1351-1361.
- [155] XU H, LIU S, WANG G, et al. OMNet: Learning overlapping mask for partial-to-partial point cloud registration [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 3132-3141.
- [156] FU K, LIU S, LUO X, et al. Robust point cloud registration framework based on deep graph matching [C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 8893-8902.
- [157] WANG Y, SOLOMON J M. Deep closest point: Learning representations for point cloud registration [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea (South), 2019: 3523-3532.
- [158] WANG Y, SOLOMON J. PRNet: Self-supervised learning for partial-to-partial registration [C]. Proceedings of the 33rd International Conference on Neural Information Processing Systems, 2019: 8814-8826.
- [159] DANG Z, WANG F, SALZMANN M. Learning 3D-3D correspondences for one-shot partial-to-partial registration [J]. ArXiv Preprint, 2020, ArXiv: 2006.04523.
- [160] CHOY C, DONG W, KOLTUN V. Deep global registration [C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 2514-2523.
- [161] HEZRONI I, DRORY A, GIRYES R, et al. DeepBBS: Deep best buddies for point cloud registration [C]. 2021 International Conference on 3D Vision, 2021: 342-351.
- [162] YE W Z J, LEE G H. RPM-Net: Robust point matching using learned features [C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11821-11830.
- [163] LU W, WAN G, ZHOU Y, et al. DeepVCP: An end-to-end deep neural network for point cloud registration [C]. 2019 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE Computer Society, 2019: 12-21.
- [164] AOKI Y, GOFORTH H, SRIVATSAN R A, et al. PointNetLK: Robust & efficient point cloud registration using PointNet [C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 7156-7165.
- [165] SARODE V, LI X, GOFORTH H, et al. PCRNet: Point cloud registration network using PointNet encoding [J]. ArXiv, 2019, ArXiv Preprint: 1908.07906.
- [166] HUANG X, MEI G, ZHANG J. Feature-metric registration: A fast semi-supervised approach for robust point cloud registration without correspondences [C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11366-11374.
- [167] AO S, HU Q, YANG B, et al. SpinNet: Learning a general surface descriptor for 3D point cloud registration [C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 11753-11762.
- [168] GOJCIC Z, ZHOU C, WEGNER J D, et al. Learning multiview 3D point cloud registration [C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 1759-1769.
- [169] BAI X, LUO Z, ZHOU L, et al. PointDSC: Robust point cloud registration using deep spatial consistency [C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 15859-15869.
- [170] WANG H, LIU X, KANG W, et al. Multi-features guidance network for partial-to-partial point cloud registration [J]. Neural Computing and Applications, 2022, 34(2): 1623-1634.
- [171] LI J, LEE G H. USIP: Unsupervised stable interest point detection from 3D point clouds [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 361-370.
- [172] KADAM P, ZHANG M, LIU S, et al. Unsupervised point cloud registration via salient points analysis (SPA) [C]. 2020 IEEE International Conference on Visual Communications and Image Processing (VCIP), 2020: 5-8.
- [173] KADAM P, ZHANG M, LIU S, et al. R-PointHop: A

- green, accurate and unsupervised point cloud registration method[J]. ArXiv Preprint, 2021, ArXiv: 2103.08129.
- [174] CHEN Z, YANG W, XU Z, et al. VK-Net: Category-level point cloud registration with unsupervised rotation invariant keypoints [C]. ICASSP 2021—2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021: 1900-1904.
- [175] 崔海华,姜涛,杜坤鹏,等. 基于深度学习位姿估计的多视结构光三维成像方法[J]. 光学学报, 2021, 41(17):72-81.
- CUI H H, JIANG T, DU K P, et al. 3D imaging method for multi-view structured light measurement via deep learning pose estimation[J]. Acta Optica Sinica, 2021, 41(17): 72-81.
- [176] 袁建英,王琼,李柏林. 利用标志点多视图约束实现结构光扫描高精度粗拼接[J]. 计算机辅助设计与图形学学报, 2015, 27(4):674-683.
- YUAN J Y, WANG Q, LI B L. Using multi-view network constraints among reference marker points to realize coarse registration in structured light system with high accuracy[J]. Journal of Computer-Aided Design & Computer Graphics, 2015, 27(4): 674-683.
- [177] 张婧. 文物碎片孔洞修复和基于断裂面的拼接方法研究[D]. 西安:西北大学, 2019.
- ZHANG J. Research on hole restoration and fracture surface based splicing of cultural relic fragments [D]. Xi'an: Northwest University, 2019.
- [178] MUNOZ D, BAGNELL J A, VANDAPEL N, et al. Contextual classification with functional max-margin markov networks [C]. 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009: 975-982.
- [179] LAI K, BO L, REN X, et al. A large-scale hierarchical multi-view RGB-D object dataset [C]. 2011 IEEE International Conference on Robotics and Automation, 2011: 1817-1824.
- [180] GEIGER A, LENZ P, URTASUN R. Are we ready for autonomous driving? The KITTI vision benchmark suite[C]. 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, 2012: 3354-3361.
- [181] POMERLEAU F, LIU M, COLAS F, et al. Challenging data sets for point cloud registration algorithms[J]. The International Journal of Robotics Research, 2012, 31(14): 1705-1711.
- [182] XIAO J, OWENS A, TORRALBA A. SUN3D: A database of big spaces reconstructed using SFM and object labels[C]. Proceedings of the IEEE International Conference on Computer Vision, 2013: 1625-1632.
- [183] WU Z, SONG S, KHOSLA A, et al. 3D ShapeNets: A deep representation for volumetric shapes [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1912-1920.
- [184] CHANG A X, FUNKHOUSER T, GUIBAS L, et al. ShapeNet: An information-rich 3D model repository[J]. ArXiv Preprint, 2015, ArXiv: 1512.03012.
- [185] 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 (ToG), 2016, 35(6): 1-12.
- [186] ARMENI I, SENER O, ZAMIR A R, et al. 3D semantic parsing of large-scale indoor spaces [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 1534-1543.
- [187] GAIDON A, WANG Q, CABON Y, et al. Virtual worlds as proxy for multi-object tracking analysis[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 4340-4349.
- [188] HACKEL T, SAVINOV N, LADICKY L, et al. Semantic3D. net: A new large-scale point cloud classification benchmark [J]. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2017: 91-98.
- [189] DAI A, CHANG A X, SAVVA M, et al. ScanNet: Richly-annotated 3D reconstructions of indoor scenes[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 5828-5839.
- [190] BEHLEY J, GARBADE M, MILIOTO A, et al. SemanticKITTI: A dataset for semantic scene understanding of lidar sequences [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 9297-9307.
- [191] 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]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 909-918.
- [192] DONG Z, LIANG F, YANG B, et al. Registration of large-scale terrestrial laser scanner point clouds: A

review and benchmark [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 163: 327-342.

- [193] CAESAR H, BANKITI V, LANG A H, et al. nuScenes: A multimodal dataset for autonomous driving[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11621-11631.
- [194] SUN P, KRETZSCHMAR H, DOTIWALLA X, et al. Scalability in perception for autonomous driving: Waymo open dataset [C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 2446-2454.

## 作者简介



**陈慧娴**, 2021 年于辽宁大学获得学士学位, 现为南京航空航天大学硕士研究生, 主要研究方向为图像处理与机器视觉。

E-mail: chenhuixian2017@163.com

**Chen Huixian** received her B.Sc. degree from Liaoning University (LNU) in 2021. She

is a currently a master student at Nanjing University of Aeronautics and Astronautics. Her current research interests include image processing and machine vision.



**吴一全**(通信作者), 1998 年于南京航空航天大学获得博士学位, 现为南京航空航天大学教授、博士生导师, 主要研究方向为视觉检测与图像测量、遥感图像处理与理解、红外目标检测与识别、视频处理与智能分析等。

E-mail: nuaaimage@163.com

**Wu Yiquan** (Corresponding author) received his Ph. D. degree from Nanjing University of Aeronautics and Astronautics in 1998. He is currently a professor and a Ph. D. advisor at Nanjing University of Aeronautics and Astronautics. His main research interests include remote sensing image processing and understanding, infrared target detection and recognition, visual detection and image measurement, video processing and intelligent analysis, et al.