计算机研究与发展
Journal of Computer Research and Development
DOI:10.7544/issn1000-1239.20210131
59(5):1160-1179,2022

基于深度学习的3维点云处理综述

李娇娇[1] 孙红岩[1] 董 雨[1] 张若晗[1] 孙晓鹏[1,2]
[1](辽宁师范大学计算机与信息技术学院计算机系统研究所 辽宁大连 116029)
[2](智能通信软件与多媒体北京市重点实验室(北京邮电大学) 北京 100876)
(1025843074@qq.com)

Survey of 3-Dimensional Point Cloud Processing Based on Deep Learning

Li Jiaojiao[1], Sun Hongyan[1], Dong Yu[1], Zhang Ruohan[1], and Sun Xiaopeng[1,2]
[1](*Institute of Computer System*, *School of Computer and Information Technology*, *Liaoning Normal University*, *Dalian*, *Liaoning* 116029)
[2](*Beijing Key Laboratory of Intelligent Telecommunications Software and Multimedia* (*Beijing University of Posts and Telecommunications*), *Beijing* 100876)

Abstract Deep learning has shown its superior performance in the structured data analysis such as 2-dimensional images. In recent years, with the development of LIDAR sensing equipment and related technologies, 3-dimensional point cloud scanning and acquisition has become more convenient. That makes the analysis and processing of unstructured point cloud data potential become an important research direction and obtain some progress in many fields such as computer graphics, robot, autonomous driving, virtual and augmented reality. A survey on the research of 3-dimensional point cloud processing of recent years is presented. Focusing on the application of deep learning in 3-dimensional point cloud shape analysis, structure extraction, detection and repair, we introduce the extraction method of point cloud topological structure, and compare the progress of the following research directions with the construction of neural networks as the main method: shape deformation, reconstruction, segmentation, classification, object tracking, scene flow estimation, object detection and pose estimation. Finally, we summarize the commonly used 3-dimensional point cloud public datasets, analyze and compare the characteristics and evaluation indicators of various point cloud processing task methods, and point out their advantages and disadvantages. The challenges and development directions of processing point cloud data based on deep learning are discussed.

Key words point cloud; deep learning; reconstruction; classification and segmentation; detection and tracking; pose estimation

摘 要 深度学习在2维图像等结构化数据处理中表现出了优越性能,对非结构化的点云数据分析处理的潜力已经成为计算机图形学的重要研究方向,并在机器人、自动驾驶、虚拟及增强现实等领域取得一定进展.通过回顾近年来3维点云处理任务的主要研究问题,围绕深度学习在3维点云形状分析、结构

收稿日期:2021-02-22;**修回日期**:2021-07-26
基金项目:国家自然科学基金项目(61472170);北京邮电大学智能通信软件与多媒体北京市重点实验室开放课题(ITSM201301)
This work was supported by the National Natural Science Foundation of China (61472170) and the Beijing Key Laboratory of Intelligent Telecommunications Software and Multimedia (Beijing University of Posts and Telecommunications) (ITSM201301).
通信作者:孙红岩(2083725178@qq.com)

提取、检测和修复等方向的应用,总结整理了典型算法.介绍了点云拓扑结构的提取方法,然后对比分析了变换、分类分割、检测跟踪、姿态估计等方向的以构建神经网络为主要研究方法的进展.最后,总结常用的3维点云公开数据集,分析对比了各类方法的特点与评价指标,指出其优势与不足,并从不同角度对基于深度学习的方法处理点云数据所面临的挑战与发展方向进行了讨论.

关键词 点云;深度学习;重建;分类分割;检测追踪;姿态估计

中图法分类号 TP391.4

随着3维传感器的迅速发展,3维数据变得无处不在,利用深度学习方法对这类数据进行语义理解和分析变得越来越重要.

不同3维数据(体素、网格等)表示下,深度学习的方法不尽相同,但这些方法应用于点云中都有一定的局限性,具体表现为:体素化方法会受到分辨率的限制;转换为2维图像的方法在形状分类和检索任务上取得了优越性能,但将其扩展到场景理解或其他3维任务(如姿态估计)有一定的困难;光谱卷积神经网络限制在流形网格;基于特征的深度神经网络会受到所提取特征表示能力的限制[1].

点云本身具有的无序性与不规则性为利用深度学习方法直接处理该类数据带来一定挑战性:1)规模局限性.现有研究方法一般针对小规模点云,而能处理大型点云的方法也需要切割处理,即将其分为小块后再处理.但切割可能会造成点云整体信息的丢失.2)遮挡.当被扫描对象被遮挡时,将直接导致信息的丢失,这为后续任务的处理带来严重影响.3)噪声.由于点云数据本身就是3维空间内的点集,噪声的存在直接影响点云模型的正确表示,在形状识别等任务中会造成精度的降低.4)旋转不变.对于同一模型,旋转不同角度仍表示同一对象,网络识别结果不应由于角度不同而产生差异.

随着近年来激光雷达等传感设备及相关技术的发展,3维点云的扫描与获取更为便捷,其处理技术在机器人、自动驾驶及其他领域的实际应用中已取得一定进展.基于深度学习的蓬勃发展,研究者提出了许多方法来解决相关领域的不同问题.本文对基于深度学习的点云处理任务进行详细阐述.

本文的主要贡献有4个方面:

1)从机器人、自动驾驶、虚拟和增强现实以及医学4个领域介绍点云处理技术的应用情况;

2)探讨点云拓扑结构与形状分析在应用于点云处理任务中的必要性,并总结对比多种算法;

3)归纳基于点云数据处理相关任务的方法,主要包括模型重建与变换、分类分割、检测跟踪与姿态估计,着重讨论基于深度学习的方法,并给出各种方法的优劣比较;

4)总结多个公开点云数据集,并分析各数据集中不同方法能处理的不同任务.

1 基本概念及应用情况

1.1 深度学习

机器学习推动现代科技手段的进步.网络的内容过滤及推荐、语音文本的转换及医学影像分析等应用程序越来越多地使用深度学习技术.

1)基础概念

深度学习善于发现高维数据中的复杂结构,因此可应用于科学、商业和医学等诸多领域.

深度学习利用多处理层组成的计算模型学习具有抽象层次的数据表示,关键在于其目标特征不是人类指定的,而是从大量数据中学习获取的,深度神经网络已经成为人工智能的基础.多层感知机或全连接网络堆叠线性层和非线性激活层,是神经网络的经典类型.卷积网络引入卷积层和池化层,在处理图像、视频和音频方面取得了突破性进展.递归网络可有效处理文本语音等具有连续性的数据.Transformer利用self-attention机制提取特征,最早用于处理自然语言[2-3].

2)框架平台

为了实现更复杂的模型,若从头开始编写代码,效率非常低,因此深度学习框架应运而生.本节介绍常用的深度学习框架,并将其汇总于表1中.

目前常用于点云处理的框架更多为TensorFlow与PyTorch,其他框架如Caffe与Jittor等也可用于处理点云,但应用较少.

1.2 点云处理任务

3维几何模型中,点云已经成为主要表达方式之一,其应用于深度学习中的处理技术已取得一定

成果.在不同任务驱动下,本文以构建神经网络为主要方法,通过分类与整理相关文献,将点云处理任务分为模型重建与变换、分类分割、检测跟踪与姿态估计几大类.本节总结其基本概念.

Table 1 Deep Learning Framework

表 1 深度学习框架

框架	出版年	底层语言	接口语言	优势	劣势
Theano	2007	Python	Python	灵活性高且架构简单	编译过程较慢
Caffe	2013	CUDA/C++	C++/Python/Matlab	速度快且性能高,可读性强	灵活性与扩展性低
CNTK	2014	C++	C++/Python/Java/C#	速度快,可扩展	官方解释文档晦涩难懂
TensorFlow	2015	C++/Python	C++/Python/Java	自带可视化工具	接口变动频繁,运行较慢
Keras	2015	Python	Python	可快速搭建模型,易扩展	灵活性低
MxNet	2015	C++	C++/Python/Juila	兼具灵活性与高效率	教学文档不够系统
Chainer	2015	Python	Python	速度较快且易于调试	版本间改动较大
PyTorch	2017	C/C++/Python	Python	设计直观,代码易懂	无可视化接口或工具
Jittor	2020	CUDA/C++	Python	元算子融合,统一计算图	目前使用较少,功能尚在完善

模型重建与变换包括形状修复、模型补全与变形.扫描获取到的数据并不能完美表征原物体的特性,很可能存在缺漏或误差,造成模型不完整、扭曲,故而需要对该模型进行处理,使其尽可能贴合原物体模型或目标模型,处理手段即为重建与变换.

分类分割主要包括分类、部件分割、语义分割与实例分割.在诸如机器人抓取等需求中,必须明确所抓取对象的分类,即需要判断其信息,判断即为对场景中对象语义信息标记与分类.

检测跟踪主要包括 3 维对象检测、场景流估计与目标跟踪.在诸如自动驾驶等应用中,需要明确路径与方向,确定追踪对象,并能依据当前状态自动调节或人为干预使其后续运动符合预期目标.

姿态估计主要包括位姿估计与手部姿态估计.前者需要确定对象的位置与方向,如工厂喷漆中,喷枪需要依据目标不断改变其位置与指向.后者则是为了理解人类肢体语言,如在体感游戏中,根据肢体变换执行相应游戏操作.

1.3 应用情况

3 维点云处理目前在实际应用中已经取得了一定的进展.本节以应用为导向,从机器人领域、自动驾驶领域及虚拟、增强现实领域及医学领域 4 个角度介绍点云处理技术的应用情况.

1.3.1 机器人领域

机器人抓取技术的核心在于目标识别和定位.2019 年 Lin 等人[4]利用深度神经网络学习物体外在形状,并训练网络在获取物体局部表面时也能成功抓取目标.

在机器人室内定位及导航技术方面,2020 年 Khanh 等人[5]设计了新的云端导航系统.云端导航下机器人能更准确地移动到目标位置.该技术可应用于位置服务需求,如盲人导航.

针对喷漆机器人的自动化操作,2019 年 Lin 等人[6]利用迭代最近点(iterative closest point, ICP)算法进行姿态估计,计算物体部件的位置误差,并重新调整机器人的方向,以完成所需的喷漆任务.2020 年 Parra 等人[7]设计了能够在地板下的空隙中进行隔热喷涂以提高建筑的强度及使用年限的机器人.他们针对地形不均匀等情况,提出定位模块.机器人依据传感器获取连续点云的信息.Yang 等人[8]基于点云模型表示的家具表面路径规划和边缘提取技术提出边缘喷涂,获取喷涂枪路径点序列和对应姿态.在家具等工件的生产流程中,该方法能够根据喷涂系统坐标系与家具姿态的不同,自适应地调整二者的坐标关系,以实现正确喷涂的目的.

1.3.2 自动驾驶领域

自动驾驶系统的性能受环境感知的影响.车辆对其环境的感知为系统的自动响应提供了基础.2017 年 Hanke 等人[9]提出采用光线追踪的汽车激光雷达传感器实现实时模型测量方法.使用由真实世界场景的测量构建的虚拟环境,能够在真实世界和虚拟世界传感器数据之间建立直接联系.2019 年 Josyula 等人[10]提出了利用机器人操作系统(robot operating system, ROS)和点云库(point cloud library, PCL)对点云进行分割的方法.它是为自动驾驶车辆和无人机的避障而开发的,具体涉及障碍物检测与跟踪.

激光雷达(light detection and ranging, LIDAR)和视觉感知是高水平(L4-L5)飞行员成功自动避障的关键因素.为了对大量数据进行点云标记,2020年Li等人[11]提出针对3维点云的标注工具,实现了点云3维包围盒坐标信息到相机与LIDAR联合标定后获得的2维图像包围盒的转换.

基于图的同步定位与建图(simultaneous localization and mapping, SLAM)在自动驾驶中应用广泛.实际驾驶环境中包含大量的运动目标,降低了扫描匹配性能.2020年Lee等人[12]利用加权无损检测(扫描匹配算法)进行图的构造,在动态环境下也具有鲁棒性.

1.3.3 虚拟、增强现实领域

为了更好地了解室内空间信息,2015年Tredinnick等人[13]创建了能够在沉浸式虚拟现实(virtual reality, VR)显示系统中以较快的交互速率可视化大规模LIDAR点云的应用程序,能够产生准确的室内环境渲染效果.2016年Bonatto等人[14]探讨了在头戴式显示设备中渲染自然场景的可能性.实时渲染是使用优化的子采样等技术来降低场景的复杂度实现的,这些技术为虚拟现实带来了良好的沉浸感.2018年Feichter等人[15]提出了在真实室内点云场景中抽取冗余信息的算法.其核心思想是从点云中识别出平面线段,并通过对边界进行三角剖分来获取内点,从而描述形状.

生成可用于训练新模型的标注已成为机器学习中独立的研究领域,它的目标是高效和高精度.标注3维点云的方法包括可视化,但这种方法是十分耗时的.2019年Wirth等人[16]提出了新的虚拟现实标注技术,它大大加快了数据标注的过程.

LTDAR为增强现实(augmented reality, AR)提供了基本的3维信息支持.2020年Liu等人[17]提出学习图像和LIDAR点云的局部特征表示,并进行匹配以建立2维与3维空间的关系.

使用手势自然用户界面(natural user interface, NUI)对于头戴式显示器和增强及虚拟现实等可穿戴设备中虚拟对象的交互至关重要.然而,它在GPU上的实现存在高延迟,会造成不自然的响应.2020年Im等人[18]提出基于点云的神经网络处理器.该处理器采用异构内核结构以加速卷积层和采样层,实现了使用NUI所必需的低延迟.

1.3.4 医学领域

医学原位可视化能够显示患者特定位置的成像数据,其目的是将特定病人的数据与3维模型相结合,如将手术模拟过程直接投影到患者的身体上,从而在实际位置显示解剖结构.2011年Placitelli等人[19]采用采样一致性初始配准算法(sample consensus initial alignment, SAC-IA),通过快速配准三元组计算相应的匹配变换,实现点云快速配准.

模拟医学图像如X射线是物理学和放射学的重要研究领域.2020年Haiderbhai等人[20]提出基于条件生成式对抗网络(conditional generative adversarial network, CGAN)的点云X射线图像估计法.通过训练CGAN结构并利用合成数据生成器中创建的数据集,可将点云转换成X射线图像.

2 模型形状结构

了解并确定高层形状结构及其关系能够使得模型感知局部和全局的结构,并能通过部件之间的排列和关系描绘形状,这是研究形状结构分析的核心课题.随着真实世界的扫描和信息的挖掘,以及设计模型规模的增大,在大量信息中进行3维几何模型的识别和分析变得越来越重要.

2.1 结构信息

对于3维物体,仅明确局部信息远远不够,更重要的是结构关系,它是理解整体3维结构的关键,利用结构关系可以更好地把握物体的语义信息.

2.1.1 拓扑结构

3维物体在局部结构之间有内在联系,而这些联系是智能推理的基本能力.明确部件之间的对称性、表面的连续性及主躯干和其他部位间的联系,即明确物体本身拓扑结构对3维物体的理解起重要作用.

现有的大多数方法都是对图像的空间或时间关系进行建模,为了捕捉点云局部区域之间的结构交互作用,2019年Duan等人[21]提出结构关系网络(structural relation network, SRN)解释点云中局部区域的结构依赖性.该方法通过计算局部结构之间的相互作用,解释它们之间的关系,从而使学习到的局部特征不仅编码3维结构,而且编码与其他局部区域的关系.相较于对局部信息的利用,2018年Deng等人[22]提出点对特征网络(point pair feature network, PPFNet),学习全局信息的局部特征描述符,以在无组织的点云中到对应点.

相邻点往往具有相似的几何结构,因此通过邻域图传播特征有助于学习更稳健的局部模式.2018年Shen等人[23]提出了2种新的操作来改进PointNet,使之更有效利用局部结构.第1种方法是定义局部

3维几何结构,它类似于处理图像的卷积核.第2种方法利用局部高维特征结构,从3维位置生成的近邻图上重复进行特征聚合.

为了学习点云内的空间拓扑结构,2019年He等人[24]提出GeoNet,针对不同任务,采用不同融合方法.具体来说,选择PU-Net用于点云上采样,PointNet++[25]则用于其他任务(重建、分类等).

2.1.2 算法性能对比分析

具体来说,文献[21]的SRN模块证明了结构关系推理在点云数据分析中的有效性.它具有很强的泛化能力,可以很容易地与现有网络相融合.它不需要特定的标签也能捕捉到高度相关的局部结构和常见的结构关系.对于具有复杂局部结构的点云数据,其效果更为显著.文献[22]学习纯几何上的局部描述符,并高度感知全局上下文,在精度、速度、对点密度以及对3维姿态变化的鲁棒性方面达到了较高的性能.其主要限制是内存占用.文献[23]能够有效地捕捉局部信息,直接利用局部几何结构.2种新的操作能够显著提高点云语义学习的性能.但是,这种方法需要尽量避免在顶层改变邻域图结构.文献[24]学习对局部和全局结构信息都进行编码的特征,可用于与其他网络架构融合以提高其性能,但数据集中像火箭这样的棒状物体只占小部分,所以GeoNet会在推理这类样例时出错.

2.2 形状信息

形状分析与识别中长期存在的问题是如何使得模型具有多样且逼真的3维形状,并具有相关语义和结构特点的能力.

2.2.1 形状分析

形状分析的目的往往不是几何意义上的,而是功能的或语义级别的.局部描述符是各种3维形状分析问题的核心,它应该对形状的结构变化保持不变,并且对丢失的数据、异常值和噪声具有鲁棒性.

2017年Huang等人[26]采用能够自动学习3维形状局部描述符的方法,不需要输入部件分割,通过学习多个形状类别,可直接生成通用的描述符.网络将几何和语义上相似的点嵌入描述符空间中,其产生的描述符可以用于各种形状分析应用.

借助多种数据格式,2017年Shafiq等人[27]提出点云到2维网格的表示方法和体系结构.现有的大多数方法在低层中使用较少的滤波器,在高层中逐渐增加其数量,但这可能丢失重要特征信息.Shafiq等人主张在低分辨率的输入层也使用大量滤波器,这不会显著影响参数的总数,还能实现更高精度.

基于层次化的思想,2017年Klokov等人[28]提出的Kd-network、2018年Xie等人[29]提出的注意力形状上下文网络(attentional shape context net, attentional SCN)以及2019年Liu等人[30]提出的RS-Conv(relation-shape convolutional neural network)和Mo等人[31]提出的StructureNet分别以不同方法实现分析模型形状信息的目的.

具体来说,Kd-network[28]在多方面模仿ConvNets[32]但使用kd-tree形成计算图、共享可学习参数,并以自下而上的方式计算层次表示.attentional SCN[29]不会删除点之间的空间关系,它通过构建形状上下文的层次结构,以解释端到端过程学习的局部和全局上下文信息.RS-Conv[30]可以将规则网格使用的卷积神经网络(convolutional neural network, CNN)扩展到不规则配置,实现点云的上下文形状感知学习.StructureNet[31]引入n元层次结构编码,从根本上避免了二值化引起的不必要的数据变化,从而大大简化了学习任务.

2.2.2 算法性能与对比分析

文献[26]在对象类别未知时也能产生有效的局部描述符.但它对局部信息和上下文都很敏感,且在生成局部描述符过程中,只依靠透视投影来获取局部表面信息,而投影得到的信息可能不够全面.此外,对于形状和拓扑结构变化显著的部件,它使用的非刚性对齐方法易于生成不精确的训练对应,而太多错误的训练对应将影响描述符的区分性能.文献[27]结合了体素表示和2维图像的优点.文献[28]内存占用小且计算效率高.但在形状分类中,对于较小的模型,每个epoch的学习时间短,达到收敛的周期数会增加.对于较大的模型,kd-tree构造的时间较长.文献[29]通过层次结构传递信息,以获取丰富的局部和全局形状信息,并据此来表示目标点的内在属性.文献[30]在法线估计任务中,可能对一些棘手的形状(如旋转楼梯)不太有效.文献[31]允许对具有多种几何和结构变化的包围盒和点云进行形状合成,可用于不同的分析任务中.然而,StructureNet是基于数据驱动的方法,它继承了数据集中数据的采样偏差.对于包含具有分离部分或非对称部分的模型,其生成效果不尽如人意.

3 模型重建与变换

由于遮挡等多种因素的限制,利用激光雷达等点云获取设备得到的数据存在几何信息和语义信息

的丢失以及拓扑结构的不确定,这直接导致了数据的质量问题.为后续任务的处理带来极大挑战.

3.1 形状修复与重建

点云的不完整给后续处理任务带来了一定的困难和挑战,这突显出点云补全作为点云预处理方法的重要性.

直接对原始点云进行形状补全与修复的方法是2019年Sarmad等人[33]提出的RL-GAN-NET及Wang等人[34]提出的渐进上采样网络、2020年Huang等人[35]提出的PF-Net及缪永伟等人[36]提出的基于生成对抗网络的方法.PF-Net,RL-GAN-NET与基于生成对抗网络的方法是对残缺点云的补全:PF-Net只输出缺失部分;RL-GAN-NET输出修复后的完整模型;基于生成对抗网络的方法生成缺失部分并与原输入数据合并得到完整模型.渐进上采样网络则是将稀疏点云变密集.

RL-GAN-NET[33]基于数据驱动填充缺失区域,通过控制生成对抗网络(generative adversarial network, GAN)将含噪声的部分点云转换成更具真实性的完整点云.基于片元的点集渐进上采样网络[34]由具有相同结构的上采样单元组成,但每个单元对应不同级别的细节,可以成功地将稀疏的输入点集逐步上采样到具有丰富几何细节的密集点集.PF-Net[35]能够从部分点云及其低分辨率特征点中提取多尺度特征,增强了网络提取语义和几何信息的能力.文献[36]为了修复补全模型形状,以生成对抗网络为基础,利用Wasserstein距离优化模型,补全形状的同时保持精细结构信息.

在不直接对原始点云进行操作的情况下,广泛使用的方法是基于图像进行的重建.2019年Nguyen等人[37]、Choi等人[38]都提出了由单一2维图像重建物体3维点云表示的方法.2种方法都能够根据输入图像对随机点集变形以生成目标对象,并具有可伸缩性,即输出点云的大小可以是任意的.

Nguyen等人[37]提出的点云变形网络(point cloud deformation network, PCDNet)基于局部特征,利用高层语义进行预测.它的整体形状特征是由AdaIN提取出来的.提取操作是对称映射,因此网络对无序点云具有不变性.Choi等人[38]利用CNN从输入图像中提取形状特征,然后利用提取的形状信息将随机初始化的点云变形为给定对象的形状.

文献[33-36]都可以完成补全点云,文献[34-35]直接对原始点云进行处理,不需要进行其余步骤,但文献[33]需要对原始点云进行降维.文献[37-38]从目标图像提取点的形状信息并根据提取的信息进行模型重建.

3.2 模型变形

点云变形过程中,缺乏有效语义的局部结构监督可能会在学习过程中积累误差,这将严重限制学习特征的可分辨性,进而影响网络在3维点云理解中的能力.本节根据不同方式,将变形问题分为直接变形与借助图像信息变形2种方式展开介绍.

直接变形原点云数据的方法中,研究思路是多样的,可以根据成对形状[39]、多角度分析[40]等多种方法实现.

一般的变形方法是单方向的.2018年Yin等人[39]提出的P2P-Net可以实现双方向的变形.变换前后的2点集可以是同一形状在不同视角或不同时间下的采样,也可以是不同形状中的采样.2019年Han等人[40]提出的多角度点云变分自编码器MAP-VAE(multi-angle point cloud variational auto-encoder)将有效的局部监督与变分约束下的全局监督相结合.

与直接基于点的变形不同,2019年Wang等人[41]提出了基于目标2维图像、3维网格或3维点云来变形网格的3维变形网络(3-dimensional deformation network, 3DN),Zhou等人[42]提出了基于图像信息的点云变形监测方法.前者通过保持原网格拓扑结构不变和对称性等性质,可以生成合理的变形,能够适应原模型和目标模型中不同密度的变化.后者利用点云颜色信息和反射强度信息的特点,将小波变换模极大值技术引入点云强度图像的特征提取中.

在变形中,文献[39]不需要成对的点以及点的对应关系,只需成对的形状即可实现变形.文献[40]通过多角度分析并分割点云,利用变分约束来促进新形状的生成.文献[41]更改3维网格曲面顶点位置并变形为目标模型.文献[42]需要将点云转换为2维强度图像再变形.

3.3 算法性能对比分析

在形状补全修复及模型重建任务中,文献[33]能够在缺失大量区域的情况下实现补全,其形状完成框架在具有噪声前提下,解决了点云数据的低可用性.文献[34]主要解决不同细节级别和点云密度的上采样问题,能够自适应地确定感受野.这种基于自适应的网络结构能够以端到端的方式在高分辨率点集上训练,从具有稀疏性和噪声的点集得到高精度的点云几何结构.文献[35]能够以部分点云作为输入并直接输出缺失部分,但它对数据集的要求较高.文献[36]能有效保证网络的收敛性和训练稳定

性.但是对于局部点较为稀疏且具有精细结构的模型,其修补效果并不理想.

文献[37-38]都是根据输入图像对随机点集进行变形,并生成任意大小点云表示的模型.前者能够简单有效地生成高质量的形状模型.然而,其输出坐标的预测不受语义形状信息和局部一致性的约束,这会降低性能.后者可训练参数的数量与点云大小无关,因此不需要额外开销,其效率较高.

在变形任务中,文献[39]可以在没有明确点与点之间对应关系的情况下实现双向性的几何变化,但它无法学习并保存输入形状的内在属性.文献[40]联合利用局部和全局自监督学习更具鉴别力的点云特征,并能够从不同角度捕捉局部区域的几何和结构信息.文献[41]可以使用现有的高质量网格模型来生成新模型,但当原模型或目标模型缺失区域较大时,变形还需要更改原模型的拓扑结构,否则会产生错误的对应点.文献[42]能够明确点云中各点之间的拓扑关系.

4 形状分类与分割

基于检索或划分的目的,对具有相似特征或相同属性的点云数据进行区域的分割或属性的分类是极其重要的.

4.1 基于体素的网络

使用体素这种规则的数据结构可以保留和表达空间分布.通常,每个体素仅包含布尔占用状态而不是其他详细的点分布.

2016 年 Qi 等人[43]对体素 CNN 和多视角 CNN 进行了改进并介绍了 2 种不同的体素 CNN 网络结构.第 1 种网络有利于对对象的细节进行研究,第 2 种网络有利于捕捉对象的全局结构.

2017 年 Tchapmi 等人[44]提出 SEGCloud,Wang 等人[45]提出 O-CNN.SEGCloud 联合基于体素的 3 维全卷积神经网络(3-dimensional fully convolutional neural networks, 3D-FCNN)和基于点的条件随机场(conditional random fields, CRF),从而在原始 3 维点空间中实现分割.O-CNN 的核心思想是用八叉树表示 3 维形状并离散化其表面,仅对 3 维形状边界所占据的稀疏八叉树进行 CNN 运算.其特殊之处在于八叉树的叶子节点存储的是法向量信息.

与 SEGCloud 类似,同样使用稀疏卷积的是 2018 年 Graham 等人[46]介绍的子流形稀疏卷积网络(submanifold sparse convolutional networks, SSCN).他们引入子流形稀疏卷积(submanifold sparse convolution, SSC)算子,并将其作为 SSCN 的基础,以稀疏体素作为输入,能够处理高维空间中的数据,并可用 3 维点云语义分割.

为了有效地编码体素中点的分布,2019 年 Meng 等人[47]提出新的体素变分自编码器(variational auto-encoder, VAE)网络 VV-NET.每个体素内的点分布由自编码器捕捉,该编码器利用径向基函数(radial basis functions, RBF),既提供了规则结构,又能获取详细的数据分布.

2020 年 Shao 等人[48]提出基于空间散列的数据结构,设计了 hash2col 和 col2hash,使得卷积和池化等 CNN 操作[49]能够有效地并行化,使用完美空间散列(perfect spatial hashing, PSH)整合 3 维形状.

文献[43]的 2 种体素 CNN 网络结构输出结果的精度值较高,但高分辨率会限制该网络的性能.文献[44]结合了神经网络(neural networks, NNs)、三线性插值(trilinear interpolation, TI)和全连接条件随机场(fully connected conditional random fields, FC-CRF)的优点,表现出相当高的性能.与"暴力"体素化方案相比,文献[45]使用的八叉树结构有效减少了占用的内存,但是也生成了许多冗余的空叶八叉树.特别是对于高分辨率模型,其内存开销相当大.文献[46]在识别大场景中的对象表现出高效率、高精度的优势.文献[47]占用内存较小且效率较高,但与其他方法相比,其精度不显优势且处理某些特定形状时可能会出错.文献[48]利用 3 维形状边界稀疏性,建立不同分辨率下模型的层次散列表,显著减少了 CNN 训练过程中占用的内存.

4.2 基于视图的网络

在基于视图的方法中,通常将点云投影到 2 维图像中,并利用 2 维 CNN 提取及融合图像特征,进而应用于后续具体任务中.

受现有深度学习网络的限制,基于多视角的方法只能从特定角度识别点云模型.因此,选择角度提取点云的所有信息是难点.2017 年 Lawin 等人[50]与 2019 年 Zhou 等人[51]分别提出不同的视角选择方法来应对挑战.为了完全覆盖渲染视图中的点云,Lawin 等人[50]控制等距角,生成具有不同俯仰角和平移距离的图像.Zhou 等人[51]提出了 MVPointNet,其视图是利用变换网络(transformer network, T-Net)[1]生成的变换矩阵来确定多个相同的旋转角度获取的,这保证了网络对几何变换的不变性.

点云包含了丰富的3维信息,不同的视图包含不同的2维信息.不同于以上只利用不同视角图像的方法,2017年Guerry等人[52]提出的SnapNet-R可同时利用2维图像和3维空间结构中的信息,2019年Jaritz等人[53]提出的MVPNet将2维图像特征聚合到3维中,2019年Yang等人[54]提出的Relation Network综合考虑了不同视图之间区域到区域和视图到视图的关系.

对于单个图像,SnapNet-R[52]生成多个视图,所有视图都对应于从不同的角度看到的场景.MVPNet[53]采用贪心算法动态选择RGB-D帧,并获取不同帧上的2维图像特征,然后将这些特征提升到3维,并将它们聚集到原始点云中以进行语义分割.对于给定视图中的某区域,Relation Network[54]从其他视图中找到匹配或相关区域,并利用来自匹配或相关区域的线索来重新增强该区域的信息.此外,其还采用注意选择机制生成各视图的重要性分数,该分数反映视图的相对辨别能力.

文献[50]仅使用颜色值或法线作为输入也能取得较高的性能.文献[51]提取中心点与邻域点之间的信息,在3维形状分类中精度较高.文献[52]证明了3维结构重建与2维语义标记是互利的.文献[53]计算了2维图像特征,这可以从高分辨率的图像中收集额外的信息,提升到3维中的2维特征包含上下文信息.文献[54]的网络结构考虑了区域到区域的关系和视图到视图的关系,对3维对象的学习能力较强.

4.3 基于点的网络

CNN处理点云的研究中,大多数方法需要对点云进行体素化或将其转化为视图等其他操作,这会带来一定的局限性.直接对点云进行处理即相当于直接处理原始数据,其优势十分显著.

基于点云数据不规则的特点,针对采样密度不确定的情况,2018年Atzmon等人[55]提出点卷积神经网络(point convolutional neural networks, PCNN),对图像CNN进行了泛化,允许调整网络结构,利用扩展算子和约束算子生成适应点云的卷积.Hermosilla等人[56]提出Monte Carlo卷积,使用Monte Carlo积分做卷积计算,利用这一概念可以组合处理来自不同层的多个采样信息.2020年Zhai等人[57]提出双输入网络(dual-input network, DINet)框架和适用于该框架的正则化方法,可以减少噪声和背景对分类任务的干扰.

对于局部信息丢失问题,2019年白静等人[58]提出的MSP-Net与2020年Hu等人[59]提出的RandLA-Net都能在网络训练过程中有效改变感受野范围.2021年杜静等人[60]引入局部残差块能够提取更多局部细节信息.

只使用最高层特征将会丢失较多底层细节信息,在满足点云覆盖的完备、空间分布的自适应性及区域之间的重叠性的要求下,文献[58]提出多尺度局部区域划分及多尺度局部特征融合算法.

昂贵的采样技术或计算繁重的预/后处理使得大多数方法只能处理小规模点云.RandLA-Net[59]使用随机采样解决规模局限性,引入局部空间编码(local spatial encoding, LocSE)模块逐步增大感受野来学习复杂的局部结构,能有效保留几何特征.文献[60]融合几何结构特征及语义特征,改进残差模块以实现点云数据复杂几何结构的提取.

基于点云本身无序性的特点,为了满足置换不变性与顺序不变性,2019年Wu等人[61]提出PointConv、Wang等人[62]提出DGCNN、Komarichev等人[63]提出环状卷积、Zhang等人[64]提出ShellNet,2020年Zhao等人[65]提出Point Transformer,2021年Guo等人[66]提出PCT.

PointConv[61]扩展到反卷积PointDeconv可以获得更好的分割结果,这是大多数现有算法不能实现的操作.DGCNN[62]显式地构造局部图并学习边的嵌入,因此能够在语义空间中对点进行分组.点云中普遍存在法向翻转,环形保护策略下,无论相邻点如何排列,其结果不变.Komarichev等人[63]将搜索区域限制在局部环形区域中.这使得相邻点序列的首尾相连,因此,可以基于任意起始位置排序.卷积运算ShellConv使用同心球的统计信息来定义代表性特征并解决点序模糊性.ShellNet[64]是在ShellConv的基础上进一步建立的.

Point Transformer[65]与PCT[66]的相同之处在于都以transformer为基础.文献[65]设计了适合于处理点云的point transformer layer,并构造以其为核心的residual point transformer block,它有助于局部特征向量之间的信息交换,为所有数据点生成新的特征向量.文献[66]的PCT编码器将输入坐标嵌入到特征空间中生成特征,继而输入注意模块中获取具有区分性的表示并学习点的语义信息.

针对点云密度不同的问题,文献[55]计算效率高,对点云中点的阶数不变,对采样密度变化鲁棒性强,但其计算量较大.文献[56]参数数量较少,但在不同规模的点云中,效率与质量方面的高性能不能

兼得.文献[57]在处理包含大量噪声和复杂背景信息的真实数据时也能表现出较高精度.文献[58]所提的 MSP-Net 是多尺度分类网络,随着神经网络深度的增加及感受野的扩大,其特征抽象程度也越高.文献[59-60]可直接处理大规模点云,前者能够很好地权衡效率和质量问题,后者注意力机制的引入及残差模块的改进,提高了网络获取更具区分性语义特征的提取能力.

针对顺序与置换不变的特点,文献[61]能够完全逼近任意 3 维点上的连续卷积,特定的反卷积操作可以获得更好的分割结果.文献[62]使用有向图表示点云的局部结构,能够更好地捕捉结构信息,但该方法的某些细节设计影响了其效率.文献[63]可以在局部环形区域上定义任意大小的卷积核,更好地捕获邻域结构,且捕获到的信息不重叠.文献[64]在不增加网络层数的情况下允许感受野更大,且解决了卷积阶数问题.文献[65]中 residual point transformer block 集成 self-attention 与线性投影,可以减少维数并加速处理过程.文献[66]用注意模块的输入和注意特征之间的偏移量来代替注意特征,提出隐式拉普拉斯算子和归一化改进,偏移注意优化过程可以近似理解为拉普拉斯过程.

4.4 算法性能对比分析

本节将从评估指标与算法详细对比分析 2 部分进行介绍.

4.4.1 评估指标

目前广泛使用的指标为准确率(accuracy, Acc)、精确率(precision, P)、召回率(recall, R)以及交并比(intersection over union, IoU).

指标计算公式中,TP(true positives)表示正类判定为正类,FP(false positives)表示负类判定为正类,FN(false negatives)表示正类判定为负类,TN(true negatives)表示负类判定为负类.

N 类对象中,第 i 类的准确率为

$$Acc_i=\frac{TP_i+TN_i}{TP_i+FP_i+FN_i+TN_i}. \tag{1}$$

N 类对象的类间平均准确率为

$$mAcc=\frac{1}{N}\sum_{i=1}^{N}Acc_i. \tag{2}$$

精确率指的是所有被判定为正类($TP+FP$)中,真实的正类(TP)所占的比例.N 类对象中,第 i 类的精确率为

$$P_i=\frac{TP_i}{TP_i+FP_i}. \tag{3}$$

N 类对象的总体精度为

$$OA=\frac{\sum_{i=1}^{N}TP_i}{\sum_{i=1}^{N}(TP_i+FP_i)}. \tag{4}$$

N 类对象中,第 i 类的交并比为

$$IoU_i=\frac{TP_i}{TP_i+FP_i+FN_i}. \tag{5}$$

所有类的平均交并比为

$$mIoU=\frac{1}{N}\sum_{i=1}^{N}IoU_i. \tag{6}$$

召回率指所有真实为正类($TP+FN$)中被判定为正类(TP)占的比例,其计算方式为

$$R=\frac{TP}{TP+FN}. \tag{7}$$

除了这些指标外,还有一个重要指标为平均精准度(average precision, AP).基于精确率和召回率即可得到 PR(precision-recall)曲线(R 值为横轴,P 值为纵轴),则 PR 曲线的线下面积即为 AP 值.注:mAP 为所有类别下 AP 的均值.

4.4.2 算法对比

文献[43-48]都是基于体素的方法.具体来说,文献[43]提出的 2 种体素 CNN 网络结构在结合数据扩充和多方向池化后,性能有显著的提升.该方法显著地改善了体素 CNN 在 3 维形状分类方面的研究现状,但更高的 3 维分辨率会限制该网络的性能.文献[44]使用了基于标准体素的 3D-FCNN,并且仍然可以使用稀疏卷积来适应体素的稀疏性.文献[45]利用了八叉树表示的稀疏性和形状的局部方向性,实现了紧凑的存储和快速的计算.但其存储和计算开销随着八叉树深度的增加呈 2 次增长,且该算法没有考虑形状的几何变化.文献[46]在识别单个模型部件或大场景中的对象时,都表现出高效率高精度的优势.文献[47]进一步使用 RBF 来计算每个体素内的局部连续表示.此外,对对称性进行了编码,并在不增加参数数量的情况下提高了网络的表达能力,获得更稳健的分割结果.VV-Net 对包含噪声的数据具有一定的鲁棒性.但对某些特定形状的输入,它得到的效果并不好.文献[48]中 PSH 的运用使得散列表的大小与输入 3 维模型的大小相同.2 种 GPU 算法使得基于散列的模型实现了 CNN 操作的并行计算,其内存开销比现有的基于八叉树的方法(如 O-CNN)小得多,运行速度较快.但所有 PSH 都是使用 CPU 生成的,使用 GPU 可进一步加速该过程.

文献[50-54]都是基于视图的方法.只利用不同视角图像的算法中,文献[50]从点云中提取不同信息(如颜色、深度值和法线)并组合多种信息作为输入,判断其对分割结果的影响.该方法证明多种信息的融合能显著提高分割性能.该方法得益于大量现成的用于图像分割和分类的数据集,这大大减少甚至消除了训练3维数据的需要.此外,该方法提高了空间分辨率和分割结果的质量.文献[51]引入了丰富的局部结构特征,这些特征包含了中心点及其邻域点之间的信息,能够更好地表示和捕捉模型的上下文结构.多个视图的融合包含了更多的点云信息,使网络在3维目标分类任务中具有更强的鲁棒性和准确性.

除了不同视角图像外,还考虑其他信息(点云信息、不同视图的联系等)的算法中,与仅使用RGB-D单幅图像相比,文献[52]利用了点云中的信息,具有更高的完备性,能够快速生成与原始相机位置不同的视点.文献[53]有效融合2维视角图像和3维点云,在将2维信息提升到3维之前,先计算了2维图像特征,证明了从多视角图像中计算图像特征的优越性.其网络训练速度较快,对密度变化的点云具有更高的鲁棒性,在遮挡情况下也能实现良好的分割.文献[54]从不同的角度有效地连接相应的区域,从而增强了单个视图图像的信息,利用视图之间的相互关系,并对这些视图进行集成以获得有区别的3维对象表示.

文献[55-66]都是基于点的处理方法.主要针对点云密度问题的算法中,文献[55]的框架由扩展算子和约束算子组成,其核心思想得到适应任意点云的卷积.文献[56]能在相邻点数目可变的感受野中直接工作.特定的结构可以在2个不同采样密度之间进行卷积,实现从较低采样到较高采样的映射,也可以降低采样分辨率.该方法在均匀与非均匀采样中都表现出优越性能.但是,它存在效率与质量的权衡:小规模点云或较小的感受野中,其计算速度很快但结果不精确;大规模点云或较大的感受野中,其结果精度较高但计算速度慢.文献[57]提出适用于DINet框架的正则化方法能有效减少点云噪声和遮挡对原始信息的干扰.文献[58]建立不同尺度的局部感受野,能随着感受野的扩大获得抽象程度更高的多尺度局部语义重要特征,其多尺度局部空间划分贴合点云空间分布,但该算法未考虑单一尺度局部区域的关系.文献[59-60]都采用随机采样解决点云规模过大的问题,但随机采样在快速采样的同时很可能会丢失关键特征.二者为弥补该问题所采取的方法也有一定的相似之处:前者引入LocSE,后者设计多特征提取模块.它们都对中心点、邻域点的3维坐标、中心点与邻域点间的欧氏距离和相对坐标进行编码,用于后续特征处理.

主要针对顺序与置换不变的算法中,文献[61]可以实现与2维卷积网络中相同的平移不变性以及点云中点的顺序不变性.此外,它可以在保证高效利用内存的同时实现改变求和顺序技术.文献[62]动态更新图的同时聚合点,它描述的是相邻点之间的边特征.文献[63]可以在具有相同大小卷积核且不增加参数的情况下覆盖较大的区域.基于环的方法可以聚集更多具有区分性的特征,能够更好地捕获形状的几何细节.文献[64]定义从内到外的卷积顺序,允许高效的邻域点查询.ShellNet具有快速的局部特征学习能力,同时能以较快的速度训练网络.文献[65]引入了可训练的、参数化的位置编码,这对后续特征转换非常重要.文献[66]采用邻域嵌入策略来改进点嵌入,增强局部上下文信息获取能力.其注意机制在获取全局特征方面是有效的,但是它可能忽略了点云学习所必需的局部几何信息.

表2与表3分别给出各算法在处理分类与分割任务的性能比较.其中,由于文献[45,48]受分辨率影响,表中给出分辨率为64^3的结果.

Table 2 Performance Comparison of Classified Tasks

表2 分类任务性能比较 %

文献	方法	ModelNet 10		ModelNet 40	
		OA	*mAcc*	*OA*	*mAcc*
文献[1]	PointNet			89.2	86.2
文献[25]	PointNet++			90.7	
文献[45]	O-CNN			89.9	
文献[48]	H-CNN			89.3	
文献[51]	MVPointNet	95.2	95.1	93.2	90.3
文献[54]	Relation Network	95.3	95.1	94.3	92.3
文献[57]	DI-PointNet			88.9	86.8
	DI-PointCNN			92.1	88.3
文献[58]	MSP-Net	94.7		91.7	
文献[61]	PointConv			92.5	
文献[62]	DGCNN			92.9	90.2
文献[63]	A-CNN	95.5	95.3	92.6	90.3
文献[64]	ShellNet			93.1	
文献[65]	Point Transformer			93.7	90.6
文献[66]	PCT			93.2	

Table 3 Performance Comparison of Segmentation Tasks

表 3 分割任务性能比较 %

文献	方法	S3DIS			Semantic3D			ScanNet		ShapeNet	KITTI
		OA	*mAcc*	*mIoU*	*OA*	*mAcc*	*mIoU*	*OA*	*mIoU*	*mIoU*	*mIoU*
文献[1]	PointNet	78.6		47.6						83.7	14.6
文献[25]	PointNet++	80.1		54.5				84.5	33.9		20.1
文献[44]	SEGCloud		57.4	48.9	88.1	73.1	61.3				36.8
文献[45]	O-CNN									85.9	
文献[47]	VV-Net	87.8		78.2							
文献[53]	MVPNet	88.1	68.7	62.4					64.1		
文献[59]	RandLA-Net	87.2	81.5	68.5	94.4		76.0				55.9
文献[60]	multi-feature	87.2	81.7	69.2	93.5	74.0					
文献[61]	PointConv								55.6	85.7	
文献[62]	DGCNN	84.1		56.1						85.2	
文献[63]	A-CNN	87.3	62.9					85.4		85.9	
文献[64]	ShellNet	87.1		66.8	93.2		69.4	85.2			
文献[65]	Point Transformer	90.8	76.5	70.4						86.6	
文献[66]	PCT		67.65	61.33						86.4	

5 目标检测与跟踪

自动驾驶、机器人设计等领域中,3 维目标检测与跟踪至关重要.自动驾驶车辆和无人机的避障等实际应用中,涉及障碍物检测与跟踪.

5.1 3 维目标跟踪

目标跟踪是推测帧的属性并预测变化,即推断对象的运动情况,可以利用预测对象的运动信息进行干预使之实际运动符合预期目标或用户要求.

为了从点云中推断出目标对象的可移动部件以及移动信息,2019 年 Yan 等人[67]提出 RPM-Net.其特定的体系结构够预测对象多个运动部件在后续帧中的运动,同时自主决定运动何时停止.

2020 年 Wang 等人[68]提出 PointTrackNet.网络中提出了新的数据关联模块,用于合并 2 帧的点特征,并关联同一对象的相应特征.首次使用 3 维 Siamese 跟踪器并应用于点云的是 Giancola 等人[69].基于 Achlioptas 等人[70]提出的形状完成网络,2019 年 Giancola 等人[69]通过使用给定对象的语义几何信息丰富重编码后的表示来提高跟踪性能.

2019 年 Burnett 等人[71]提出 aUToTrack,使用贪婪算法进行数据关联和扩展卡尔曼滤波(extended Kalman filter, EKF)跟踪目标的位置和速度.Simon 等人[72]融合 2 维语义信息及 LIDAR 数据,还引入了缩放旋转平移分数(scale-rotation-translation score, SRTs),该方法可更好地利用时间信息并提高多目标跟踪的精度.

文献[67]可以从开始帧和结束帧的移动部分导出变化范围,故参数中不含变换范围,减少了参数个数.文献[68]提供的跟踪关联信息有助于减少目标短期消失的影响,其性能比较稳定,但是当汽车被严重遮挡时,结果会出现问题.文献[69]解决了相似性度量、模型更新以及遮挡处理 3 方面的问题,但该方法直接利用对称性来完善汽车整体形状会导致更多噪声.文献[71]实际需要计算被检测物体的质心,这种方法能有效检测行人,但对于汽车来说,其结果并不准确.文献[72]提出的 SRTs 可用于快速检测目标,提高了准确性和鲁棒性.

5.2 3 维场景流估计

机器人和人机交互中的应用可以从了解动态环境中点的 3 维运动,即场景流中受益.以往对场景流的研究方法主要集中于立体图像和 RGB-D 图像作为输入,很少有人尝试从点云中直接估计.

2019 年 Behl 等人[73]提出 PointFlowNet,网络联合预测 3 维场景流以及物体的 3 维包围盒和刚体运动.Gu 等人[74]提出 HPLFlowNet,可以有效地处理非结构化数据,也可以从点云中恢复结构化信息.能在不牺牲性能的前提下节省计算成本.Liu 等人[75]提出 FlowNet3D.由于每个点都不是“独立”的,相邻

点会形成有意义的信息,故而FlowNet3D网络嵌入层会学习点的几何相似性和空间关系.

文献[73]先检测出object并计算出ego motion和scene flow,再去回归各个object的motion,它从非结构化点云中直接估计3维场景流.文献[74-75]的整体结构类似,都是下采样-融合-上采样,直接拟合出scene flow.

5.3 3维目标检测与识别

在城市环境中部署自动型车辆是一项艰巨的技术挑战,需要实时检测移动物体,如车辆和行人.为了在大规模点云中实现实时检测,研究者针对不同需求提出多种方法.

2019年Shi等人[76]提出PointRCNN,将场景中的点云基于包围盒生成真实分割掩模,分割前景点的同时生成少量高质量的包围盒预选结果.在标准坐标中优化预选结果来获得最终检测结果.

2019年Lang等人[77]提出编码器PointPillars.它学习在pillars中组织的点云表示,通过操作pillar,无需手动调整垂直方向的组合.由于所有的关键操作都可以表示为2维卷积,所以仅使用2维卷积就能实现端到端的3维点云学习.

考虑到模型的通用性,2019年Yang等人[78]提出STD,利用球形锚生成精确的预测,保留足够的上下文信息.PointPool生成的规范化坐标使模型在几何变化下具有鲁棒性.box预测网络模块消除定位精度与分类得分之间的差异,有效提高性能.

2019年Liu等人[79]提出大规模场景描述网络(large-scale place description network, LPD-Net).该网络采用自适应局部特征提取方法得到点云的局部特征.此外,特征空间和笛卡儿空间的融合能够进一步揭示局部特征的空间分布,归纳学习整个点云的结构信息.

为了克服一般网络中点云规模较小的局限性,2019年Paigwar等人[80]提出Attentional PointNet.利用Attentional机制进行检测能够在大规模且杂乱无章的环境下重点关注感兴趣的对象.

2020年Shi等人[81]提出PV-RCNN.它执行2步策略:第1步采用体素CNN进行体素特征学习和精确的位置生成,以节省后续计算并对具有代表性的场景特征进行编码;第2步提取特征,聚集特征可以联合用于后续的置信度预测和进一步细化.

文献[76]生成的预选结果数量少且质量高.文献[77]能够利用点云的全部信息,其计算速度较快.文献[78]能够将点特征从稀疏表示转换为紧凑表示,且用时较短.文献[79]充分考虑点云的局部结构,自适应地将局部特征作为输入,在不同天气条件下仍能体现出健壮性.文献[80]不必处理全部点云,但预处理步骤使得计算成本较大.文献[81]结合基于体素的与基于PointNet的优势,能够学习更具鉴别力的点云特征.

5.4 算法性能对比分析

跟踪算法中,文献[67]主要关注的是物体部件的跟踪,文献[68]与文献[69]则主要检测同一物体在不同时间的状态.文献[67]的优势在于可以同时预测多个运动部件及其各自的运动信息,进而产生基于运动的分割.该方法实现高精度的前提是输入对象的几何结构明确,否则很有可能会生成不完美的运动序列.文献[68]在快速变化的情况下,如突然刹车或转弯,其结果仍可靠.但是当目标被严重遮挡时,其结果并不可靠.由于大多数模型(如汽车模型)只能从单侧看到,文献[69]利用对称性完善汽车形状的方法未必是有效的.文献[71]的处理方法较简单且用时较短,在CPU上运行时间不超过75 ms.它能在检测行人时达到较高性能.但用于拥挤道路的自动驾驶时,其采用的质心估计对于汽车并不准确.文献[72]同时利用2维信息与3维LIDAR数据,且使用的SRTs指标可缩短训练时间.

场景流估计算法中,文献[73]联合3维场景流和刚性运动进行预测,其效率较高且处理不同运动时具有鲁棒性.文献[74]与文献[75]都以端到端的方式从点云中学习场景流.前者从非结构化的点云中恢复结构化,在生成的网格上进行计算,后者则是在点云的连续帧中计算.

检测算法中,文献[76]不会在量化过程中丢失信息,也不需要依赖2维检测来估计3维包围盒,故而可以充分利用3维信息.文献[77]的处理速度较快,计算效率较高.文献[78]具有较高的计算效率和较少的计算量,能够同时集成基于点和基于体素的优点.文献[79]引入局部特征作为网络输入,有助于充分了解输入点云的局部结构.文献[80]能够有效地获取数据的3维几何信息.但是,将点云裁剪成较小区域等预处理步骤增加了计算成本.文献[81]结合了基于体素与基于PointNet的优点,不仅保留了精确的位置,而且编码了丰富的场景上下文信息.

表4给出KITTI数据集下不同算法处理跟踪任务的性能对比.指标为多目标跟踪准确度(multi-object tracking accurancy, MOTA)、多目标跟踪精确度(multi-object tracking precision, MOTP)、目标大部分被跟踪到的轨迹占比(mostly tracked,

MT)、目标大部分跟丢的轨迹占比(mostly lost, ML)、ID改变总数量(ID switches, IDS)、跟踪过程中被打断的次数(fragmentation, FRAG)及每秒帧数(frames per second, FPS).

Table 4 Performance Comparison of Tracking Tasks

表 4 处理跟踪任务性能对比

文献	方法	框架	MOTA↑/%	MOTP↑/%	MT↑/%	ML↓/%	IDS↓	FRAG↓	FPS↑
文献[68]	PointTrackNet	TensorFlow	68.23	76.57	60.62	12.31	111	725	
文献[71]	aUToTrack		82.25	80.52	72.62	3.54	1025	1402	100
文献[72]	Complexer-YOLO		75.70	78.46	58.00	5.08	1186	2092	100

注:"↑"表示值越大性能越好,"↓"表示值越小性能越好.

表5给出在KITTI数据集下3维检测框(3-dimensional detection benchmark, 3D)、BEV视图下检测框(bird eye view detection benchmark, BEV)与检测目标旋转角度(average orientation similarity detection benchmark, AOS)的检测结果.其中,评估指标为AP,IoU阈值为:汽车0.7,行人和自行车0.5.

Table 5 Performance Comparison of Detecting Tasks

表 5 处理检测任务性能对比 %

指标	方法	框架	汽车			行人			自行车		
			较易	中等	较难	较易	中等	较难	较易	中等	较难
3D	Complexer-YOLO[72]		55.63	49.44	44.13	19.45	15.32	14.80	28.36	23.48	22.85
	PointRCNN[76]	PyTorch	84.32	75.42	67.86						
	PointPillars[77]	PyTorch	79.05	74.99	68.30	52.08	43.53	41.49	75.78	59.07	52.92
	STD[78]	TensorFlow	86.61	77.63	76.06	53.08	44.24	41.97	78.89	62.53	55.77
	Attentional PointNet[80]	PyTorch	58.62	52.28	47.23						
	PV-RCNN[81]		90.25	81.43	76.82	52.17	43.29	40.29	78.60	63.71	57.65
BEV	Complexer-YOLO[72]		74.23	66.07	65.70	22.00	20.88	20.81	36.12	30.16	26.01
	PointRCNN[76]	PyTorch	89.28	86.04	79.02						
	PointPillars[77]	PyTorch	88.35	86.10	79.83	58.66	50.23	47.19	79.14	62.25	56.00
	STD[78]	TensorFlow	89.66	87.76	86.89	60.99	51.39	45.89	81.04	65.32	57.85
	PV-RCNN[81]		94.98	90.65	86.14	59.86	50.57	46.74	82.49	68.89	62.41
AOS	Complexer-YOLO[72]		87.97	79.08	78.75	37.80	31.80	31.26	64.51	56.32	56.23
	PointRCNN[76]	PyTorch	90.76	89.55	80.76						
	PointPillars[77]	PyTorch	90.19	88.76	86.38	58.05	49.66	47.88	82.43	68.16	61.96

6 姿态估计

3维姿态估计即确定目标物体的方位指向问题,在机器人、动作跟踪和相机定标等领域都有应用.

6.1 位姿估计

解决3维可视化问题的中间步骤一般是确定3维局部特征,位姿估计是其中最突出的问题.

2017年Elbaz等人[82]提出的LORAX采用了可以处理不同大小点云的设置,并设计了对大规模扫描数据有效的算法.2019年Speciale等人[83]将原始3维点提升到随机方向的3维线上,仅存储3维线和3维点的关联特征描述符,这类映射被称为3维线云.2019年Zhang等人[84]从目标点云中自动提取关键点,生成对刚性变换不变的逐点特征,利用层次式神经网络预测参考姿态对应的关键点坐标.最后计算出当前姿态与参考姿态之间的相对变换.

2018年Deng等人[85]提出了PPF-FoldNet,通过点对特征(point pair feature, PPF)对局部3维几何编码,建立了理论上的旋转不变性,同时兼顾点的稀疏性和置换不变性,能很好地处理密度变化.

考虑到成对配准描述符也应该为局部旋转的计算提供线索,2019年Deng等人[86]提出端到端的配准方法.这种算法在PPF-FoldNet[85]的工作基础上,通过学习位姿变换将3维结构与6自由度运动解

耦.该方法基于数据驱动来解决2点云配准问题.

2020年Kurobe等人[87]提出CorsNet,连接局部特征与全局特征,不直接聚集特征,而是回归点云之间的对应关系,比传统方法集成更多信息.

文献[82]解决了2点云之间点数相差数倍的问题,它简单、快速,并且具备扩展性,但在极端情况下,其结果会出错.文献[83]只使用了一个几何约束,其准确性与召回率可以与传统方法媲美,但这种方法的速度较慢.文献[84]需要较少的训练数据,因此对于没有纹理的对象,它更快、更精确.文献[85]继承了多个网络框架的优点,且充分利用点云稀疏性,能够快速提取描述符.文献[86]提高了成对配准的技术水平且减少了运行时间.文献[87]结合了局部与全局特征,从平移和旋转的角度而言准确性较高.表6上半部分给出位姿估计算法的核心方法及优势对比分析.

Table 6 Comparison of Pose Estimation Methods

表6 姿势估计方法对比

任务	文献	出版年	方法	优势	框架
位姿估计	文献[82]	2017	LORAX	性能高且稳健性强	TensorFlow
	文献[83]	2019	3维线云	准确率高且稳健性强	
	文献[84]	2019	层次结构	需较少训练数据且精度高	
	文献[85]	2018	PPF-FoldNet	能够快速提取描述符	
	文献[86]	2019	利用关键点信息	速度快且泛化力强	
	文献[87]	2020	CorsNet	充分利用信息	
手部姿态估计	文献[88]	2018	SHPR-Net	成本低且鲁棒性强	TensorFlow
	文献[89]	2018	堆叠PointNet模块	兼具高性能与准确率	PyTorch
	文献[90]	2019	PEL	内存占用小	TensorFlow
	文献[91]	2019	SO-HandNet	性能高	PyTorch
	文献[92]	2018	Hand PointNet	准确率高	PyTorch
	文献[93]	2020	NARHT	鲁棒性强	PyTorch

6.2 手部姿态估计

点云作为更简单有效的数据表示方法,其输入的点集和输出的手部姿态共享相同表示域,有利于学习如何将输入数据映射到输出姿态上.

为了直接从点云中估计手部姿态,同样以手部3维点云为输入,2018年Chen等人[88]提出语义手部姿态回归网络(semantic hand pose regression network, SHPR-Net),通过学习输入数据的变换矩阵和输出姿态的逆矩阵应对几何变换的挑战.Ge等人[89]提出的方法输出反映手部关节的每点贴近度和方向的heat-maps和单位向量场,并利用加权融合从估计的heat-maps和单位向量场中推断出手部关节位置.2019年Li等人[90]提出的方法以置换等变层(permutation equivariant layer, PEL)为基本单元,构建了基于PEL的残差网络模型.且手部姿态是利用点对姿势的投票方案来获得的,这避免了使用最大池化层提取特征而导致的信息丢失.

现有的手部姿态估计方法大多依赖于训练集,而在训练数据上标注手部3维姿态费时费力.2019年Chen等人[91]提出的SO-HandNet旨在利用未注记数据以半监督的方式获得精确的3维手部姿态估计.通过自组织映射(self-organizing map, SOM)模拟点的空间分布,然后对单个点和SOM节点进行层次化特征提取,最终生成输入点云的判别特征.

2018年Ge等人[92]提出Hand PointNet,提出的精细化网络可以进一步挖掘原始点云中更精细的细节,能够回归出更精确的指尖位置.Huang等人[93]认为学习算法不仅要研究数据的内在相关性,而且要充分利用手部关节之间的结构相关性及其与输入数据的相关性.基于此,2020年他们提出非自回归手部transformer(non-autoregressive hand transformer, NARHT),以关节特征的形式提供参考手部姿态,利用其固有的相关性来逼近输出姿态.

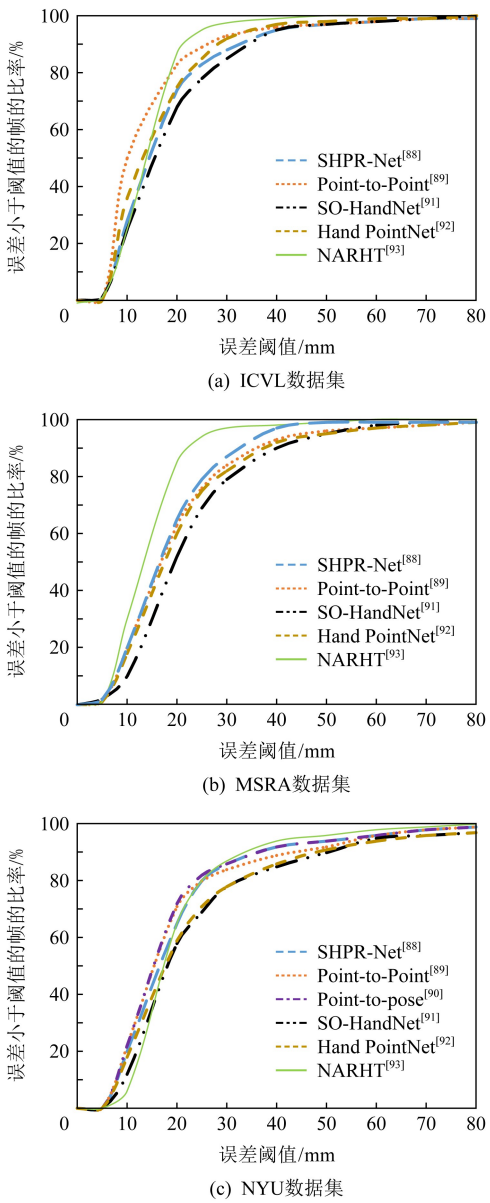
文献[88]对点云的几何变换具有鲁棒性.文献[89]能够很好地捕捉空间中点云的结构信息.文献[90]较利用体素的方法占用内存更少,但其效率不如基于深度图像的方法.文献[91]的特征编码器能够揭示输入点云的空间分布.文献[92]能够捕捉复杂的手部结构,并精确地回归出手部姿态的低维表示.文献[93]采用新的non-autoregressive结构学习

机制来代替 transformer 的自回归分解，在解码过程中提供必要的姿态信息.表 6 下半部分给出手部姿态估计算法的核心方法及优势对比分析.

6.3 算法性能对比分析

位姿估计方法中，核心问题是找到旋转矩阵与平移矩阵.文献[83,85-86]都利用了 RANSAC 迭代算法.其中，文献[83]实现了鲁棒、准确的 6 自由度姿态估计.文献[85]是无监督、高精度、6 自由度变换不变的网络.文献[86]在挑战成对配准的真实数据集方面优于现有技术，具有更好的泛化能力且速度更快.文献[82]的 LORAX 能够并行实现，效率较高，适合实时应用.它对随机噪声、密度变化不敏感，并且其鲁棒性仅在极端水平下才会恶化.文献[84]使用较少的训练图像实现了较高的准确性.文献[87]提出的 CorsNet 回归的是对应关系，而不是直接姿态变化.

手部姿态估计方法中，文献[88]可获得更具代表性的特征.SHPR-Net 可以在不改变网络结构的前提下扩展到多视点的手部姿态估计，这需要将多视点的深度数据融合到点云上.然而，融合后的点云也会受到噪声的影响.文献[89]可以更好地利用深度图像中的 3 维空间信息，捕捉 3 维点云的局部结构，并且能够集中学习手部点云的有效特征，从而进行精确的 3 维手部姿态估计.文献[90]与基于体素化的方法相比，需要更少的内存.但与基于深度图像的方法相比，需要更多的计算时间和内存.文献[91]使用半监督的方式对网络进行训练，其性能可与全监督的方法相媲美.文献[92]有效利用深度图中的信息，以较少的网络参数捕获更多的手部细节及结构，并准确地估计其 3 维姿态.文献[93]首次结合结构化手部姿势估计与基于 transformer 的自然语言处理领域的转换框架.引入参考手部姿势为输出关节提供等效依赖关系.文献[89]的模型大小为 17.2 MB.其中 11.1 MB 用于点对点回归网络，它是分层 PointNet；6.1 MB 用于附加的回归模块，它由 3 个全连层组成.文献[90]有 2 种版本，回归版本为 38 MB，检测版本为 44 MB.文献[91]中，手部特征编码器(hand feature encoder, HFE)、手部特征解码器(hand feature decoder, HFD)和手部特征估计器(hand pose estimator, HPE)的大小分别为 8.1 MB，74 MB，8.5 MB.由于只在测试阶段使用 HFE 和 HPE，所以其网络模型大小为 16.6 MB.文献[92]的模型大小为 10.3 MB，其中回归网络为 9.2 MB，指尖精细网络为 1.1 MB.不同方法在 3 个数据集上的性能对比分析如图 1 所示：



(a) ICVL数据集

(b) MSRA数据集

(c) NYU数据集

Fig. 1 Performance comparison of hand pose estimation methods

图 1 手部姿态估计方法的性能对比

7 总 结

本文总结了近年来多种点云处理任务的方法，特别侧重于基于深度学习的工作，为读者提供了最新的研究进展.

大多数关于点云的综述类文章都集中于讨论点云分类分割处理任务.如文献[94-95]只讨论了语义分割任务；文献[96-97]增加了目标检测和分类任务的研究分析.其中，文献[97]只用 1 节内容简要介绍分类、分割及目标检测三大任务，更关注于处理点云

数据的深度学习方法,而不依据处理任务对其进行划分讨论.本文则考虑多种点云处理任务,包括模型重建与变换、分类分割、检测跟踪与姿态估计等.在模型分割分类中,由于大部分算法有用于实现点云分类与分割的功能,不同于文献[96-97]将分类与分割作为2种类别分开讨论,本文将它们统一考虑,并根据基于体素、基于视图与基于点三大主流方法对其划分并展开讨论,明确给出各算法可处理的任务.

目前,已经有大量学者对点云处理任务进行研究并依据任务的不同提出多种方法,但这些方法或多或少都有一定的局限性.本文基于这些算法的不足总结点云处理任务所面临的挑战与发展趋势.

1) 数据方面

大部分方法只在现有的数据集上进行实验,而对于新获取的数据并不适用.这很大程度上是由于新获取的数据无法实现多角度、全方位的完美匹配,而且不同平台获得的数据难以融合,无法达到统一的标准.对于融合后的点云,具有鲁棒性和区分性特征的提取有一定的难度,未来的研究可以从特征提取方面入手.

数据集尺度不均衡是由于真实复杂场景中检测及识别小目标较为困难.未来研究工作可人工生成小目标样本,增大数据集中小目标所占比例,进而在网络训练中提高其识别检测能力.

数据质量对网络(如transformers)的泛化性和鲁棒性的影响较大[2].点云的几何位置存在误差时,可以通过已知控制点对其进行几何矫正.当使用激光扫描获取数据时,除了考虑扫描距离和入射角度的问题,还可以进行强度矫正,通过不同方法改善点云的质量.

随着3维扫描技术的发展,大规模点云的获取已不是难点,挑战性在于如何对其进行处理.此外,算法精度依赖大批量的数据集[98],目前还没有比较好的解决手段.

2) 性质方面

点云是3维空间内点的集合,它没有提供邻域信息,故而大部分方法需要依据不同的邻域查询方法确定点的邻域,这将导致算法增加额外的计算成本.点云不能显式地表达目标结构以及空间拓扑关系.此外,当目标被遮挡或重叠时,不能依据几何关系确定拓扑结构,给后续处理任务带来一定难度.

针对点云的不规则性及无序性,将其应用于深度神经网络中进行相关任务的处理需要做数据形式的转换,如体素化[40].但这些转换操作不但增加了计算量,而且很可能在转换的过程中丢失信息,所以直接的点云处理方法是重要的研究方向.

3) 网络结构方面

① 基于快速和轻量级的模型.为了达到理想效果,目前的算法倾向于使用含大量参数的较大的神经网络结构,导致计算复杂度高、内存占用大、速度慢等问题.因此,设计快速且轻量级的网络架构具有较大的应用价值[99-100].

② 网络结构的改良.优化网络结构可使同一网络处理多种任务,能够很大程度地降低复杂度[2].还可以考虑与其他网络结构结合[45]来实现优化目的.

4) 应用方面

室外场景信息较多、结构复杂,所以目前大多数方法着重于相对简单的室内场景的分析.然而自动驾驶[12]等技术的研究无法在室内场景中完成,所以未来的研究方向可侧重于构建适用于室外场景的网络模型.

现有分割方法大都用于单个物体的部件分割[1]或场景中同类对象的语义分割[25].而真实场景中目标类别众多、结构复杂,对同类对象的不同个体分割是3维形态检测(文物、古建监测)的重要手段.

现有的大多数算法主要利用静态场景中获取的数据,在地震检测等实际应用中,设计能够应对变化场景的算法具有重要应用价值.利用时序上下文信息可作为其研究方向[99].

计算机视觉中的有效性通常与效率相关,它决定模型是否可用于实际应用中[100],因此在二者之间实现更好的平衡是未来研究中有意义的课题.

作者贡献声明:李娇娇负责调研文献、撰写并修改全文;孙红岩负责检查论文并提出指导意见;董雨和张若晗负责检索、归纳、整理相关文献;孙晓鹏负责确定论文思路、设计文章框架.

参 考 文 献

[1] Charles R Q, Su Hao, Mo Kaichun, et al. PointNet: Deep learning on point sets for 3D classification and segmentation [C] //Proc of the 30th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 77-85

[2] Han Kai, Wang Yunhe, Chen Hanting, et al. A survey on visual transformer [J]. arXiv preprint, arXiv: 2012.12556, 2021

[3] LeCun Y, Bengio Y, Hinton G. Deep learning [J]. Nature, 2015, 521(7553): 436-444

[4] Lin H, Cong M N. Inference of 6-DOF robot grasps using point cloud data [C] //Proc of the 19th Int Conf on Control, Automation and Systems. Piscataway, NJ: IEEE, 2019: 944-948

[5] Khanh T T, Hoang Hai T, Nguyen V, et al. The practice of cloud-based navigation system for indoor robot [C/OL] //Proc of the 14th Int Conf on Ubiquitous Information Management and Communication. Piscataway, NJ: IEEE, 2020 [2021-02-03]. https://ieeexplore.ieee.org/document/9001709

[6] Lin Weiyang, Anwar A, Li Zhan, et al. Recognition and pose estimation of auto parts for an autonomous spray painting robot [J]. IEEE Transactions on Industrial Informatics, 2019, 15 (3): 1709-1719

[7] Parra C, Cebollada S, Payá L, et al. A novel method to estimate the position of a mobile robot in underfloor environments using RGB-D point clouds [J]. IEEE Access, 2020, 8: 9084-9101

[8] Yang Jubo, Kong Minxiu, Wang Rundong. A path planning algorithm of spray robot based on 3D point cloud [C] //Proc of the 2nd Int Conf on Communications, Information System and Computer Engineering. Piscataway, NJ: IEEE, 2020: 353-360

[9] Hanke T, Schaermann A, Geiger M, et al. Generation and validation of virtual point cloud data for automated driving systems [C/OL] //Proc of the 20th Int Conf on Intelligent Transportation Systems. Piscataway, NJ: IEEE, 2017[2021-02-03]. https://ieeexplore.ieee.org/document/8317864

[10] Josyula A, Anand B, Rajalakshmi P. Fast object segmentation pipeline for point clouds using robot operating system [C] // Proc of the 5th World Forum on Internet of Things. Piscataway, NJ: IEEE, 2019: 915-919

[11] Li Minghui, Zhang Yanning. 3D point cloud labeling tool for driving automatically [C] //Proc of the 12th Asia-Pacific Signal and Information Processing Association Annual Summit and Conf. Piscataway, NJ: IEEE, 2020: 1666-1672

[12] Lee S, Kim C, Cho S, et al. Robust 3-dimension point cloud mapping in dynamic environment using point-wise static probability-based NDT scan-matching [J]. IEEE Access, 2020, 8: 175563-175575

[13] Tredinnick R, Broecker M, Ponto K. Experiencing interior environments: New approaches for the immersive display of large-scale point cloud data [C] //Proc of the 22nd IEEE Virtual Reality. Piscataway, NJ: IEEE, 2015: 297-298

[14] Bonatto D, Rogge S, Schenkel A, et al. Explorations for real-time point cloud rendering of natural scenes in virtual reality [C/OL] //Proc of the 2016 Int Conf on 3D Imaging. Piscataway, NJ: IEEE, 2016 [2021-02-03]. https://ieeexplore.ieee.org/document/7823453

[15] Feichter S, Hlavacs H. Planar simplification of indoor point-cloud environments [C] //Proc of the 1st IEEE Int Conf on Artificial Intelligence and Virtual Reality. Piscataway, NJ: IEEE, 2018: 274-281

[16] Wirth F, Quehl J, Ota J, et al. PointAtMe: Efficient 3D point cloud labeling in virtual reality [C] //Proc of the 30th IEEE Intelligent Vehicles Symp. Piscataway, NJ: IEEE, 2019: 1693-1698

[17] Liu Weiquan, Lai Baiqi, Wang Cheng, et al. Learning to match 2D images and 3D LIDAR point clouds for outdoor augmented reality [C] //Proc of the 2020 IEEE Conf on Virtual Reality and 3D User Interfaces Abstracts and Workshops. Piscataway, NJ: IEEE, 2020: 654-655

[18] Im D, Kang S, Han D, et al. A 4.45 ms low-latency 3D point-cloud-based neural network processor for hand pose estimation in immersive wearable devices [C/OL] //Proc of the 2020 IEEE Symp on VLSI Circuits. Piscataway, NJ: IEEE, 2020 [2021-02-03]. https://ieeexplore.ieee.org/document/9162895

[19] Placitelli A P, Gallo L. 3D point cloud sensors for low-cost medical in-situ visualization [C] //Proc of the 2011 IEEE Int Conf on Bioinformatics and Biomedicine Workshops. Piscataway, NJ: IEEE, 2011: 596-597

[20] Haiderbhai M, Ledesma S, Navab N, et al. Generating X-ray images from point clouds using conditional generative adversarial networks [C] //Proc of the 42nd Annual Int Conf of the IEEE Engineering in Medicine and Biology Society. Piscataway, NJ: IEEE, 2020: 1588-1591

[21] Duan Yueqi, Zheng Yu, Lu Jiwen, et al. Structural relational reasoning of point clouds [C] //Proc of the 32nd IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2019: 949-958

[22] Deng Haowen, Birdal T, Ilic S. PPFNet: Global context aware local features for robust 3D point matching [C] //Proc of the 31st IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 195-205

[23] Shen Yiru, Feng Chen, Yang Yaoqing, et al. Mining point cloud local structures by kernel correlation and graph pooling [C] //Proc of the 31st IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 4548-4557

[24] He Tong, Huang Haibin, Yi Li, et al. GeoNet: Deep geodesic networks for point cloud analysis [C] //Proc of the 32nd IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2019: 6881-6890

[25] Charles R Q, Yi Li, Su Hao, et al. PointNet++: Deep hierarchical feature learning on point sets in a metric space [C] //Proc of the 31st Conf on Neural Information Processing Systems. Cambridge, MA: MIT Press, 2017: 5100-5109

[26] Huang Haibin, Kalogerakis E, Chaudhuri S, et al. Learning local shape descriptors from part correspondences with multiview convolutional networks [J]. ACM Transactions on Graphics, 2017, 37(1): 6:1-6:14

[27] Shafiq U, Taj M, Ali M. More for less: Insights into convolutional nets for 3D point cloud recognition [C] //Proc of the 24th IEEE Int Conf on Image Processing. Piscataway, NJ: IEEE, 2017: 1607-1611

[28] Klokov R, Lempitsky V. Escape from cells: Deep kd-networks for the recognition of 3D point cloud model [C] // Proc of the 16th IEEE Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2017: 863-872

[29] Xie Saining, Liu Sainan, Chen Zeyu, et al. Attentional ShapeContextNet for point cloud recognition [C] //Proc of the 31st IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 4606-4615

[30] Liu Yongcheng, Fan Bin, Xiang Shiming, et al. Relation-shape convolutional neural network for point cloud analysis [C] //Proc of the 32nd IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2019: 8887-8896

[31] Mo Kaichun, Guerrero P, Yi Li, et al. StructureNet: Hierarchical graph networks for 3D shape generation [J]. ACM Transactions on Graphics, 2019, 38(6): 242:1-242:19

[32] Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324

[33] Sarmad M, Lee H J, Kim Y M. RL-GAN-Net: A reinforcement learning agent controlled gan network for real-time point cloud shape completion [C] //Proc of the 32nd IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2019: 5891-5900

[34] Wang Yifan, Wu Shihao, Huang Hui, et al. Patch-based progressive 3D point set upsampling [C] //Proc of the 32nd IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2019: 5951-5960

[35] Huang Zitian, Yu Yikuan, Xu Jiawen, et al. PF-Net: Point fractal network for 3D point cloud completion [C] //Proc of the 33rd IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2020: 7659-7667

[36] Miao Yongwei, Liu Jiazong, Chen Jiahui, et al. Structure-preserving shape completion of 3D point clouds with generative adversarial network [J]. SCIENTIA SINICA Informationis, 2020, 50(5): 675-691 (in Chinese)
(缪永伟, 刘家宗, 陈佳慧, 等. 基于生成对抗网络的点云形状保结构补全[J]. 中国科学: 信息科学, 2020, 50(5): 675-691)

[37] Nguyen D, Choi S, Kim W, et al. GraphX-convolution for point cloud deformation in 2D-to-3D conversion [C] //Proc of the 17th IEEE/CVF Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2019: 8627-8636

[38] Choi S, Nguyen A, Kim J, et al. Pointcloud deformation for single image 3D reconstruction [C] //Proc of the 26th IEEE Int Conf on Image Processing. Piscataway, NJ: IEEE, 2019: 2379-2383

[39] Yin Kangxue, Huang Hui, Daniel C, et al. P2P-NET: Bidirectional point displacement for shape transform [J]. ACM Transactions on Graphics, 2018, 37(4): 152:1-152:13

[40] Han Zhizhong, Wang Xiyang, Liu Yushen, et al. Multi-angle point cloud-VAE: Unsupervised feature learning for 3D point clouds from multiple angles by joint self-reconstruction and half-to-half prediction [C] //Proc of the 17th IEEE/CVF Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2019: 10441-10450

[41] Wang Weiyue, Ceylan D, Mech R, et al. 3DN: 3D deformation network [C] //Proc of the 32nd IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2019: 1038-1046

[42] Zhou Baoxing, Shan Yuhui, Yao Shoufeng. Deformation monitoring of similar material model based on point cloud feature extraction [J]. The Journal of Engineering, 2019, 2019(20): 6790-6794

[43] Qi C R, Su Hao, Nießner M, et al. Volumetric and multi-view CNNs for object classification on 3D data [C] //Proc of the 29th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 5648-5656

[44] Tchapmi L, Choy C, Armeni I, et al. SEGCloud: Semantic segmentation of 3D point clouds [C] //Proc of the 5th Int Conf on 3D Vision. Piscataway, NJ: IEEE, 2017: 537-547

[45] Wang Pengshuai, Liu Yang, Guo Yuxiao, et al. O-CNN: Octree-based convolutional neural networks for 3D shape analysis [J]. ACM Transactions on Graphics, 2017, 36(4): 72:1-72:11

[46] Graham B, Engelcke M, Maaten L V D. 3D semantic segmentation with submanifold sparse convolutional networks [C] //Proc of the 31st IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 9224-9232

[47] Meng H, Gao Lin, Lai Yukun, et al. VV-Net: Voxel vae net with group convolutions for point cloud segmentation [C] //Proc of the 17th IEEE/CVF Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2019: 8499-8507

[48] Shao Tianjia, Yang Yin, Weng Yanlin, et al. H-CNN: Spatial hashing based CNN for 3D shape analysis [J]. IEEE Transactions on Visualization and Computer Graphics, 2020, 26(7): 2403-2416

[49] Lefebvre S, Hoppe H. Perfect spatial hashing [J]. ACM Transactions on Graphics, 2006, 25(3): 579-588

[50] Lawin F J, Danelljan M, Tosteberg P, et al. Deep projective 3D semantic segmentation [C] //Proc of the 17th Int Conf on Computer Analysis of Images and Patterns. Berlin: Springer, 2017: 95-107

[51] Zhou Weiguo, Jiang Xin, Liu Yunhui. MVPointNet: Multi-view network for 3D object based on point cloud [J]. IEEE Sensors Journal, 2019, 19(24): 12145-12152

[52] Guerry J, Boulch A, Saux B L, et al. SnapNet-R: Consistent 3D multi-view semantic labeling for robotics [C] //Proc of the 16th IEEE Int Conf on Computer Vision Workshops. Piscataway, NJ: IEEE, 2017: 669-678

[53] Jaritz M, Gu Jiayuan, Su Hao. Multi-view pointnet for 3D scene understanding [C] //Proc of the 17th IEEE/CVF Int Conf on Computer Vision Workshops. Piscataway, NJ: IEEE, 2019: 3995-4003

[54] Yang Ze, Wang Liwei. Learning relationships for multi-view 3D object recognition [C] //Proc of the 17th IEEE/CVF Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2019: 7504-7513

[55] Atzmon M, Maron H, Lipman Y, et al. Point convolutional neural networks by extension operators [J]. ACM Transactions on Graphics, 2018, 37(4): 71:1-71:12

[56] Hermosilla P, Ritschel T, Vázquez P, et al. Monte Carlo convolution for learning on non-uniformly sampled point clouds [J]. ACM Transactions on Graphics, 2018, 37(6): 235:1-235:12

[57] Zhai Ruifeng, Li Xueyan, Wang Zhenxin, et al. Point cloud classification model based on a dual-input deep network framework [J]. IEEE Access, 2020, 8: 55991-55999

[58] Bai Jing, Xu Haojun. MSP-Net: Multi-scale point cloud classification network [J]. Journal of Computer-Aided Design & Computer Graphics, 2019, 31(11): 1927-1934 (in Chinese)
(白静, 徐浩钧. MSP-Net: 多尺度点云分类网络[J]. 计算机辅助设计与图形学学报, 2019, 31(11): 1927-1934)

[59] Hu Qingyong, Yang Bo, Xie Linhai, et al. RandLA-Net: Efficient semantic segmentation of large-scale point clouds [C] //Proc of the 33rd IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2020: 11105-11114

[60] Du Jing, Cai Guorong. Point cloud semantic segmentation method based on multi-feature fusion and residual optimization [J]. Journal of Image and Graphics, 2021, 26(5): 1105-1116 (in Chinese)
(杜静, 蔡国榕. 多特征融合与残差优化的点云语义分割方法[J]. 中国图象图形学报, 2021, 26(5): 1105-1116)

[61] Wu Wenxuan, Qi Zhongang, Li Fuxin. PointConv: Deep convolutional networks on 3D point clouds [C] //Proc of the 32nd IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2019: 9613-9622

[62] Wang Yue, Sun Yongbin, Liu Ziwei, et al. Dynamic graph CNN for learning on point cloud [J]. ACM Transactions on Graphics, 2019, 1(1): 1:1-1:13

[63] Komarichev A, Zhong Zichun, Hua Jing. A-CNN: Annularly convolutional neural networks on point clouds [C] //Proc of the 32nd IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2019: 7413-7422

[64] Zhang Zhiyuan, Hua B, Yeung S. ShellNet: Efficient point cloud convolutional neural networks using concentric shells statistics [C] //Proc of the 17th IEEE/CVF Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2019: 1607-1616

[65] Zhao Hengshuang, Jiang Li, Jia Jiaya, et al. Point transformer [J]. arXiv preprint, arXiv: 2012.09164, 2020

[66] Guo Menghao, Cai Junxiong, Liu Zhengning, et al. PCT: Point cloud transformer [J]. Computational Visual Media, 2021, 7(2): 187-199

[67] Yan Zihao, Hu Ruizhen, Yan Xingguang, et al. RPM-Net: Recurrent prediction of motion and parts from point cloud [J]. ACM Transactions on Graphics, 2019, 38(6): 240:1-240:15

[68] Wang Sukai, Sun Yuxiang, Liu Chengju, 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

[69] Giancola S, Zarzar J, Ghanem B. Leveraging shape completion for 3D siamese tracking [C] //Proc of the 32nd IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2019: 1359-1368

[70] Achlioptas P, Diamanti O, Mitliagkas I, et al. Learning representations and generative models for 3D point clouds [C] //Proc of the 35th Int Conf on Machine Learning. New York: ACM, 2018: 40-49

[71] Burnett K, Samavi S, Waslander S, et al. aUToTrack: A lightweight object detection and tracking system for the SAE autodrive challenge [C] //Proc of the 16th Conf on Computer and Robot Vision. Piscataway, NJ: IEEE, 2019: 209-216

[72] Simon M, Amende K, Kraus A, et al. Complexer-YOLO: Real-time 3D object detection and tracking on semantic point clouds [C] //Proc of the 32nd IEEE/CVF Conf on Computer Vision and Pattern Recognition Workshops. Piscataway, NJ: IEEE, 2019: 1190-1199

[73] Behl A, Paschalidou D, Donné S, et al. PointFlowNet: Learning representations for rigid motion estimation from point clouds [C] //Proc of the 32nd IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2019: 7954-7963

[74] Gu Xiuye, Wang Yijie, Wu Chongruo, et al. HPLFlowNet: Hierarchical permutohedral lattice FlowNet for scene flow estimation on large-scale point clouds [C] //Proc of the 32nd IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2019: 3249-3258

[75] Liu Xingyu, Qi C R, Guibas L J, et al. FlowNet3D: Learning scene flow in 3D point clouds [C] //Proc of the 32nd IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2019: 529-537

[76] Shi Shaoshuai, Wang Xiaogang, Li Hongsheng. PointRCNN: 3D object proposal generation and detection from point cloud [C] //Proc of the 32nd IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2019: 770-779

[77] Lang A H, Vora S, Caesar H, et al. PointPillars: Fast encoders for object detection from point clouds [C] //Proc of the 32nd IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2019: 12689-12697

[78] Yang Zetong, Sun Yanan, Liu Shu, et al. STD: Sparse-to-dense 3D object detector for point cloud [C] //Proc of the 17th IEEE/CVF Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2019: 1951-1960

[79] Liu Zhe, Zhou Shunbo, Suo Chuanzhe, et al. LPD-Net: 3D point cloud learning for large-scale place recognition and environment analysis [C] //Proc of the 17th IEEE/CVF Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2019: 2831-2840

[80] Paigwar A, Erkent O, Wolf C, et al. Attentional PointNet for 3D-object detection in point clouds [C] //Proc of the 32nd IEEE/CVF Conf on Computer Vision and Pattern Recognition Workshops. Piscataway, NJ: IEEE, 2019: 1297-1306

[81] Shi Shaoshuai, Guo Chaoxu, Jiang Li, et al. PV-RCNN: Point-voxel feature set abstraction for 3D object detection [C] //Proc of the 33rd IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2020: 10526-10535

[82] Elbaz G, Avraham T, Fischer A. 3D point cloud registration for localization using a deep neural network auto-encoder [C] //Proc of the 30th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 2472-2481

[83] Speciale P, Schönberger J L, Kang S B, et al. Privacy preserving image-based localization [C] //Proc of the 32nd IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2019: 5488-5498

[84] Zhang Weiyi, Qi Chenkun. Pose estimation by key points registration in point cloud [C] //Proc of the 3rd Int Symp on Autonomous Systems. Piscataway, NJ: IEEE, 2019: 65-68

[85] Deng Haowen, Birdal T, Ilic S. PPF-FoldNet: Unsupervised learning of rotation invariant 3D local descriptors [C] //Proc of the 15th European Conf on Computer Vision. Berlin: Springer, 2018: 620-638

[86] Deng Haowen, Birdal T, Ilic S. 3D local features for direct pairwise registration [C] //Proc of the 32nd IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2019: 3239-3248

[87] Kurobe A, Sekikawa Y, Ishikawa K, et al. CorsNet: 3D point cloud registration by deep neural network [J]. IEEE Robotics and Automation Letters, 2020, 5(3): 3960-3966

[88] Chen Xinghao, Wang Guijin, Zhang Cairong, et al. SHPR-Net: Deep semantic hand pose regression from point clouds [J]. IEEE Access, 2018, 6: 43425-43439

[89] Ge Liuhao, Ren Zhou, Yuan Junsong. Point-to-point regression pointnet for 3D hand pose estimation [C] //Proc of the 15th European Conf on Computer Vision. Berlin: Springer, 2018: 489-505

[90] Li Shile, Lee D. Point-to-pose voting based hand pose estimation using residual permutation equivariant layer [C] //Proc of the 32nd IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2019: 11919-11928

[91] Chen Yujin, Tu Zhigang, Ge Liuhao, et al. SO-HandNet: Self-organizing network for 3D hand pose estimation with semi-supervised learning [C] //Proc of the 17th IEEE/CVF Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2019: 6960-6969

[92] Ge Liuhao, Cai Yujun, Weng Junwu, et al. Hand PointNet: 3D hand pose estimation using point sets [C] //Proc of the 31st IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 8417-8426

[93] Huang Lin, Tan Jianchao, Liu Ji, et al. Hand-Transformer: Non-autoregressive structured modeling for 3D hand pose estimation [C] //Proc of the 16th European Conf on Computer Vision. Berlin: Springer, 2020: 17-33

[94] Zhang Jiaying, Zhao Xiaoli, Chen Zheng, et al. A review of deep learning-based semantic segmentation for point cloud [J]. IEEE Access, 2019, 7: 179118-179133

[95] Xie Yuxing, Tian Jiaojiao, Zhu Xiaoxiang. Linking points with labels in 3D: A review of point cloud semantic segmentation [J]. IEEE Geoscience and Remote Sensing Magazine, 2020, 8(4): 38-59

[96] Guo Yulan, Wang Hanyun, Hu Qingyong, et al. Deep learning for 3D point clouds: A survey [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(12): 4338-4364

[97] Bello S A, Yu Shangshu, Wang Cheng. Review: Deep learning on 3D point clouds [J]. arXiv preprint, arXiv: 2001.06280, 2020

[98] Xu Bingbing, Cen Keting, Huang Junjie, et al. A survey on graph convolutional neural network [J]. Chinese Journal of Computers, 2020, 43(5): 755-780 (in Chinese)
(徐冰冰, 岑科廷, 黄俊杰, 等. 图卷积神经网络综述[J]. 计算机学报, 2020, 43(5): 755-780)

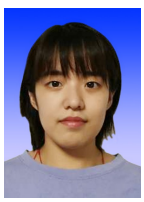
[99] Zhang Rui, Li Jintao. A survey on algorithm research of scene parsing based on deep learning [J]. Journal of Computer Research and Development, 2020, 57(4): 859-875 (in Chinese)
(张蕊, 李锦涛. 基于深度学习的场景分割算法研究综述[J]. 计算机研究与发展, 2020, 57(4): 859-875)

[100] Zhao Yongqiang, Rao Yuan, Dong Shipeng, et al. Survey on deep learning object detection [J]. Journal of Image and Graphics, 2020, 25(4): 629-654 (in Chinese)
(赵永强, 饶元, 董世鹏, 等. 深度学习目标检测方法综述[J]. 中国图象图形学报, 2020, 25(4): 629-654)

Li Jiaojiao, born in 1997. Master candidate. Her main research interest is computer graphics.
李娇娇,1997年生.硕士研究生.主要研究方向为计算机图形学.

Sun Hongyan, born in 1968. Associate professor. Her main research interests include computer graphics, virtual reality.
孙红岩,1968年生.副教授.主要研究方向为计算机图形学、虚拟现实.

Dong Yu, born in 1998. Master candidate. His main research interest is computer graphics.
董　雨,1998年生.硕士研究生.主要研究方向为计算机图形学.



Zhang Ruohan, born in 1998. Master candidate. Her main research interest is computer graphics.
张若晗,1998年生.硕士研究生.主要研究方向为计算机图形学.

Sun Xiaopeng, born in 1968. PhD, professor, PhD supervisor. Senior member of CCF. His main research interests include computer graphics, virtual reality, computer vision, etc.
孙晓鹏,1968年生.博士,教授,博士生导师,CCF高级会员.主要研究方向为计算机图形学、虚拟现实、计算机视觉等.