

基于深度学习的点云分类方法综述

文沛^{1,2}, 程英蕾^{1*}, 余旺盛¹

¹空军工程大学信息与导航学院, 陕西 西安 710077;

²中国人民解放军 93575 部队, 河北 承德 067000

摘要 点云作为一种重要的 3D 数据类型,随着 3D 采集技术的发展已被广泛用于多个应用场景。深度学习因其处理大型数据集的高效性、提取特征的自主性,成为点云分类研究的主导方法。首先对点云分类方法的研究现状进行了介绍,接着重点对基于深度学习的点云分类的主要方法和最新方法进行了阐述。根据数据处理方式对点云分类方法进行归类,总结对比了每类方法的主要思想和优缺点,并详细介绍了部分代表性、创新性算法的实现过程。最后,对点云分类面临的挑战及未来研究方向进行了展望。

关键词 图像处理; 点云分类; 深度学习; 卷积神经网络; 语义分割

中图分类号 TP391

文献标志码 A

doi: 10.3788/LOP202158.1600003

Point Cloud Classification Methods Based on Deep Learning: A Review

Wen Pei^{1,2}, Cheng Yinglei^{1*}, Yu Wangsheng¹

¹Information and Navigation College, Air Force Engineering University, Xi'an, Shaanxi 710077, China;

²The 93575 Unit of PLA, Chengde, Hebei 067000, China

Abstract As an important three-dimensional (3D) data type, point cloud has been widely used in many applications with the development of 3D acquisition technology. Owing to its high efficiency in processing large-scale data sets and the autonomy of extracting features, deep learning has become the leading method for investigating the latest studies in a point cloud classification. This paper introduces the current research status of the point cloud classification methods. Furthermore, some main and latest methods of point cloud classification based on deep learning are analyzed and classified according to the data processing method. Additionally, this paper summarizes the key ideas, advantages, and disadvantages of each type of method and discusses the realization process of some representative and innovative algorithms in detail. Finally, the challenges and future research directions of the point cloud classification are outlined.

Key words image processing; point cloud classification; deep learning; convolutional neural network; semantic segmentation

OCIS codes 100.3008; 100.4996; 100.6890

1 引 言

近年来,随着三维(3D)传感器技术的飞速发展,3D点云数据的获取变得越来越便捷。3D点云也因其丰富的几何、形状、尺度信息,在场景理解中

发挥重要作用。在众多的点云识别任务中,点云分类几十年来一直活跃于摄影测量和遥感界等研究领域,并已成为智能车辆^[1]、自动驾驶^[2]、电力线路检测^[3]、森林监测^[4]、三维重构^[5]、文化遗产保护^[6]等许多应用的重要组成部分。值得一提的是,相比于

收稿日期: 2020-10-19; 修回日期: 2020-11-24; 录用日期: 2020-12-03

基金项目: 国家自然科学基金青年科学基金项目(61703423)

通信作者: *2412288250@qq.com

摄影测量和遥感中的点云分类^[7-8],某些情况下称其为点标记^[9-10]而言,点云语义分割一词在计算机视觉中广泛使用,尤其在近几年比较活跃的深度学习^[11-14]领域中。尽管如此,由于 3D 点云的无序性、非结构化、密度分布不均,以及传感器噪声高、场景复杂等因素,3D 点云分类依然是一个极具挑战性的问题。

一般来说,点云分类任务分为两个步骤:从局部和全局提取具有代表性的点特征;利用学习到的特征将每个点划分为预定义的语义类别。早期的研究主要集中于用手工设计的规则提取特征,然后使用基于机器学习的分类器预测每个点的语义标签,如高斯混合模型^[15-16]、支持向量机(SVM)^[17-18]、AdaBoost^[19-20]、随机森林(RF)^[21-22]等,这些方法忽略了邻域点之间的关系,分类结果容易产生噪声。一些研究尝试通过综合上下文信息来解决此类问题,如条件随机场(CRF)^[7,23]、马尔可夫随机场(MRF)^[24-25]等,一定程度上提高了分类性能。然而,利用手工设计规则提取的特征表达能力较弱,模型的泛化能力在处理复杂场景(如大规模野外场景)时无法达到令人满意的效果;此外,最优分类器的选择和组合复杂,严重依赖专业经验,限制了此类方法应用于各种真实场景时的灵活性^[26]。

早在 1971 年,就有学者提出了基于数据分组处理方法(GMDH)训练的具有 8 层深度的神经网络模型^[27],后由于计算机的计算能力受限导致此类研究陷入低谷。随着深度学习的再度兴起,越来越多使用神经网络处理应用问题的模型[卷积神经网络(CNN)^[28]、递归神经网络^[29]、深度信念网络^[30]等]不断涌现并取得了可观效果,研究者们逐渐考虑基于深度学习来处理 3D 点云数据的有关问题。起初研究者尝试将点云转换为适合利用卷积神经网络进行处理的规则结构(即多视图和体素网格),但这些方法容易导致信息丢失和计算复杂的问题。随着 2017 年斯坦福大学 Charles 等^[11]提出的 PointNet 网络模型在对象分类、语义分割、场景解析等方面取得突破性进展,直接处理原始点云开始变得越来越受人们欢迎,并逐渐占据主导地位。

目前已有一些对基于深度学习的点云语义分割进行总结分析的综述性文章^[31-34]。本文在前人工作的基础上进行了丰富和完善,梳理总结了现有基于深度神经网络的代表性点云分类方法的基本思路和关键技术,并添加了最近提出的新方法,最后对点云分类技术未来的研究方向做出了展望。

2 点云处理介绍

2.1 点云分割与点云分类

点云分割与分类的研究由来已久,涉及多个应用场景和领域。为避免混淆,本节就点云分割与分类的区别和联系进行简要说明。

点云分割是将原始 3D 点云按照相似属性进行分组的过程,而点云分类是为每个点或组分配语义标签的过程。对二者而言,特征提取过程相似,很多网络并未将其完全割裂,在共用一套特征提取方法后再区分具体的分割任务和分类任务。不同的是:点云分割旨在将原始 3D 点云分割为不重叠的区域,这一过程不需要监督的先验知识,因此传递的结果没有强大的语义信息;而点云分类通常通过有监督的学习方法来实现,包括传统有监督的机器学习和深度学习^[31],是用某个类标记每个点或分割后的每个组,并为其赋予某种语义,因此点云分类也称为点云语义分割或点标记。

2.2 评价指标

为了定量评估点云分类算法的性能,研究者们提出多种评价指标,包括执行时间、空间复杂度、精度,其中精度是关键指标,一般采用准确率(Accuracy)、精确率(Precision)、召回率(Recall)来评价。准确率指正确预测的样本数占总预测样本数的比值;精确率指正确预测的正样本数占所有预测为正样本数量的比值;召回率指正确预测的正样本数占真正样本总数的比值,计算公式为

$$V_{\text{acc}} = \frac{N_{\text{TP}} + N_{\text{TN}}}{N_{\text{TP}} + N_{\text{TN}} + N_{\text{FP}} + N_{\text{FN}}}, \quad (1)$$

$$V_{\text{pre}} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FP}}}, \quad (2)$$

$$V_{\text{rec}} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FN}}}, \quad (3)$$

式中: V_{acc} 为准确率的值; V_{pre} 为精确率的值; V_{rec} 为召回率的值; N_{TP} 为预测为正例,实际为正例的样本数; N_{TN} 为预测为正例,实际为负例的样本数; N_{FP} 为预测为负例,实际为负例的样本数; N_{FN} 为预测为负例,实际为正例的样本数。

目前主流的点云分类精度衡量指标包括总体精度(OA)、平均精度(MA)、平均交并比(mIoU),本文将选取这三种评价指标对不同分类方法进行对比,以供相关研究者参考分析。

3 基于深度学习的点云分类方法

近年来,研究者们提出大量的基于深度学习的

模型来处理点云分类问题,这些模型较传统算法性能更优。本章根据三维点云数据处理方式,将基于深度学习的点云分类方法分为基于规则表示的点云分类方法和基于原始点云的分类方法。

3.1 基于规则表示的点云分类方法

将不规则的点云进行规则化表示,通常有两种方法:1)多视图投影,即通过多视角投影到二维图像;2)体素化,即利用占用网格将环境状态表示为 3D 网格。

3.1.1 基于多视图的点云分类方法

早在 1995 年, Murase 等^[35]在视觉识别领域通过自动代表姿势和照明参数获得大量二维(2D)图像,进而获得低维子空间。在地理信息科学中, Haala 等^[36]将机载激光扫描提供的数据与建筑物

现有的 2D 平面图结合,实现了 3D 数据的自动捕获。受 2D 图像的启发, Su 等^[37]首次在 2015 年提出多视图卷积神经网络(MVCNN)。其主要思想为设置不同位置、角度的视点(虚拟相机)捕获多个视角下的 2D 视图,并将其分别通过卷积层(CNN1),然后通过池化层(Pooling)聚合成紧凑的 3D 形状描述符,最后将聚合后的特征输入网络剩余部分,返回分割和分类结果, MVCNN 原理如图 1 所示。虽然与传统方法相比, MVCNN 在分割和分类任务中表现出较好的效果,计算效率有所提升,但视点位置事先设定使其无法动态选择视图;同时,由于忽略了大量关键的几何空间信息, MVCNN 分割和分类的精度也受到影响,不适用于大规模复杂场景。

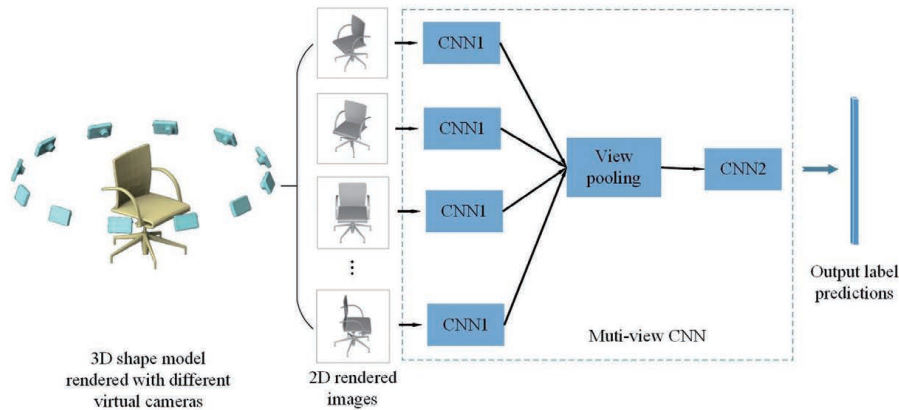


图 1 MVCNN 点云分类和分割框架图

Fig. 1 Architecture of MVCNN for point cloud classification and segmentation

在多视图处理过程中,不同视图对最终的形状描述符贡献不同,相似视图做出的贡献类似,而差异大的视图更有利于形状识别。为进一步挖掘更具区分性的信息, Feng 等^[38]在 MVCNN 的基础上提出了组视图卷积神经网络(GVCNN)。其主要思想为先用全卷积网络(FCN)提取视图级别描述符,再由分组模块学习视图之间的关联信息和区分信息,并根据每个视图基于内容的区分度将多视图划分为不同组,生成组级别描述符,然后进一步加权生成形状级别描述符,最后输入全连接层(FC)完成分类任务, GVCNN 原理如图 2 所示。通过使用分层的“视图-组-形状”描述框架,可以在组级别中发现重要且更具区分性的信息,并在形状描述符中予以强调,分类性能显著提升,但与其他基于多视图的方法一样,该方法依赖于视点位置和角度的选择。

由于 MVCNN 在进行多视图的聚类操作时采

用最大池化,忽略了其他非最大元素,部分有用信息丢失。为了克服这一问题, Wang 等^[39]设计了一个包含循环聚类池化(RCP)模块的卷积神经网络(RCPCNN),模块中引入了显性集概念。在视图相似度的基础上找出优势集,执行显性集聚类,然后以循环方式在池化层中多次迭代更新。采用循环聚类池化模块使 RCP 模块能对特征向量本身进行迭代完善,具有一定的自适应性,一定程度上提升了分类性能。

Yu 等^[40]则提出了一种多视图协调双线性网络(MHBN),借助不同视图之间可能局部存在的补充信息,计算相关补丁对相似度,解耦不相关补丁对,用双线性池化来获得合理的相似度评分,用相似度来度量每个匹配对中两个对象之间的相关性,将协调后的双线性池串联到网络层中,构建多视图协调双线性网络, MHBN 原理如图 3 所示。与其他方式相比, MHBN 采用双线性池,有效抑制了不良匹配

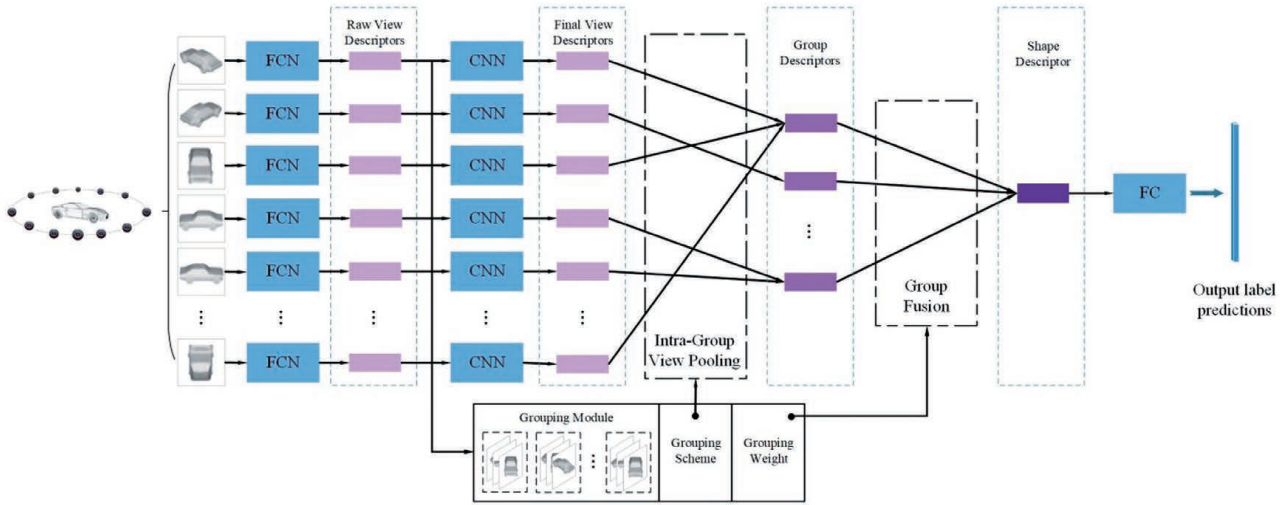


图 2 GVCNN 点云分类和分割框架图

Fig. 2 Architecture of GVCNN for point cloud classification and segmentation

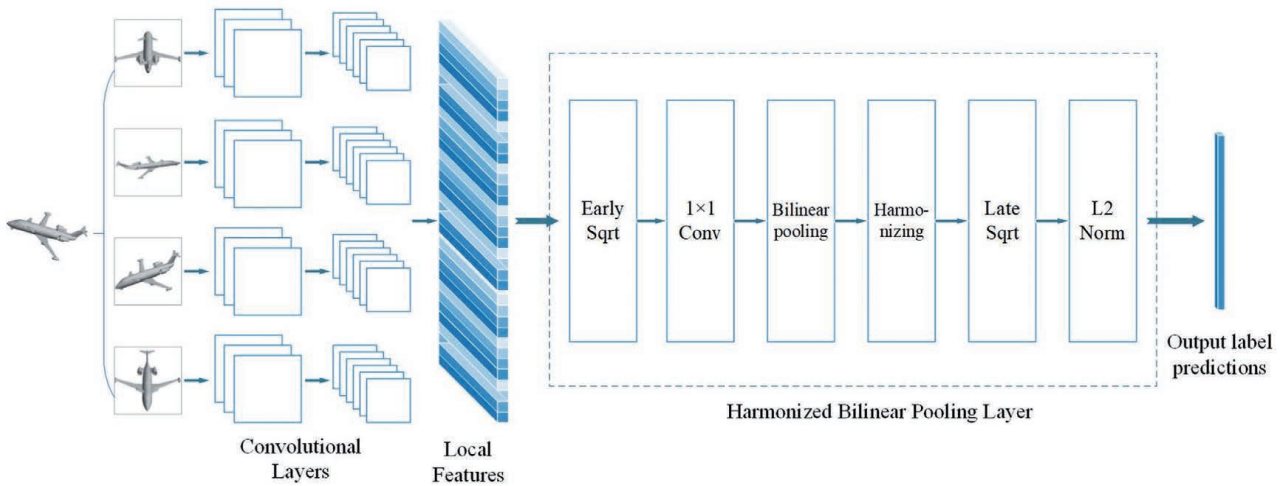


图 3 MHBN 点云分类和分割框架图

Fig. 3 Architecture of MHBN for point cloud classification and segmentation

对,且不受分类器选择的约束,对合并的双线性特征进行协调运算,获得了更具判别性的 3D 对象表示,在极大提高计算和存储效率的同时也显著提高了分类精度。

随着 RGB-D 传感器的发展和应用^[41],对 RGB-D 数据的处理也引起了研究者的关注。Boulch 等^[42]提出了 SnapNet 网络,选择不同视点生成 RGB 视图和包含几何特征的深度融合视图,然后使用全卷积网络对每对 2D 视图进行逐像素的语义标记,最后将像素标记反投影到原始的 3D 点云,在设计选择视图策略时关注多个局部视图,并根据最终投票结果来汇总局部预测,避免了过多的数据集误差,改善了信息丢失的问题,提高了分类精度。针对机器人上的应用,Boulch 等对 SnapNet 进行了扩展和改进,设计了 SnapNet-R 网络^[43],进一步改善了

分类性能,并提出了适用于 RGB-D 单视图数据的投影策略。

为充分利用多视图之间的相关性,Ma 等^[44]将长短期记忆(LSTM)网络与 CNN 结合在一起,将从不同视点渲染得到的多个视图图像构建成一个视图序列,并使用 2D CNN 网络提取视图序列中每个图像的低级特征,将其视为时间序列输入 LSTM 网络,再通过序列投票层将其聚合为形状描述符。该模型充分发挥了 CNN 和 LSTM 的优势,有效提升了多视图形状描述符的判别能力。此外,Biasutti 等^[45]提出先根据给定 3D 邻域点提取局部特征,再将这些特征投影到 2D 空间,构建了 LU-Net 模型,得到高质量分类结果。最近,Inigo 等提出的 3D-MiniNet^[46]先通过多视图投影从 3D 数据中获取局部和全局上下文信息,再输入 2D 全卷积神经网络

(FCNN)预测语义标签,最后将预测的 2D 语义标签重新投影至 3D 空间,以更快更高效的方式获得了高质量的结果。

尽管与传统的人工提取特征分类相比,将点云投影为多视图的方法在点云分割和分类任务中取得了良好的效果,但仍存在几个方面的明显缺陷。首先,对 3D 点云从多个角度进行投影,生成多个 2D 视图时,视点需预先设定,这意味着视点位置和角度的选择会对最终的分类结果造成很大影响,一定程度上并未摆脱人工设置的影子;其次,2D 投影仅限于对对象表面进行建模,无法捕获 3D 内部结构,一定程度上破坏了 3D 数据的内在几何关系,无法充分利用三维空间的上下文信息,造成信息丢失,且不适用于大规模场景;此外,基于多视图的方法在提取特征时通常需要进行预训练和微调,增加了工作负担。

3.1.2 基于体素的点云分类方法

最早的将非结构化的点云几何数据体素化为网格数据并应用于 3D 卷积神经网络的模型是 Maturana 等^[47]于 2015 年提出的 VoxNet,首先采用占用网格算法将环境状态表示为多个 3D 网格,每个网格对应一个体素,并将每个网格单元的每个值进行归一化处理,然后输入网络的卷积层来创建特征图,并对不重叠的体素块进行最大池化处理。Wu 等^[48]则以数据驱动的方式学习所有 3D 体素的复杂联合分布,提出了一种卷积深度信念网络 3D ShapeNet,将 3D 几何形状表示为 3D 体素网格上二进制变量的概率分布,每个 3D 网格都表示为二进制张量,然后输入三层卷积滤波器来提取特征。VoxNet 和 3D ShapeNet 都在一定程度上解决了点云非结构化的问题,但仍存在计算时占用内存大的问题,随着分辨率的增加甚至呈立方增长,结构的紧

凑表示问题亟待解决。

用灵活的八叉树结构替代固定分辨率的体素是有效的方法之一。受双边卷积层(BCL)网格-八叉树混合结构应用的启发,Riegler 等^[49]考虑 3D 数据本质上的稀疏性,提出了一种具有自适应空间划分能力的 OctNet,将 3D 空间分层划分为一组不平衡的八叉树结构,每个叶节点都存储一个池化特征表示,依据数据密度对 3D 空间进行划分,最后根据输入的 3D 结构动态地集中计算存储资源,极大地降低了内存消耗,缩短了运行时间。Wang 等^[50]提出的 O-CNN 仅在 3D 形状表面边界所占据的稀疏八分圆上执行卷积操作,利用八叉树表示的稀疏性和形状的局部方向来合理分配内存并提高计算效率。Xu 等^[51]提出一种无需学习的 3D 点云分割策略(VGS),基于八叉树结构对点云体素化,在局部上下文信息的基础上使用图论进行体素和超体素的聚类,并结合感知分组定律以纯几何的方式对 3D 点云进行分割处理,该方法对复杂场景和具有非平面表面的对象具有较好的分类效果。虽然基于八叉树的方法减少了 3D 表示形式的内存占用量,但访问八叉树中的任意元素需要从根节点开始遍历,直到到达所需的单元格为止,这对于高分辨率的八叉树而言代价过于昂贵。

类似地,其他索引结构也被合理应用到识别任务中。Klokov 等^[52]提出使用 Kd-tree 来构建 3D 点云的结构图并设计了 Kd-Net,以前馈自下而上的方式计算层次表示。Zeng 等^[53]提出的 3D ContextNet 采用 Kd-tree 结构细分 3D 空间,然后使用局部信息和全局上下文信息对特征进行编码学习,再递归使用多层感知机(MLP)和池化操作来逐步进行抽象判别表示,完成特征聚合,其原理如图 4 所示。由于

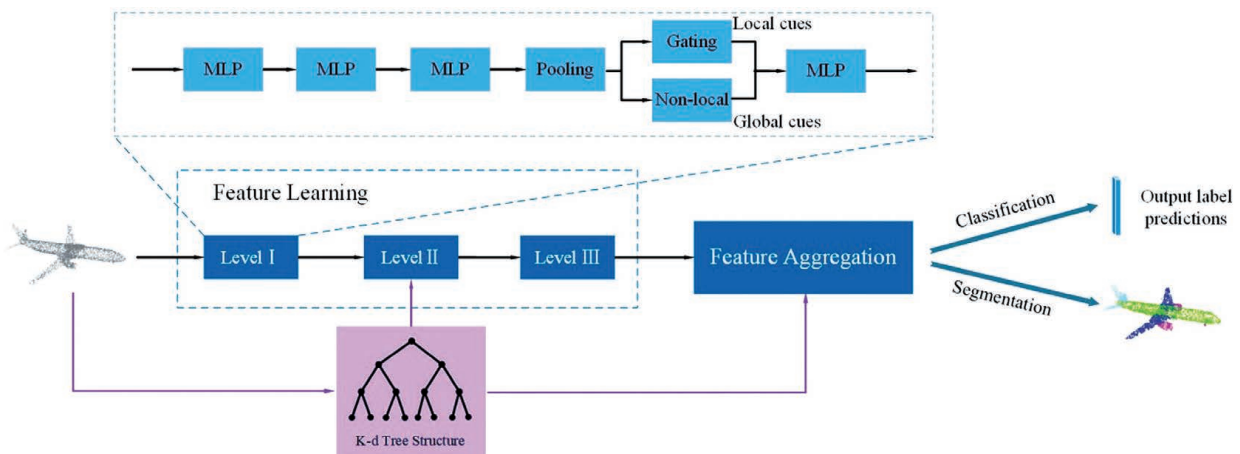


图 4 3D ContextNet 点云分类和分割框架图

Fig. 4 Architecture of 3D ContextNet for point cloud classification and segmentation

Kd-tree 具有较好的索引能力和结构化 3D 数据的能力, Kd-Net 和 3D ContextNet 在训练和测试时内存占用量均较少, 提高了计算效率。

基于八叉树和 Kd-tree 结构的方法虽然一定程度上减少了内存消耗, 但容易受体素边界的影响, 且未能充分利用局部几何结构。为此, Wang 等^[54]提出了多尺度体素化技术, 并进一步提出了多尺度卷积网络 MSNet, 将点云划分为多个尺度的体素, 从而基于多尺度体素提取不同空间分辨率的多尺度特征, 以预测类别概率, 同时, 构造具有空间一致性的 CRF 模型来获得全局上下文信息, 以实现全局优化, 最后组合局部特征和全局上下文信息, 实现对点云的准确分类。该方法在遥感、测绘数据等领域获得了不错的语义分割效果。

为了解决体素化网格内存消耗大、训练时间长的问题, 研究者们做了许多改进性工作。Li 等^[55]提出一种场探测模型 FPNN, 将 3D 空间表示为 3D 矢量场后作为网络的输入, 使用场探测滤波器从 3D

矢量场中有效地提取特征, 用场探测滤波器代替了 CNN 中的卷积层, 计算复杂度仅取决于场探测滤波器和采样点的个数, 不会随输入分辨率大小的改变而改变。Le 等^[56]提出一种由点和网格集成的混合模型 PointGrid, 在每个网格单元中使用简单的点量化策略对恒定数量的点进行采样, 使网络提取局部几何特征。PointGrid 具有比体积网格更好的伸缩性, 同时能更好地表示局部几何形状细节, 避免信息的丢失, 且无需高分辨率的网格即可良好地工作。Tchapmi 等^[57]以 3D-FCNN 为基础提出了 SEGCloud, 先对 3D 点云体素化并通过 3D-FCNN 生成粗略的下采样体素标签, 然后输入三线性插值层, 将粗略的体素标签插值回 3D 点, 最后将原始 3D 点特征与插值后所得类标签结合在一起, 使用全连接条件随机场 (FC-CRF) 进行最终类标签的推理, 获得更具细粒度的类分布, 并使用全卷积网络共享计算, 一定程度上减少了计算开销, 其原理如图 5 所示。

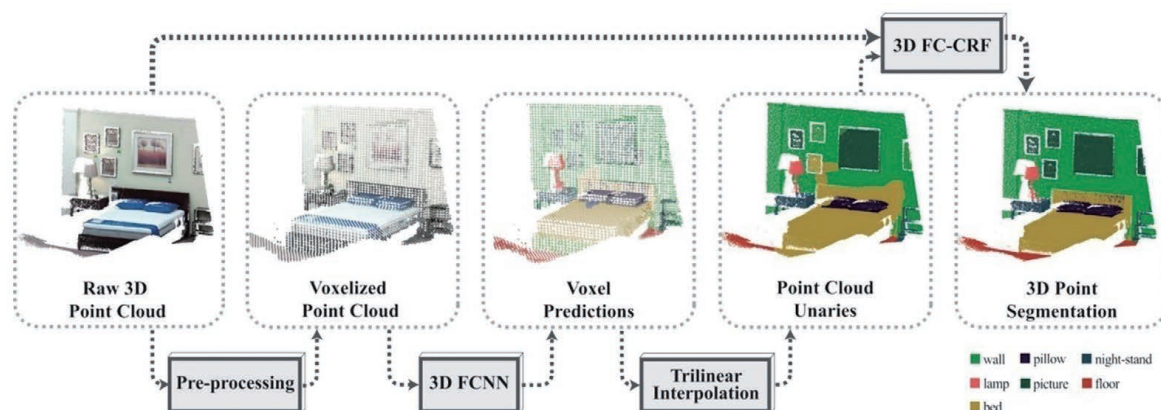


图 5 SEGCloud 点云语义分割框架图^[57]

Fig. 5 Architecture of SEGCloud for point cloud semantic segmentation^[57]

此外, Meng 等^[58]结合变分自动编码器 (VAE) 提出了 VV-NET, 将非结构化点云转换为体素网格后, 将每个体素进一步细分为子体素, 并对子体素内的稀疏点样本进行插值处理, 使用编码-解码的方式突破二进制体素的局限, 增强了捕获点分布的能力。Hegde 等^[59]结合了体素化网络和多视图网络, 提出 FusionNet 模型, 在全连接层融合两种网络, 其性能较单独网络有明显提升。

虽然将 3D 点云转换为体素网格解决了 3D 点云非结构化的问题, 较好地完成了分类分割任务, 但仍然存在以下几个方面的问题。首先, 点云从根本上来讲只是对象表面的点的集合, 并非整个实体对象的“3D 数据”, 对象表面以外的 3D 空间中存在大

量的“空数据”^[54], 简单地转换容易导致大量不必要的计算; 其次, 由于体素化之后的卷积操作通常需要规则的体素网格作为输入, 故体素大小成为限制整体精度的一个因素, 低分辨率的体素可能导致有用信息丢失, 而高分辨率的体素可能导致庞大复杂的计算量和严重的存储器占用情况。

表 1 汇总了基于规则化表示的点云分类方法的提出年份、核心思想、应用场景, 以及其在主要数据集上的表现性能。由表 1 可以看出, 基于规则化表示的方法大多在 ModelNet40/10 等数据集上实现, 并取得了较好的对象识别和分类效果, 其中 MHBN^[40]、文献^[44]、Kd-Net^[52]、PointGrid^[56]等算法表现突出。SnapNet^[42]、SEGCloud^[57]、VV-net^[58]

表 1 基于规则化表示的方法比较
Table 1 Comparison for methods based on regular representation

Method	Year	Key idea	Application scenario	Dataset	Accuracy / %		
					OA	MA	mIoU
MVCNN ^[37]	2015	Learning to recognize 3D shapes from a collection of their rendered views on 2D images	3D shape recognition	ModelNet40	90.10	—	—
RPCNN ^[38]	2017	Introducing a view clustering and pooling layer based on dominant sets	3D object recognition	ModelNet40	93.80	—	—
SnapNet ^[42]	2017	Transferring the very impressive results of 2D deep segmentation networks to 3D	3D semantic segmentation	Semantic 3D SUN RGB-D	88.60 —	70.80 67.40	59.10 —
SnapNet-R ^[43]	2017	Using 3D-coherent synthesis of scene observations and mixing them in a multi-view framework for 3D labeling	Semantic labeling of the scene perceived by a robot	SUN RGB-D	78.04	—	39.61
GVCNN ^[38]	2018	Using a grouping strategy	3D shape classification and retrieval	ModelNet40	93.10	—	—
MHBN ^[40]	2018	Aggregating local convolutional features through bilinear pooling	3D object recognition	ModelNet40 ModelNet10	94.91 92.23	—	—
In the Ref. [44]	2019	Combining CNNs with LSTM to exploit the correlative information from multiple views	3D shape recognition and 3D shape retrieval	ModelNet40 ModelNet10	91.05 95.29	—	—
LU-Net ^[45]	2019	Embedding 3D local features in 2D range-images; using a U-Net	Solving the image processing problem of 3D LiDAR point cloud	KITTI	—	—	55.40
3D-MiniNet ^[46]	2020	Combining 3D and 2D learning layers; learning the 2D representation through a novel projection	Fast and efficient for 3D LiDAR point cloud	Semantic-KITTI KITTI	— —	—	55.80 58.00
VoxNet ^[47]	2015	Integrating a volumetric occupancy grid representation with a supervised 3D CNN	Real-time object recognition	ModelNet40 ModelNet10	85.90 —	83.00 92.00	— —
3D ShapeNet ^[48]	2015	Using a convolutional deep belief network to represent a geometric 3D shape as a probability distribution of binary variables on a 3D voxel grid	Joint object recognition and shape completion from 2.5D depth maps	ModelNet40 ModelNet10	84.70 —	77.30 83.50	— —

表 1(续)

Method	Year	Key idea	Application scenario	Dataset	Accuracy / %		
					OA	MA	mIoU
OctNet ^[49]	2016	Using a set of unbalanced octrees to exploit the sparsity in the input data to hierarchically partition the space	3D object classification, orientation estimation and point cloud labeling	ModelNet40	86.50	—	—
FPNN ^[55]	2016	Representing 3D spaces as volumetric fields, using field probing filters to extract features	3D object recognition	ModelNet40	87.50	—	—
FusionNet ^[59]	2016	Using both voxel and pixel representations for training relatively weak classifiers	3D CAD models classification	ModelNet40 ModelNet10	90.80 93.11	—	—
O-CNN ^[50]	2017	Storing the octant information and CNN features into the graphics memory and executing the entire O-CNN training and evaluation on the GPU	3D object classification, shape retrieval, and shape segmentation	ModelNet40	90.60	—	85.90
Kd-Net ^[52]	2017	Performing multiplicative transformations and sharing parameters according to the subdivisions of the point clouds imposed onto them by kd-trees	3D shape classification, and shape part segmentation	ModelNet40 ModelNet10	91.80 94.00	—	— 77.20
3D ContextNet ^[53]	2017	Exploiting the local and global contextual cues imposed by the implicit space partition of the K-d tree for feature learning	3D object classification and part segmentation	S3DIS	84.90	74.50	55.60
SEGCLOUD ^[57]	2017	Combining 3D-FCNN, trilinear interpolation (TI), and fully connected conditional random fields (FC-CRF).	3D semantic segmentation (indoor and outdoor scenes)	Semantic 3D S3DIS NYUv2 KITTI	— — 66.82 —	73.08 57.35 56.43 49.46	60.30 48.92 43.45 36.78
MSNet ^[54]	2018	Multi-scale voxelization	Adaptive and robust point cloud classification	MLS TLS ALS	83.18 98.24 97.02	— — —	— — —
PointGrid ^[56]	2018	Incorporating a constant number of points within each grid cell	3D visual recognition	ModelNet40 ShapeNet	92.00 86.10	88.90 80.50	— 86.40
VV-net ^[58]	2018	Using a kernel-based interpolated variational autoencoder (VAE) architecture to encode the local geometry within each voxel	3D object segmentation into parts and scenes segmentation into individual objects; normal estimation	ShapeNet S3DIS	— 87.78	— —	87.40 78.22

等算法选用了 Semantic 3D 或 S3DIS 数据集,可用于大规模室外或室内场景分割,而除 2018 年提出的 VV-net 网络性能较好外, SnapNet、SEGCloud 算法的 MA 和 mIoU 均较低,体现出基于规则表示的方法在大规模场景中表现性能的局限性。

3.2 基于原始点云的分类方法

为了降低计算复杂度,充分利用 3D 点云数据的特性,研究者们开始考虑直接对原始点云数据进行处理,Charles 等^[11]提出的 PointNet 网络开创了此类研究工作的先河。与先将点云规则化表示的方法不同,PointNet 直接将点云作为输入,使用 T-Net 模块对输入点矩阵进行仿射变换,使用共享 MLP 学习每点特征,使用最大池化层聚合全局特征,以较低的代价、较好的效果解决了点云无序性、置换不变性和旋转不变性的问题。尽管 PointNet 网络在点云分类或分割任务中取得了突破性进展,但仍存在

无法捕获局部邻域信息的缺陷。为此,研究者们提出了一系列解决方案,引入了许多专用神经模块。本节整理了 70 多篇代表文献并将其归为 4 类:1) 邻域特征池化;2) 图形信息传递;3) 基于内核的卷积;4) 基于注意力机制。

3.2.1 邻域特征池化的方法

为了进一步提取更具细粒度的局部特征, Qi 等^[60]基于 PointNet 进行扩展,提出了 PointNet++ 网络框架。其主要思想为在原网络中构建局部分层模块,每层由采样层、分组层、特征提取层(PointNet)组成,通过采样层选择局部邻域质心,通过分组层构建局部邻域子集,通过 PointNet 层获取局部邻域特征向量,沿多分辨率层次结构捕获局部特征,其原理如图 6 所示。PointNet++ 虽然一定程度上解决了非均匀点采样和局部特征提取的问题,但依然没有考虑点与点之间的结构信息(如距离、方向等)。

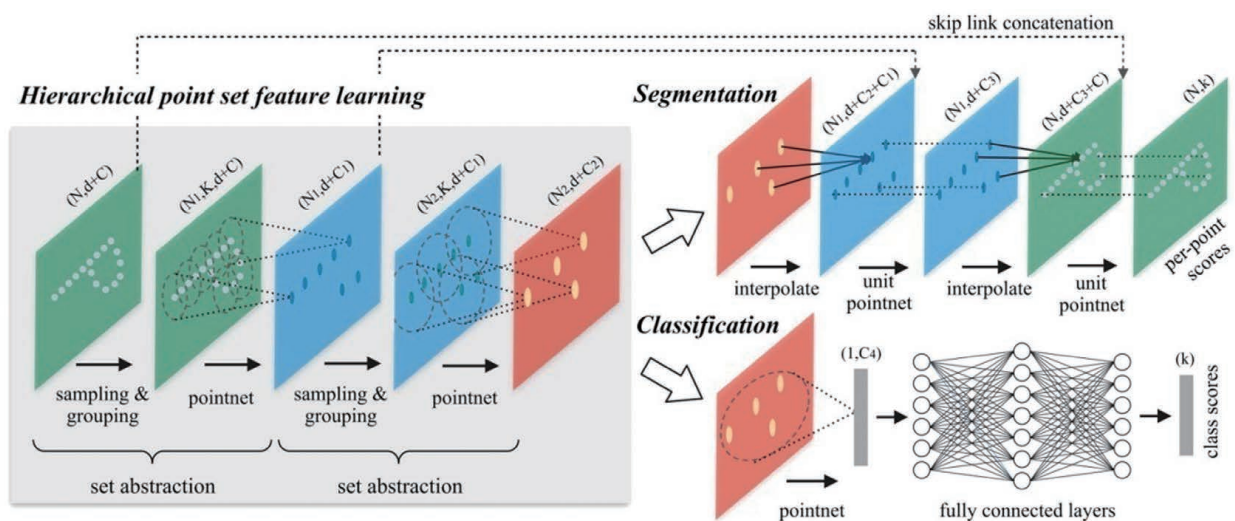


图 6 PointNet++ 点云分类和分割框架图^[60]

Fig. 6 Architecture of PointNet++ for point cloud classification and segmentation^[60]

为了同时实现方向感知和尺度感知, Jiang 等^[61]提出一种可嵌入各种基于 PointNet 网络的 PointSIFT 模块,该模块采用方向编码卷积(OEC)集成来自 8 个定向的信息,获得编码定向信息的表示,并通过堆叠多个方向编码单元来实现多尺度表示。PointSIFT 模块应用于整体框架的每一层,在下采样阶段将其与集合抽象模块(SA)结合用于层次特征提取,在上采样阶段将其与特征传播模块(FP)结合用于获取具有一定密度的特征,最后输入全连接层完成语义标签预测。Li 等^[62]通过构建自组织图(SOM)对点云的空间分布进行建模,提出了 SO-Net 网络,该网络先通过 SOM 生成输入点

云的二维表示并基于 SOM 节点进行归一化处理,然后将点特征聚合为 SOM 节点特征,再采用平均池化将 SOM 节点特征聚合为全局特征向量,最后通过一个具有完全连接分支和卷积分支的并行分支网络,从全局特征中恢复出可表示输入点云的单个特征向量,在点云分类和形状检索方面取得良好效果。

受空间上下文信息机制的启发, Ma 等^[63]提出了 3DMAX-Net,通过将单尺度下采样单元(DSU)与上采样单元(USU)相结合引入单尺度特征学习块(FLB),将不同尺度的 FLB 集成为多尺度特征学习块(MS-FLB),最后通过局部-全局特征聚合模块

进行特征聚合。Zhao 等^[64]通过连接交互局部邻域内的所有成对点来获取局部上下文信息,提出了 PointWeb 网络,通过插入自适应特征调整(AFA)模块构建局部全连接网络,从点对差异中学习点特征向量并实现自适应调整。Qiu 等^[65]引入自适应扩展点分组算法,提出一种密度分辨率网络,网络由

全分辨率(FR)分支和多分辨率(MR)分支组成,在 FR 分支中由一组误差最小化模块提取具有不同尺度特征空间的特征图,在 MR 分支中采用最远点采样(FPS)和特征传播(FP)分别进行上采样和下采样操作,提取多分辨率特征图,然后融合 FR 和 MR 的特征图,获取更全面的信息,其原理如图 7 所示。

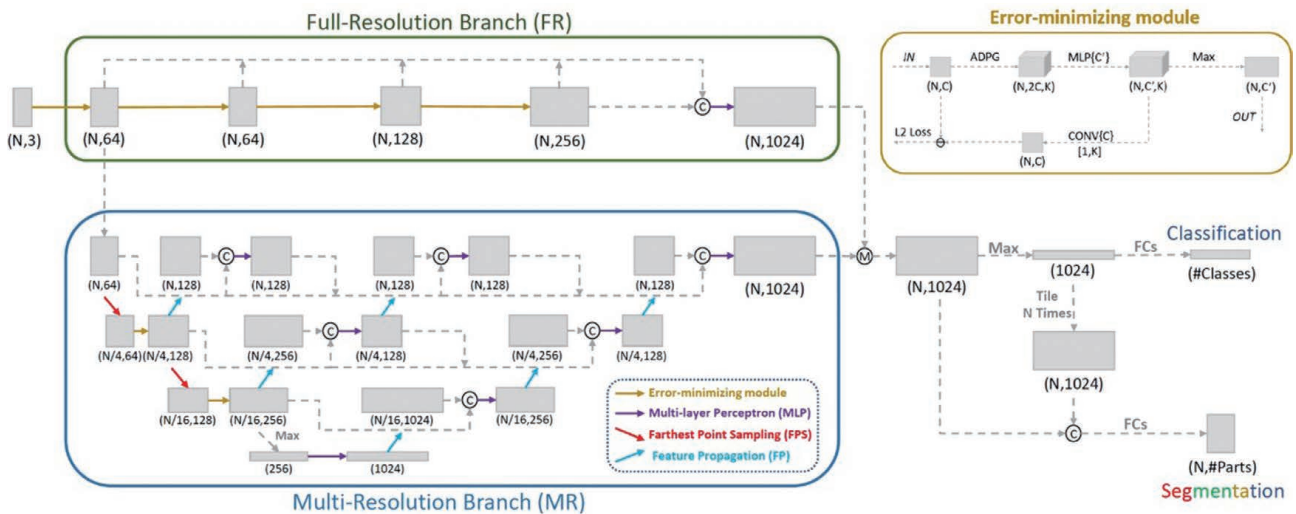


图 7 密度分辨率网络点云分类和分割框架图^[65]

Fig. 7 Architecture of dense-resolution network for point cloud classification and segmentation^[65]

为打破现有大多数方法无法直接适用于大规模点云的壁垒, Hu 等^[66]提出了一种轻量级网络 RandLA-Net,设计了一个局部特征聚合模块,通过逐渐增加每个神经层的感受野大小来有效学习复杂的局部结构。先对每个 3D 点引入局部空间编码(LocSE)单元以保留局部几何结构信息,再采用注

意力机制聚合有用的局部特征,并通过堆叠多个 LocSE 单元和池化层来增加每个点的有效感受野,其原理如图 8 所示。RandLA-Net 使用简单快速的随机采用方法降低点密度,采用局部特征聚合模块保留突出特征,在展现良好分类效果的同时提高了计算存储效率。

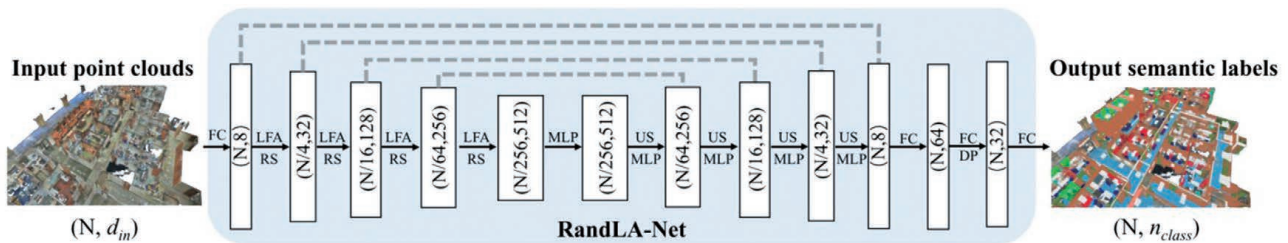


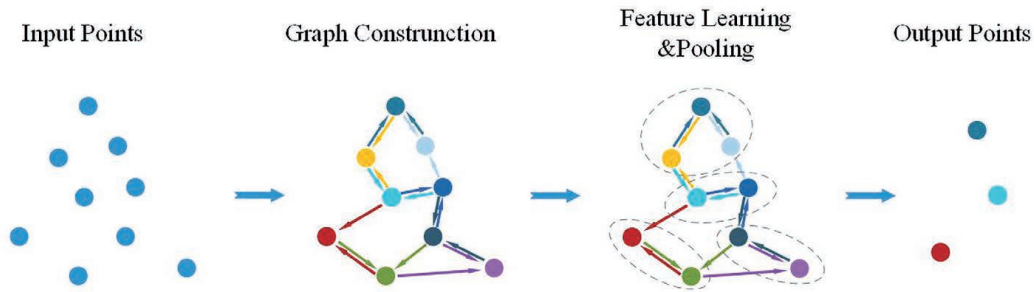
图 8 RandLA-Net 点云语义分割框架图^[66]

Fig. 8 Architecture of RandLA-Net for point cloud semantic segmentation^[66]

3.2.2 图形信息传递的方法

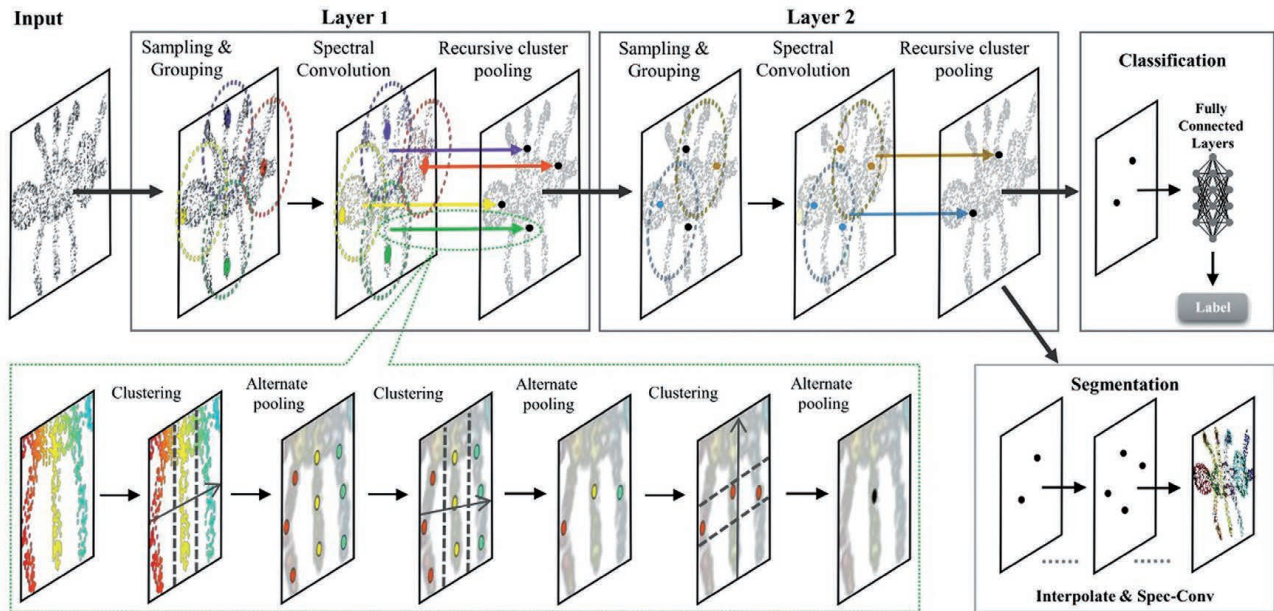
最早的图神经网络(GNN)由 Scarselli 等^[67]提出,此后 Bruna 等^[68]首次介绍了用图建模的非欧氏域的卷积神经网络,研究了用于执行局部滤波的空间域和频谱域两种网络架构, Kipf 等^[69]进一步提出图卷积神经网络(GCNN),在半监督分类任务中表现良好。受边缘标签应用的启发, Simonovsky 等^[70]提出一种利用边缘标签在空间域(即局部图邻域)上执行边缘条件卷积(ECC)的方法,并使用非对

称边缘函数来描述局部点之间的关系。Landrieu 等^[71]引入一种具有丰富边缘特征的属性有向图——超点图(SPG),基于门控图神经网络和边缘条件卷积(ECC)获取上下文信息,在大规模点云分割中表现良好。此后, Landrieu 等^[72]又提出 3D 点云过度分割策略,进一步提高了分割精度。 Jiang 等^[73]构建了一个边缘分支,将点特征以分层的方式聚合到边缘分支中生成边缘特征,以增强局部特征的描述能力。图 9 展示了基于图的网络原理。

图 9 基于图的神经网络原理示意图^[31]Fig. 9 Schematic diagram of a graph-based network^[31]

由于 PointNet++ 对每个点样本特征的学习都是在局部邻域中孤立进行的,且在特征聚合时采用贪婪的最大池化策略,容易导致有用信息丢失,具有一定局限性。为解决此类问题,Wang 等^[12]用递

归聚类策略代替最大池化策略提出了 SpecGCN,对每个点邻域图进行频谱图卷积,再进行递归聚合,经过几层采样、频谱图卷积、聚类池化后得到分类或分割结果,其原理如图 10 所示。

图 10 SpecGCN 点云分类和分割框架图^[12]Fig. 10 Architecture of SpecGCN for point cloud classification and segmentation^[12]

为进一步扩展 GCNN 的任务适用范围,Te 等^[74]提出一种正则化图卷积神经网络(RGCNN),该网络主要由三个正则化图卷积层构成,每层包含图构造、图卷积和特征滤波,通过在损失函数中添加图形信号平滑度以规范每个卷积层,更新图的拉普拉斯矩阵,以自适应地捕获动态图结构,采用切比雪夫近似以降低计算复杂度,在点云分类和分割任务中都表现出良好效果,且对点云密度和噪声具有较强的鲁棒性。Wang 等^[75]提出一种边缘卷积(EdgeConv)模块并设计了动态图卷积神经网络(DGCNN),该模块可以在保持置换不变性的同时捕获局部邻域的几何特征。

为进一步降低模型复杂度,提升网络性能,

Zhang 等^[76]在 DGCNN 基础上进行改进,提出了一种链接动态图卷积神经网络(LDGCNN),该网络来自不同动态图的层次特征链接在一起,通过使用当前索引从先前特征中获取有用的边缘矢量来学习新特征,有效避免了深度神经网络梯度消失的问题。Liu 等^[77]则提出一种动态点聚集模块(DPAM)来执行灵活点采样、分组、合并操作。

受深度卷积^[78]应用的启发,Liang 等^[79]提出一种层次深度图卷积神经网络(HDGCN),设计了一个基于深度图卷积和逐点卷积的 DGConv 单元,先采用逐点卷积聚合每个点的输入特征,再使用深度图卷积提取局部特征,然后再用逐点卷积学习不同层次特征,最后聚合全局特征。受一些谱卷积方法的

启发, Lu 等^[80]将 Chebyshev 多项式作为邻域图滤波器扩展到 3D 点云中, 基于 PointNet 提出了 PointNGCNN, 通过构建邻域图来描述邻域内点的空间关系, 应用邻域图滤波器分别提取特征空间和笛卡儿空间中的邻域特征信息和空间分布信息。

Zhang 等^[81]提出一种全方向图神经网络 LKPO-GNN, 将 3D 无序点转换为 1D 有序序列, 在 LKPO-GNN 模块中采用全方向局部 k-NNs 模式的图表示, 学习点云丰富的局部拓扑结构, 然后使用球查询模块 (Ball Query) 提取点的丰富特征信息, 交替应用 LKPO-GNN 模块和 Ball Query 模块精简

3D 点云场景中的中心点并丰富中心点聚合的特征信息, 其原理如图 11 所示。LKPO-GNN 从特征向量中学习获得全方向 k-NNs 模式图的顶点, 具有良好的空间几何信息表现能力, 这与大多数点云上的图卷积方法都使用 k-NN 搜索来查找中心点邻域不同。为进一步降低后续卷积层的计算成本, Nezhadarya 等^[82]提出一种自适应的下采样层——临界点层 (CPL), 根据点对全局最大池化的贡献水平自适应地保留信息量最大的点, 滤除非关键点, 将 CPL 嵌入任意图卷积神经网络可高效完成分类或分割任务。

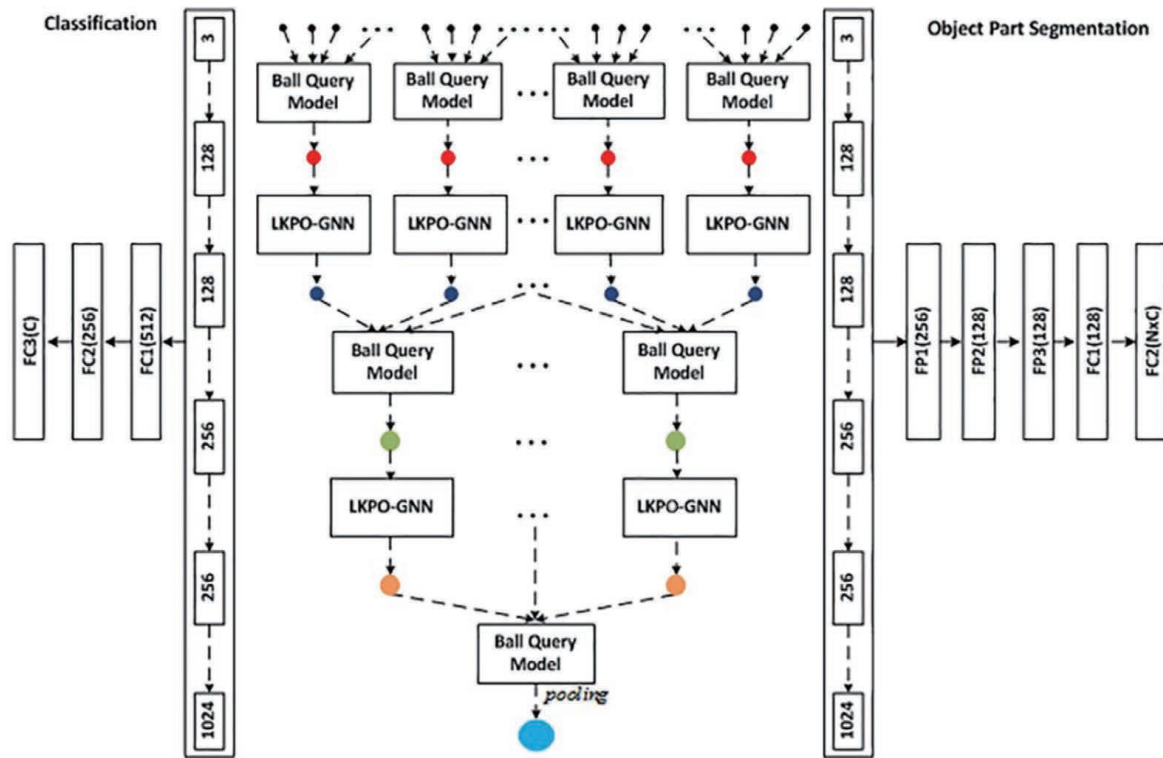


图 11 LKPO-GNN 点云分类和分割框架图^[81]

Fig. 11 Architecture of LKPO-GNN for point cloud classification and segmentation^[81]

3.2.3 基于内核卷积的方法

根据卷积内核的类型, 可将 3D 点云的卷积分为离散卷积和连续卷积^[31], 如图 12 所示。

3.2.3.1 离散卷积

对数据点来说, 在密集的网络结构中点的自然分散特性使得采用离散卷积对其进行处理更有效

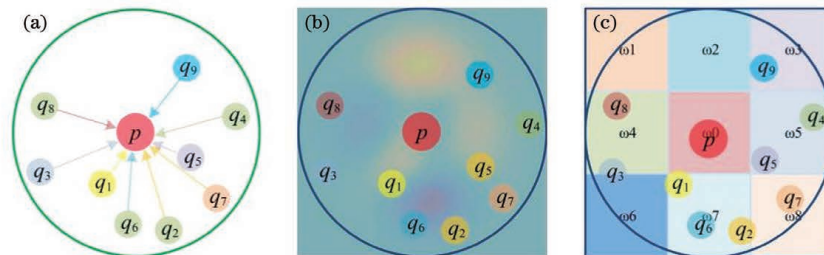


图 12 点的局部邻域连续和离散卷积^[31]。(a) 点的局部邻域; (b) 3D 连续卷积; (c) 3D 离散卷积

Fig. 12 An illustration of the continuous and discrete convolutions for local neighborhoods of a point^[31]. (a) Local neighborhoods of a point; (b) 3D continuous convolution; (c) 3D discrete convolution

性,离散卷积也成为现代深度学习架构的最基本组成部分,这类方法在规则网格上定义卷积核,其中相邻点的权重与相对于中心点的偏移量有关。

由于点云数据的无序性,常规卷积操作很难直接应用到点云数据上。为解决这类问题,Hua 等^[13]将非均匀的 3D 点云以特定顺序进行排序并输入网格,定义每个点具有相同权重、固定大小的卷积核,将即时查询的最近邻点合并到核单元中,再对每个核单元内的所有点进行逐点卷积运算,最后对所有单元的平均特征进行加权求和。受空间变换网络的启发,Li 等^[83]考虑将 X 变换应用于输入点坐标并提出 X-conv 卷积算子,用 MLP 学习输入点特征的加权和置换,再使用常规卷积来提取局部特征,并采用“子体积监督”思想进一步解决了过拟合问题。

为了获取更丰富的几何结构信息,Lan 等^[84]通过向量分解对局部邻域的几何结构进行建模,提出了 GeoConv,先将两点之间的边缘特征分解到三个固定的正交基中,并用与方向有关的权重矩阵提取沿每个方向的特征,然后基于边缘向量和基之间的角度聚合特征,提取局部邻域几何特征。通过正交分解边缘特征,减小了点云的绝对坐标的方差,且使网络独立学习各个方向的边缘特征,在保留丰富几何特征的同时减小了计算复杂度。

为了解决基于图结构的卷积运算密度敏感^[75]的问题,Mao 等^[85]提出一种内插卷积算子 InterpConv,为每个卷积核定义一组可以放置在任意位置的离散内核权重向量,采用三线性插值逆过程将输入点特征反向分配给相邻内核权重坐标,然后将输入点和内核权重归一化为单位长度的立方体,该方法一定程度上消除了点云密度带来的影响。为了减少多尺度体系结构^[60]中进行多尺度邻域搜索时重叠点导致的冗余信息,并避免方向选择带来的负面影响,Komarichev 等^[86]创新性地提出一种环形卷积并设计了环形卷积神经网络(A-CNN),与多尺度策略相比,基于环的结构在查询点的邻域中没有重复的相邻点,在大型场景数据集的应用中体现出良好的鲁棒性。

在有监督模糊聚类神经网络(SFCNN)中,Rao 等^[87]提出一种自适应球体投影模块,通过引入规则的二十面体晶格来离散球形邻域,并对球形晶格顶点到其相邻像素点间的边缘特征进行卷积,该方法具有良好的鲁棒性和稳健性。与 SFCNN 中基于球面域实现旋转不变性不同,Zhang 等^[88]提出一个 RICONV 算子,结合简单的几何关系构建分箱策略,

在单个卷积中处理旋转不变性和点排序问题。Lei 等^[89]则通过将 3D 球形邻域进行基于加权矩阵的划分,定义了球形卷积核,系统地对点邻域进行了量化处理,由于球形卷积核与空间位置相互关联,有效避免了网络训练过程中动态核的生成,故在高分辨率 3D 点云的对象分类和分割中表现出良好效果。此外,Engelmann 等^[90]通过全面的消融性研究证明感受野的大小与 3D 点云处理任务性能直接相关,提出膨胀点卷积(DPC),通过聚合膨胀的邻域要素来扩大感受野大小,提升了网络性能。

3.2.3.2 连续卷积

由于标准离散卷积在许多实际应用中的局限性,研究者们逐渐在连续空间上定义卷积核,其中相邻点的权重与相对于中心点的空间分布有关。Wang 等^[91]提出一种基于参数连续卷积层网络(PCCN),其卷积核函数由 MLP 参数化,并跨越整个连续向量空间。该网络只要支持关系可计算,则参数连续卷积可以处理任意数据结构。Xu 等^[92]引入一个 SpiderConv 单元,提出 SpiderCNN,将卷积运算从常规网格扩展到不规则点集,同时通过构造一个点局部邻域极坐标半径变量的阶跃函数来对局部邻域内的空间几何信息进行编码,提取深层次的语义信息。Wu 等^[93]将 3D 连续卷积算子进行蒙特卡罗近似的扩展,提出了一种使用非均匀采样对 3D 点云进行卷积的算子 PointConv,对每个卷积滤波器使用 MLP 近似加权函数,然后用密度函数对学习的加权函数进行重新加权,实现置换不变性和平移不变性。

与 Shen 等^[94]应用点云注册技术、使用没有任何权重的内核点来学习局部几何特征的不同,Thomas 等^[95]使用一组内核点来定义每个内核权重的应用邻域,提出一种可扩展到可形变卷积的核点卷积算子 KPConv,提升了网络适应场景对象的能力。为了捕获更丰富的形状上下文信息,Liu 等^[96]提出一种可以学习点之间的几何拓扑关系的卷积算子 RS-Conv,并基于该算子设计了关系形状卷积神经网络(RS-CNN)。在 RS-Conv 中,局部子集的卷积权重将学习从预定义的低级几何关系向量到点之间高级抽象关系的映射,依此可对点的空间布局进行推理,实现强大的形状上下文感知。

此外,为了解决 3D 卷积网络面临的旋转不变性问题,Thomas 等^[97]提出了张量场网络,认为可将标量、向量和高阶张量作为每一层滤波器的输入和输出,将滤波器限制为可学习的径向函数和球谐函数的乘积。受非交换谐波技术的启发,Cohen 等^[98]

提出一种具有较强泛化能力的球形 CNN, 定义了一种兼具表达性和旋转不变性的球面互相关, 并采用广义快速傅里叶变换 (FFT) 对球体和旋转组执行互相关操作。受 PCNN 网络扩展约束机制的启发, Poulencard 等^[99]通过引入球形谐波核定义了一个旋转不变卷积算子 SPHConv 并提出 SPHNet 网络, 由扩展运算符将点云上的函数扩展为体积函数, 再

在体积函数上应用 SPHConv 以实现旋转不变性。与 KPConv 使用显式的径向基函数 (RBF) 来关联输入和内核不同, Boulch 等^[100]则考虑用 MLP 来学习关联函数, 提出了 ConvPoint, 先对空间局部邻域进行搜索, 再对每个邻域执行基于 MLP 的卷积运算。图 13 为几种典型离散卷积和连续卷积算子原理示意图。

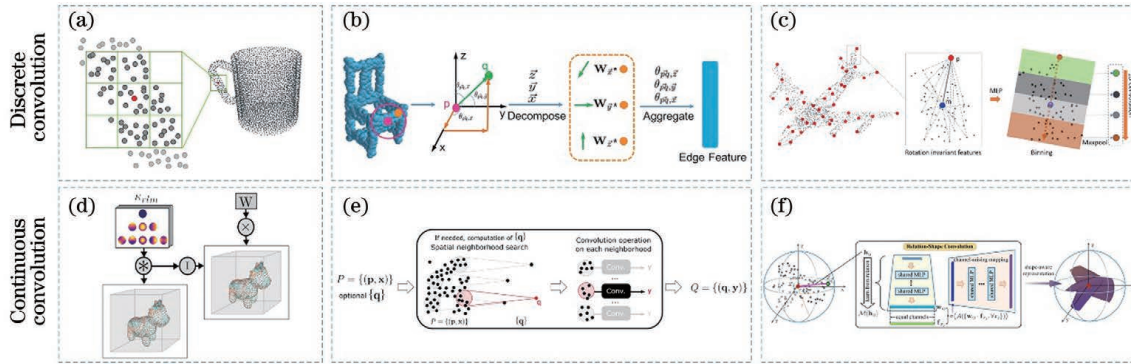


图 13 几种典型卷积原理示意图。(a) Pointwise Conv^[13]; (b) GeoConv^[84]; (c) RICov^[88]; (d) SPHConv^[99]; (e) Convpoint 卷积层^[100]; (f) RS-Conv^[96]

Fig. 13 Schematics of several typical convolutions. (a) Pointwise Conv^[13]; (b) GeoConv^[84]; (c) RICov^[88]; (d) SPHConv^[99]; (e) convolutional layer of Convpoint^[100]; (f) RS-Conv^[96]

3.2.4 基于注意力机制的方法

人类感知的一个重要特性是不会一次全盘处理整个场景, 而是有选择地将注意力集中在视觉空间的各个部分上^[101], 由此, 注意力机制成为研究者们重点研究的内容, 其在自然语言处理^[102]、对象检

测^[103]、语义分割^[104]、三维重构^[105]等领域发挥了重要作用。其中, 自我注意机制是通过考虑每个点的自身几何信息学习自我系数, 邻域注意机制则通过考虑邻域关注局部系数。图 14 为自我注意机制和邻域注意机制的原理示意图。

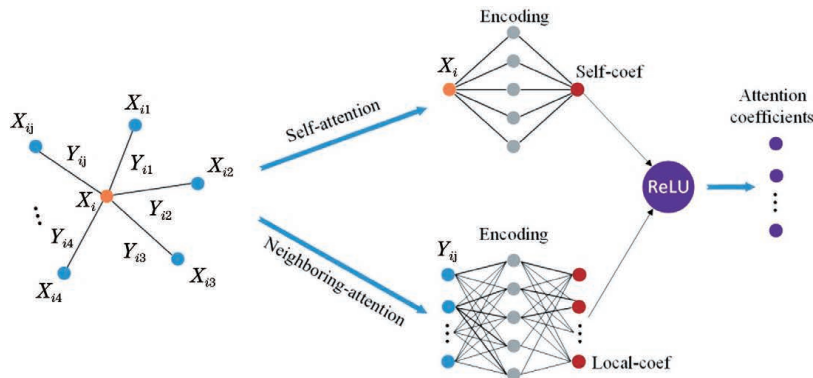


图 14 注意力系数生成原理

Fig. 14 Principle of attention coefficients generation

注意力机制的好处之一是允许处理可变大小的输入, 其着眼于输入中最相关的部分并对其做出决策。Veličković 等^[14]提出图注意层并构建了图注意力网络 (GAT), 根据节点特征执行自我关注得到注意力系数, 将标准化的注意力系数与其对应的特征向量进行线性组合获取每个节点的最终输出特征, 完成对邻域内不同节点的注意力分配。与 GAT 不同, 在

Wang 等^[106]提出的 GAC 中, 注意力权重的产生不仅取决于邻域的空间位置, 还取决于其动态学习的特征差异。为了保证不同通道特征的相互独立性, 不仅为不同相邻点分配注意力权重, 还为不同特征通道分配适当的注意力权重, 再通过加权计算出每个点特征, 然后进行图池化以降低每个特征通道中的分辨率, 最后将学习到的特征逐层插回, 完成逐点标签分配。

Chen 等^[107]引入 GAPLayer, 通过将图注意力机制嵌入堆叠的 MLP 来学习局部几何特征, 提出了 GAPNet, 对邻域内的点分配不同注意力权重来学习点特征, 然后引入多头注意力(MHA)机制来聚合从不同 GAPLayer 输出的特征, 获取多注意特征和多图特征。Yang 等^[108]则用一种组混洗注意力(GSA)机制来代替昂贵的多头注意力(MHA), 开发了基于点云推理的点注意力变压器(PAT)。

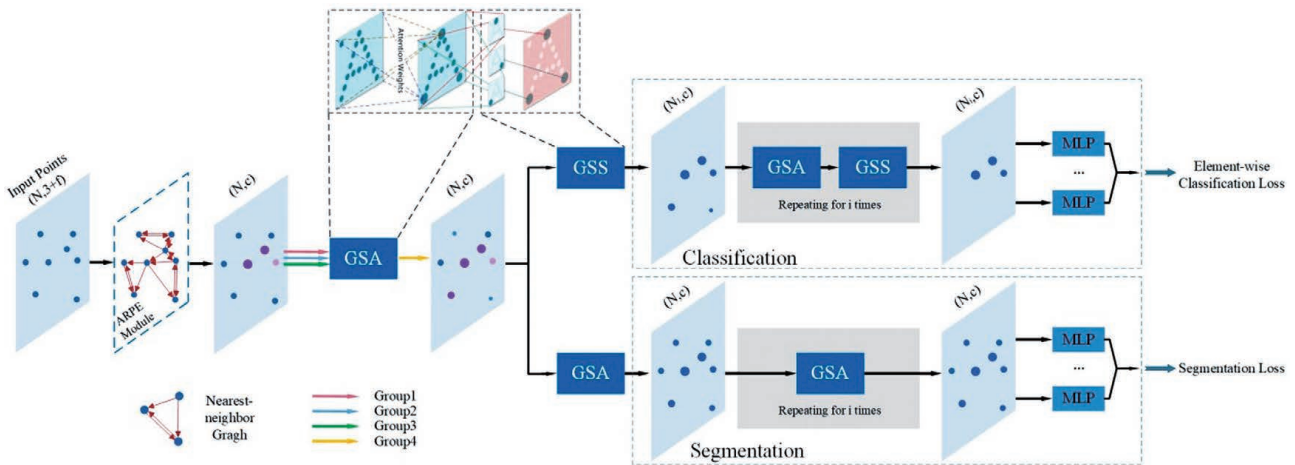


图 15 PATNet 点云分类和分割框架图

Fig. 15 Architecture of PATNet for point cloud classification and segmentation

受空间金字塔池在聚合不同邻域上下文信息中的出色表现的启发, Li 等^[109]考虑将注意力机制与空间金字塔结合来获取丰富的语义信息, 提出了特征金字塔注意(FPA)模块, 通过实现 U 形结构融合三种不同金字塔尺度下的上下文特征, 并为高层次特征图提供像素级别注意力, 同时还提出全局注意力上采样(GAU)模块, 采用全局平均池化提取高层次特征, 为低层次特征图提供指导信息, 提高了分割精度。Kang 等^[110]则基于金字塔注意网络(PAN), 引入图形嵌入模块(GEM), 提出了 PryamNet, GEM 将点云投影到图上并使用协方差矩阵描述局部邻域特征, PAN 采用不同大小的卷积核对特征图进行下采样, 融合具有不同语义强度的相邻点特征, 进一步提高局部特征提取的准确性。

为充分利用点云的空间分布信息, Chen 等^[111]设计了一个局部空间感知(LSA)层以对局部邻域的几何结构进行准确而稳健的建模, 并提出了 LSA-Net 网络, 先采用分层的局部空间特征提取器(SFE)对输入点云进行抽象表示, 获取高维空间信息, 然后根据局部邻域的空间关系分层地生成空间分布权重(SDW), 该方法具有更强大的空间信息提取功能。与一些通过多尺度特征融合来捕获上下文

首先结合输入点的绝对位置和所有剩余点的相对位置构建绝对-相对位置嵌入(ARPE)模块, 通过该模块将输入点嵌入到高级表示中, 使某些点具有代表性, 然后将点特征分组并对每点特征应用 GSA 来挖掘特征集中元素间的关系, 根据任务用一个排列不变的 Gumbel 子集采样(GSS)选择有代表性的点子集, 最后通过共享 MLP 完成分类或分割任务, 有效降低了网络的计算成本, 其原理如图 15 所示。

信息的方法不同, Fu 等^[112]提出双重注意力网络 DANet, 通过位置注意力模块和通道注意力模块分别对空间和通道维度的语义依赖性进行建模, 并有选择地聚合局部邻域特征。此外, Xie 等^[113]将自我注意机制引入形状上下文, 提出了 A-SCN 模型, 该模型能自动执行上下文区域选择、特征聚合和特征转换的过程。Zhao 等^[114]提出一种基于注意力的分数优化(ASR)模块, 根据每个点的初始分割得分计算各自权重, 然后通过使用自适应确定的权重合并相邻点的分数来优化初始分数, 再基于改进的分数来选择语义标签, ASR 模块能嵌入任意深度神经网络中, 极大地减少了基础网络初始分割中的歧义点和噪声。Feng 等^[115]提出了一种局部注意边缘卷积(LEA-Conv)层, 基于多向搜索策略构造局部图, 将注意力系数分配给图的每条边, 然后将中心点特征聚合为其相邻节点的加权总和, 获得细粒度的局部几何特征, 并设计了一个逐点空间注意模块, 生成所有点的相互依赖性矩阵, 通过在点集要素之间建立关联来捕获全局的依存关系, 打破了局部细节与远程上下文信息难以兼顾的限制。

表 2 汇总了基于原始点云分类方法的提出年份、核心思想、应用场景, 以及在主要数据集上的表

表 2 基于原始点云的方法比较
Table 2 Comparison for methods based on original point clouds

Method	Year	Key idea	Application scenario	Dataset	Accuracy / %		
					OA	MA	mIoU
PointNet ^[1]	2017	Using a single symmetric function, max pooling	3D object classification, part segmentation, scene semantic parsing	ModelNet40	89.2	86.2	—
PointNet ^[1]	2017	Processing a set of points sampled in a metric space using a hierarchical fashion	Processing point sets sampled in a metric space	ModelNet40	91.9	—	—
PointSIFT ^[61]	2018	Stacking several orientation-encoding units to achieve multi-scale representation	Improving 3D shape representation	S3DIS	88.72	—	70.23
SO-Net ^[62]	2018	Building a self-organizing map (SOM) to model the spatial distribution of point cloud	Point cloud reconstruction, classification, object part segmentation and shape retrieval	ModelNet40	93.4	—	—
3DMAX-Net ^[63]	2018	Multi-scale contextual feature learning, local and global feature aggregation	3D semantic segmentation on large-scale point clouds	S3DIS	79.5	—	47.5
PointWeb ^[64]	2019	Using adaptive feature adjustment (AFA) module to find the interaction between points	3D point cloud segmentation and classification	ModelNet40	92.3	89.4	—
In Ref. [65]	2020	Learning local point features from point cloud in different resolutions	Point cloud analysis	ModelNet40	93.1	—	—
RandLA-Net ^[66]	2020	Using random point sampling instead of complex point selection approaches	more 3D semantic segmentation on large-scale point clouds	Semantic 3D	94.8	—	77.4
In Ref. [70]	2017	Performing convolutions over local neighborhoods exploiting edge labels	Graph classification	ModelNet40	—	87.4	—
SPG ^[71]	2018	Capturing the organization of 3D point clouds by superpoint graph (SPG)	3D semantic segmentation on large-scale point clouds	Semantic3D	92.9	—	76.2
				S3DIS	85.5	73.0	62.1

表 2(续)

Method	Year	Key idea	Application scenario	Dataset	Accuracy / %		
					OA	MA	mIoU
SpecGCN ^[12]	2018	Leveraging the power of spectral graph CNNs in the PointNet++ framework while adopting a different pooling strategy	3D point cloud segmentation and classification	ModelNet40	91.5	—	—
			—	ShapeNet	—	—	84.6
RGCNN ^[74]	2018	Adding graph-signal smoothness a prior in the loss function	3D point cloud segmentation and classification	ModelNet40	90.5	87.3	—
DGCNN ^[75]	2018	Using EdgeConv to capture and exploit fine-grained geometric properties of point clouds	3D point cloud segmentation and classification	ModelNet40	92.2	90.2	—
LDGCNN ^[76]	2019	Removing the transformation network; linking hierarchical features from different dynamic graphs	3D point cloud segmentation and classification	ShapeNet	—	—	85.1
			—	ModelNet40	92.9	90.3	—
In Ref. [72]	2019	Using a simple point embedding network and a new graph-structured loss function	3D semantic segmentation on large-scale point clouds	S3DIS	87.9	78.3	68.4
			—	vKITTI	84.3	67.3	52.0
In Ref. [77]	2019	Stacking DPAM module to gradually agglomerate points	3D point cloud segmentation and classification	ModelNet40	91.9	—	—
			—	ModelNet10	94.6	—	—
HDGCN ^[79]	2019	Combining the hierarchical structure and the DGConv block to extract both local and global features of point clouds hierarchically	3D semantic segmentation (indoor and outdoor scenes)	Paris-Lille-3D	—	—	68.30
			—	ModelNet40	92.8	—	—
PointNGCNN ^[80]	2020	Using the Chebyshev polynomials as the graph filters to extract features in the neighborhood of each point	potential geometric information of 3D objects	ShapeNet	—	—	85.6
			—	S3DIS	87.3	—	—
LKPO-GNN ^[81]	2020	Using LKPO-GNN to select multi-directional k-NNs to form the local topological structure of a centroid	deeper representation	ScanNet	84.9	—	—
			—	ModelNet40	91.4	88.9	—
CPL-Net ^[82]	2020	Using critical points layer (CPL) to reduce the number of points in an unordered point cloud and retain the important (critical) ones	feature	ShapeNet	—	—	85.6
			—	S3DIS	85.8	—	64.6
—	—	—	—	ScanNet	85.3	—	58.4
—	—	—	—	ModelNet40	92.41	90.53	—

表 2(续)

Method	Year	Key idea	Application scenario	Dataset	Accuracy / %		
					OA	MA	mIoU
Pointwise CNN ^[13]	2017	Pointwise convolution which can be applied at each point in a point cloud to learn point-wise features	3D semantic segmentation and object recognition	S3DIS	—	74.1	—
PointCNN ^[83]	2018	X-Conv; weighting and permuting input points and features before processed by a typical convolution	Leveraging spatially-local correlation from data represented in point cloud	ModelNet40	92.5	88.8	—
PCCN ^[91]	2018	Exploiting parameterized kernel functions which span the full continuous vector space	Point cloud segmentation (indoor and outdoor scenes), lidar motion estimation of driving scenes	Stanford Large-Scale 3D Indoor Scene Dataset Driving Scenes Dataset	—	67.01	58.27
SpiderCNN ^[92]	2018	SpiderConv; extending convolutional operations from regular grids to irregular point sets	3D point cloud segmentation and classification	ModelNet40	92.4	—	—
GeoCNN ^[84]	2019	GeoConv; modeling the geometric structure of points by a decomposition and aggregation method based on vector decomposition	3D shape classification, segmentation and object detection	ShapeNet	—	—	85.3
InterpCNN ^[85]	2019	Interp Conv; using discrete convolutional kernels and an interpolation function to explicitly measure geometric relations between input point clouds and kernel-weight coordinates	3D shape classification, object part segmentation and indoor scene semantic parsing	ModelNet40 S3DIS ShapeNet	93.0 88.7	—	— 66.7
A-CNN ^[86]	2019	Capturing the local neighborhood geometry of each point by specifying the (regular and dilated) ring-shaped structures and directions in the computation	Object classification, part segmentation, and semantic segmentation in large-scale scenes	ModelNet40 ModelNet10 S3DIS ShapeNet ScanNet	92.6 95.5 87.3	90.3 95.3	— — — 86.1
In Ref. [88]	2019	RICConv; using low-level rotation invariant geometric features such as distances and angles to design a convolution operator for point cloud learning	3D object classification and segmentation	ModelNet40 ShapeNet Driving Scenes Dataset	86.5 — 95.45	— —	— 75.5 58.06

表 2(续)

Method	Year	Key idea	Application scenario	Dataset	Accuracy / %		
					OA	MA	mIoU
Kernel-based convolution	2019	PointConv; taking the positions of point clouds as input and learning an MLP to approximate a weight function, then applying an inverse density scale on the learned weights to compensate the non-uniform sampling	3D semantic segmentation;	ModelNet40	92.5	—	—
			convolutional networks in 2D images	ShapeNet	—	—	85.7
			of a similar structure	ScanNet	—	—	55.6
Kernel-based convolution	2019	KPConv; using a set of kernel points to define the area where each kernel weight is applied	Adapting to the geometry of the scene objects	ModelNet40	92.9	—	—
			3D shape deep learning tasks	ShapeNet	—	—	86.4
			kernel at different layers of the network	ModelNet40	87.7	—	—
Kernel-based convolution	2019	Using continuous convolution and a hierarchical data representation structure based on a search tree	Large scale indoor and outdoor semantic segmentation	ModelNet40	92.5	89.6	—
			semantic segmentation	ShapeNet	—	—	85.8
			constraint among points	ModelNet40	93.6	—	—
Attention-based methods	2020	RS-Conv; learning from the geometric topology information in 3D point cloud	Encoding meaningful shape information in 3D point cloud	ShapeNet	—	—	86.2
			3D point cloud classification and segmentation	ShapeNet	—	—	84.6
			segmentation	ShapeNet	—	—	84.6
Attention-based methods	2018	Combining Feature Pyramid Attention (FPA) module and Global Attention Upsample (GAU)	3D point cloud semantic segmentation (urban scenes)	PASCAL VOC 2012	95.7	—	84.0
			3D object classification and semantic segmentation	Cityscapes	—	—	78.6
			3D shape classification and part segmentation	ModelNet40	91.5	88.3	—
Attention-based methods	2019	Combining Graph Embedding Module (GEM) and Pyramid Attention Network (PAN)	3D object classification and semantic segmentation	S3DIS	85.6	—	55.6
			3D shape classification and part segmentation	ShapeNet	—	—	83.9
			3D shape classification and part segmentation	ModelNet40	92.4	89.7	—
Attention-based methods	2019	GAPLayer; embedding graph attention mechanism within stacked Multi-Layer-Perceptron (MLP) layers to learn local geometric representations	3D shape classification and part segmentation	ShapeNet	—	—	84.7
			3D shape classification and part segmentation	ShapeNet	—	—	84.7
			3D shape classification and part segmentation	ShapeNet	—	—	84.7

表 2(续)

Method	Year	Key idea	Application scenario	Dataset	Accuracy / %		
					OA	MA	mIoU
PAT ^[108]	2019	Using a parameter-efficient Group Shuffle Attention (GSA) to replace the costly Multi-Head Attention; Gumbel Subset Sampling (GSS)	Hierarchical multiple instance learning	ModelNet40	91.7	—	—
				S3DIS	—	—	64.28
LSANet ^[111]	2019	Generating Spatial Distribution Weights (SDWs) 3D object classification, part and semantic segmentation, hierarchically based on the spatial relationship in segmentation, and local region for spatial independent operations		ModelNet40	92.3	89.2	—
				S3DIS	86.8	—	62.2
				ShapeNet	—	—	83.2
				ScanNet	85.1	—	—
Attention-based methods	2019	Attention-based score refinement (ASR) module	Improving the segmentation accuracy	ShapeNet	—	—	85.6
					Semantic3D	91.9	—
GACNet ^[106]	2020	Assigning proper attentional weights to different neighboring points	3D point cloud segmentation	S3DIS	87.79	—	62.85
					S3DIS	88.95	66.3
In Ref. [115]	2020	Local Attention-Edge Convolution (LAE-Conv); constructing a local graph based on the neighborhood points searched in multi-directions	Predicting dense labels for 3D point cloud segmentation	ShapeNet	—	—	85.9
					ScanNet	88.6	—

现性能。

由于直接输入原始点云数据具有信息完整的特性,基于原始点云的方法相较于基于规则表示的方法优势明显,算法性能良好,应用场景多样,许多算法在多个数据集上均有良好表现。在选用 ShapeNet 数据集的算法中,mIoU 绝大多数达到 84% 以上; PointSIFT^[61]、3DMAX-Net^[63]、RandLA-Net^[66]、SPG^[71]、文献^[72]、Pointwise CNN^[13]、GACNet^[106] 等算法针对大规模场景都具有良好的性能,其中 SPG 在 Semantic 3D 数据集上 OA 达到 92.9%, RandLA-Net 则达到 94.8%。尽管基于原始点云的方法有许多优势,也是未来研究趋势,但其网络模型复杂,执行时间长,模型的优化是此类方法的难点。

4 结束语

本文系统回顾了 3D 点云分类方法概况,对点云分割与分类的区别和联系进行了阐述,并介绍了点云分类算法性能的评价指标,重点对近几年提出的基于深度学习的点云分类方法进行了细致的总结、归纳和分析。尽管当前基于深度学习的有关研究取得了丰硕成果,但不论是基于规则表示的点云还是基于原始点云,现有方法仍不同程度地存在一定的局限性,如何突破这些局限将是未来研究的热点。本章基于前面章节的内容,对点云分类的未来研究方向进行展望:

1) 当前大多数方法都旨在改善某一方法在某一方面缺陷,且大多针对特定场景或在特定数据集上取得良好效果,很难为实际应用选择合适方法,缺乏一定的灵活性和普适性。2020 年,Wang 等^[116] 提出一种基于神经结构搜索(NAS)的 RSNNet 搜索框架,采用两级优化策略,在搜索阶段设计一个分层的基本搜索空间,并基于梯度下降法交替优化体系结构和模型参数,在训练阶段对搜索到的网络进行二次训练,进一步优化模型参数,实现了面向任务需求的最优网络的自动搜索。自适应地调整网络结构可能成为未来点云分类算法的研究方向之一。

2) 现有点云分类方法很难同时兼顾精度和计算效率,一般都是通过实验在二者中寻求最好的平衡。2020 年,Li 等^[117] 提出一种自适应的数据增强框架 PointAugment,能在训练过程中根据样本本属性和网络能力动态调整样本的扩充;Ehsan 等^[82] 提出一种自适应的下采样层来保留关键点,滤除非关键点。由此,设计具有强大的样本感知能力,并能根据任务需求和任务不同阶段情况来自适应调整数据体

量的网络将具有重要意义。

3) 尽管点云分类和语义分割是遥感领域和计算机视觉领域的两种称呼,但由于遥感领域和计算机视觉领域所关注的点云数据不同(遥感更关注机载或稀疏点云等,而计算机视觉更关注密集完整的点云),需要的数据集不同(遥感应用一般需要具有更复杂和特定地面对象类别的大规模数据,而计算机视觉通常在对象类别有限的小区域数据集上进行测试),对传感器噪声的敏感程度不同(遥感中传感器噪声不可避免,而计算机视觉算法对噪声关注度并不高),很多计算机视觉研究提出的算法并不能直接用于遥感,一定程度上限制了遥感领域点云分类技术的进步,大规模复杂场景的点云分类技术在未来有待进一步研究发展。

4) 随着传感器技术的快速发展,多源数据融合已成为遥感领域的研究热点^[118,119],新一代 TomoSAR 点云^[120] 的出现提高了对相关算法的需求,针对多源点云数据的处理将是未来的发展趋势。

参 考 文 献

- [1] Kim K, Kim C, Jang C, et al. Deep learning-based dynamic object classification using LiDAR point cloud augmented by layer-based accumulation for intelligent vehicles [J]. *Expert Systems with Applications*, 2021, 167: 113861.
- [2] Zermas D, Izzat I, Papanikolopoulos N. Fast segmentation of 3D point clouds: a paradigm on LiDAR data for autonomous vehicle applications[C]// 2017 IEEE International Conference on Robotics and Automation (ICRA), May 29-June 3, 2017, Singapore. New York: IEEE Press, 2017: 5067-5073.
- [3] Ene L T, Næsset E, Gobakken T, et al. Large-scale estimation of change in above ground biomass in miombo woodlands using airborne laser scanning and national forest inventory data[J]. *Remote Sensing of Environment*, 2017, 188: 106-117.
- [4] Polewski P, Yao W, Heurich M, et al. Detection of fallen trees in ALS point clouds using anormalized cut approach trained by simulation[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2015, 105: 252-271.
- [5] Yang B S, Huang R G, Li J P, et al. Automated reconstruction of building LoDs from airborne LiDAR point clouds using an improved morphological scale space[J]. *Remote Sensing*, 2016, 9(1): 14.
- [6] Pan Y, Dong Y Q, Wang D L, et al. Three-dimensional reconstruction of structural surface

- model of heritage bridges using UAV-based photogrammetric point clouds[J]. *Remote Sensing*, 2019, 11(10): 1204.
- [7] Weinmann M, Schmidt A, Mallet C, et al. Contextual classification of point cloud data by exploiting individual 3D neighbourhoods [J]. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2015, II-3/W4: 271-278.
- [8] Wang Z, Zhang L Q, Fang T, et al. A multiscale and hierarchical feature extraction method for terrestrial laser scanning point cloud classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2015, 53(5): 2409-2425.
- [9] Koppula H S, Anand A, Joachims T, et al. Semantic labeling of 3D point clouds for indoor scenes [J]. In *Advances in Neural Information Processing Systems*, 2011, 1: 244-252.
- [10] Lu Y, Rasmussen C. Simplified Markov random fields for efficient semantic labeling of 3D point clouds[C]//2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, October 7-12, 2012, Vilamoura-Algarve, Portugal. New York: IEEE Press, 2012: 2690-2697.
- [11] Charles R Q, Hao S, Mo K C, et al. PointNet: deep learning on point sets for 3D classification and segmentation [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 77-85.
- [12] Wang C, Samari B, Siddiqi K. Local spectral graph convolution for point set feature learning[M]//Ferrari V, Hebert M, Sminchisescu C, et al. *Computer vision-ECCV 2018. Lecture notes in computer science*. Cham: Springer, 2018, 11208: 56-71.
- [13] Hua B S, Tran M K, Yeung S K. Pointwise convolutional neural networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 984-993.
- [14] Veličković P, Cucurull G, Casanova A, et al. Graph attention networks[EB/OL]. (2017-10-30)[2020-10-15]. <https://arxiv.org/abs/1710.10903>.
- [15] Lalonde J F, Unnikrishnan R, Vandapel N, et al. Scale selection for classification of point-sampled 3D surfaces[C]//Fifth International Conference on 3-D Digital Imaging and Modeling (3DIM'05), June 13-16, 2005, Ottawa, ON, Canada. New York: IEEE Press, 2005: 285-292.
- [16] Lalonde J F, Vandapel N, Huber D F, et al. Natural terrain classification using three-dimensional lidar data for ground robot mobility[J]. *Journal of Field Robotics*, 2006, 23(10): 839-861.
- [17] Gao Z H, Liu X W. Support vector machine and object-oriented classification for urban impervious surface extraction from satellite imagery [C] // 2014 The Third International Conference on Agro-Geoinformatics, August 11-14, 2014, Beijing, China. New York: IEEE Press, 2014: 1-5.
- [18] Colgan M, Baldeck C, Féret J B, et al. Mapping savanna tree species at ecosystem scales using support vector machine classification and BRDF correction on airborne hyperspectral and LiDAR data [J]. *Remote Sensing*, 2012, 4(11): 3462-3480.
- [19] Miao X, Heaton J S. A comparison of random forest and Adaboost tree in ecosystem classification in east Mojave Desert[C]//2010 18th International Conference on Geoinformatics, June 18-20, 2010, Beijing, China. New York: IEEE Press, 2010: 1-6.
- [20] Chan J C W, Paelinckx D. Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery [J]. *Remote Sensing of Environment*, 2008, 112(6): 2999-3011.
- [21] Gan Z, Zhong L, Li Y F, et al. A random forest based method for urban object classification using lidar data and aerial imagery [C] // 2015 23rd International Conference on Geoinformatics, June 19-21, 2015, Wuhan, China. New York: IEEE Press, 2016, 6: 19-21.
- [22] Wang C S, Shu Q Q, Wang X Y, et al. A random forest classifier based on pixel comparison features for urban LiDAR data[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2019, 148: 75-86.
- [23] Niemeyer J, Rottensteiner F, Soergel U. Contextual classification of lidar data and building object detection in urban areas [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2014, 87: 152-165.
- [24] 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, June 20-25, 2009, Miami, FL, USA. New York: IEEE Press, 2009: 975-982.
- [25] Shapovalov R, Velizhev A, Barinova O. Non-associative Markov networks for 3D point cloud classification[J]. In *Proc ISPRS Commission 755 III Symp (PCV)*, 2010, 38: 103-108.
- [26] Li W Z, Wang F D, Xia G S. A geometry-attentional

- network for ALS point cloud classification[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020, 164: 26-40.
- [27] Ivakhnenko A G. Polynomial theory of complex systems[J]. *IEEE Transactions on Systems, Man, and Cybernetics*, 1971, SMC-1(4): 364-378.
- [28] Hatt M, Parmar C, Qi J Y, et al. Machine (deep) learning methods for image processing and radiomics [J]. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 2019, 3(2): 104-108.
- [29] Kong W C, Dong Z Y, Jia Y W, et al. Short-term residential load forecasting based on LSTM recurrent neural network [J]. *IEEE Transactions on Smart Grid*, 2019, 10(1): 841-851.
- [30] Sokkhey P, Okazaki T. Development and optimization of deep belief networks applied for academic performance prediction with larger datasets [J]. *IEEE Transactions on Smart Processing & Computing*, 2020, 9(4): 298-311.
- [31] Guo Y L, Wang H Y, Hu Q Y, et al. Deep learning for 3D point clouds: a survey[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(12): 5434-5449.
- [32] Lateef F, Ruichek Y. Survey on semantic segmentation using deep learning techniques [J]. *Neurocomputing*, 2019, 338: 321-348.
- [33] Xie Y X, Tian J J, Zhu X X. 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.
- [34] Zhang J Y, Zhao X L, Chen Z, et al. A review of deep learning-based semantic segmentation for point cloud[J]. *IEEE Access*, 2019, 7: 179118-179133.
- [35] Murase H, Nayar S K. Visual learning and recognition of 3-D objects from appearance [J]. *International Journal of Computer Vision*, 1995, 14(1): 5-24.
- [36] Haala N, Brenner C. Virtual city models from laser altimeter and 2D map data [J]. *Photogrammetric Engineering & Remote Sensing*, 1999, 65(7): 787-795.
- [37] Su H, Maji S, Kalogerakis E, et al. Multi-view convolutional neural networks for 3D shape recognition[C]//2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 945-953.
- [38] Feng Y F, Zhang Z Z, Zhao X B, et al. GVCNN: group-view convolutional neural networks for 3D shape recognition [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 264-272.
- [39] Wang C, Pelillo M, Siddiqi K. Dominant set clustering and pooling for multi-view 3D object recognition[C]//Proceedings of the British Machine Vision Conference 2017, September, 2017, London, UK. British: British Machine Vision Association, 2017: 1-12.
- [40] Yu T, Meng J J, Yuan J S. Multi-view harmonized bilinear network for 3D object recognition[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 186-194.
- [41] Endres F, Hess J, Sturm J, et al. 3-D mapping with an RGB-D camera [J]. *IEEE Transactions on Robotics*, 2014, 30(1): 177-187.
- [42] 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.
- [43] Guerry J, Boulch A, le Saux B, et al. SnapNet-R: consistent 3D multi-view semantic labeling for robotics[C]//2017 IEEE International Conference on Computer Vision Workshops (ICCVW), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 669-678.
- [44] Ma C, Guo Y L, Yang J G, et al. Learning multi-view representation with LSTM for 3-D shape recognition and retrieval [J]. *IEEE Transactions on Multimedia*, 2019, 21(5): 1169-1182.
- [45] Biasutti P, Lepetit V, Aujol J F, et al. LU-net: an efficient network for 3D LiDAR point cloud semantic segmentation based on end-to-end-learned 3D features and U-net[C]//2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), October 27-28, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 942-950.
- [46] Alonso I, Riazuelo L, Montesano L, et al. 3D-MiniNet: learning a 2D representation from point clouds for fast and efficient 3D LIDAR semantic segmentation [J]. *IEEE Robotics and Automation Letters*, 2020, 5(4): 5432-5439.
- [47] Maturana D, Scherer S. VoxNet: a 3D convolutional neural network for real-time object recognition[C]//2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), September 28-October 2, 2015, Hamburg, Germany. New York: IEEE Press, 2015: 922-928.
- [48] Wu Z R, Song S R, Khosla A, et al. 3D ShapeNets: a deep representation for volumetric shapes[C]//2015

- IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 1912-1920.
- [49] Riegler G, Ulusoy A O, Geiger A. OctNet: learning deep 3D representations at high resolutions[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6620-6629.
- [50] Wang P S, Liu Y, Guo Y X, et al. O-CNN: octree-based convolutional neural networks for 3D shape analysis[J]. *ACM Transactions on Graphics*, 2017, 36(4): 1-11.
- [51] Xu Y, Hoegner L, Tutas S, et al. Voxel- and graph-based point cloud segmentation of 3D scenes using perceptual grouping laws[J]. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2017, IV-1/W1: 43-50.
- [52] Klovov R, Lempitsky V. Escape from cells: deep Kd-networks for the recognition of 3D point cloud models[C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 863-872.
- [53] Zeng W, Gevers T. 3DContextNet: K-d tree guided hierarchical learning of point clouds using local and global contextual cues[M]//Leal-Taixé L, Roth S. *Computer vision-ECCV 2018 workshops. Lecture notes in computer science*. Cham: Springer, 2019, 11131: 314-330.
- [54] Wang L, Huang Y C, Shan J, et al. MSNet: multi-scale convolutional network for point cloud classification [J]. *Remote Sensing*, 2018, 10(4): 612.
- [55] Li Y Y, Pirk S, Su H, et al. FPNN: field probing neural networks for 3D data[C]//NIPS'16: Proceedings of the 30th International Conference on Neural Information Processing Systems, December 5, 2016, Red Hook, NY, United States. New York: ACM, 2016: 307-315.
- [56] Le T, Duan Y. PointGrid: a deep network for 3D shape understanding[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 9204-9214.
- [57] Tchapmi L, Choy C, Armeni I, et al. SEGcloud: semantic segmentation of 3D point clouds[C]//2017 International Conference on 3D Vision (3DV), October 10-12, 2017, Qingdao, China. New York: IEEE Press, 2017: 537-547.
- [58] Meng H Y, Gao L, Lai Y K, et al. VV-net: voxel VAE net with group convolutions for point cloud segmentation [C] // 2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 8499-8507.
- [59] Hegde V, Zadeh R. FusionNet: 3D object classification using multiple data representations[EB/OL]. (2016-07-19)[2020-10-15]. <https://arxiv.org/abs/1607.05695>.
- [60] Qi C R, Yi L, Su H, et al. PointNet++: deep hierarchical feature learning on point sets in a metric space[C]//NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems, December 4, 2017, Red Hook, NY, United States. New York: ACM, 2017: 5099-5108.
- [61] Jiang M, Wu Y, Zhao T, et al. PointSIFT: a SIFT-like network module for 3D point cloud semantic segmentation[EB/OL]. (2018-07-02)[2020-10-15]. <https://arxiv.org/abs/1807.00652>.
- [62] Li J X, Chen B M, Lee G H. SO-net: self-organizing network for point cloud analysis [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 9397-9406.
- [63] Ma Y X, Guo Y L, Lei Y J, et al. 3DMAX-net: a multi-scale spatial contextual network for 3D point cloud semantic segmentation[C]//2018 24th International Conference on Pattern Recognition (ICPR), August 20-24, 2018, Beijing, China. New York: IEEE Press, 2018: 1560-1566.
- [64] Zhao H S, 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 (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 5560-5568.
- [65] Qiu S, Anwar S, Barnes N. Dense-resolution network for point cloud classification and segmentation[EB/OL]. (2020-05-14)[2020-10-15]. <https://arxiv.org/abs/2005.06734?context=cs.CV>.
- [66] Hu Q Y, Yang B, Xie L H, et al. RandLA-Net: efficient semantic segmentation of large-scale point clouds[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 11105-11114.
- [67] Scarselli F, Gori M, Tsoi A C, et al. The graph neural network model [J]. *IEEE Transactions on Neural Networks*, 2009, 20(1): 61-80.

- [68] Bruna J, Zaremba W, Szlam A, et al. Spectral networks and locally connected networks on graphs [EB/OL]. (2013-12-21) [2020-10-15]. <https://arxiv.org/abs/1312.6203>.
- [69] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks [EB/OL]. (2016-09-09) [2020-10-15]. <https://arxiv.org/abs/1609.02907>.
- [70] Simonovsky M, Komodakis N. Dynamic edge-conditioned filters in convolutional neural networks on graphs [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 29-38.
- [71] Landrieu L, Simonovsky M. Large-scale point cloud semantic segmentation with superpoint graphs [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 4558-4567.
- [72] Landrieu L, Boussaha M. Point cloud over segmentation with graph-structured deep metric learning [C] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 7432-7441.
- [73] Jiang L, Zhao H S, Liu S, et al. Hierarchical point-edge interaction network for point cloud semantic segmentation [C] // 2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 10432-10440.
- [74] Te G S, Hu W, Zheng A M, et al. RGCNN: regularized graph CNN for point cloud segmentation [C] // Proceedings of the 26th ACM international conference on Multimedia, October 15, 2018, Seoul Republic of Korea. New York: ACM, 2018: 746-754.
- [75] Wang Y, Sun Y B, Liu Z W, et al. Dynamic graph CNN for learning on point clouds [J]. *ACM Transactions on Graphics*, 2019, 38(5): 1-12.
- [76] Zhang K G, Hao M, Wang J, et al. Linked dynamic graph CNN: learning on point cloud via linking hierarchical features [EB/OL]. (2019-04-22) [2020-10-15]. <https://arxiv.org/abs/1904.10014>.
- [77] Liu J X, Ni B B, Li C Y, et al. Dynamic points agglomeration for hierarchical point sets learning [C] // 2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 7545-7554.
- [78] Li G H, Müller M, Thabet A, et al. DeepGCNs: can GCNs go as deep as CNNs? [C] // 2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 9266-9275.
- [79] Liang Z D, Yang M, Deng L Y, et al. Hierarchical depthwise graph convolutional neural network for 3D semantic segmentation of point clouds [C] // 2019 International Conference on Robotics and Automation (ICRA), May 20-24, 2019, Montreal, QC, Canada. New York: IEEE Press, 2019: 8152-8158.
- [80] Lu Q, Chen C, Xie W J, et al. PointNGCNN: deep convolutional networks on 3D point clouds with neighborhood graph filters [J]. *Computers & Graphics*, 2020, 86: 42-51.
- [81] Zhang W J, Su S Z, Wang B Z, et al. Local k-NNs pattern in omni-direction graph convolution neural network for 3D point clouds [J]. *Neurocomputing*, 2020, 413: 487-498.
- [82] Nezhadarya E, Taghavi E, Razani R, et al. Adaptive hierarchical down-sampling for point cloud classification [C] // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 12953-12961.
- [83] Li Y Y, Bu R, Sun M, et al. PointCNN: convolution on χ -transformed points [EB/OL]. (2018-01-23) [2020-10-15]. <https://arxiv.org/abs/1801.07791>.
- [84] Lan S Y, Yu R C, Yu G, et al. Modeling local geometric structure of 3D point clouds using geo-CNN [C] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 998-1008.
- [85] Mao J G, Wang X G, Li H S. Interpolated convolutional networks for 3D point cloud understanding [C] // 2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 1578-1587.
- [86] Komarichev A, Zhong Z C, Hua J. A-CNN: annularly convolutional neural networks on point clouds [C] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 7413-7422.
- [87] Rao Y M, Lu J W, Zhou J. Spherical fractal convolutional neural networks for point cloud recognition [C] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),

- June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 452-460.
- [88] Zhang Z Y, Hua B S, Rosen D W, et al. Rotation invariant convolutions for 3D point clouds deep learning[C]//2019 International Conference on 3D Vision (3DV), September 16-19, 2019, Quebec City, QC, Canada. New York: IEEE Press, 2019: 204-213.
- [89] Lei H, Akhtar N, Mian A. Octree guided CNN with spherical kernels for 3D point clouds [C] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 9623-9632.
- [90] Engelmann F, Kontogianni T, Leibe B. Dilated point convolutions: on the receptive field size of point convolutions on 3D point clouds[J]. (2019-07-28) [2020-10-15]. <https://arxiv.org/abs/1907.12046>.
- [91] Wang S L, Suo S, Ma W C, et al. Deep parametric continuous convolutional neural networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 2589-2597.
- [92] Xu Y F, Fan T Q, Xu M Y, et al. SpiderCNN: deep learning on point sets with parameterized convolutional filters [M] // Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11212: 90-105.
- [93] Wu W X, Qi Z, Fuxin L. PointConv: deep convolutional networks on 3D point clouds[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 9613-9622.
- [94] Shen Y R, Feng C, Yang Y Q, et al. Mining point cloud local structures by kernel correlation and graph pooling[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 4548-4557.
- [95] Thomas H, Qi C R, Deschaud J E, et al. KPConv: flexible and deformable convolution for point clouds [C] // 2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 6410-6419.
- [96] Liu Y C, Fan B, Xiang S M, et al. Relation-shape convolutional neural network for point cloud analysis [C] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 8887-8896.
- [97] Thomas N, Smidt T, Kearnes S M, et al. Tensor field networks: rotation- and translation-equivariant neural networks for 3D point clouds [EB/OL]. (2018-02-22) [2020-10-15]. <https://arxiv.org/abs/1802.08219?context=cs>. NE.
- [98] Cohen T S, Geiger M, Koehler J, et al. Spherical CNNs [EB/OL]. (2018-01-30) [2020-10-15]. <https://arxiv.org/abs/1801.10130>.
- [99] Poulernard A, Rakotosaona M J, Ponty Y, et al. Effective rotation-invariant point CNN with spherical harmonics kernels [C] // In 2019 International Conference on 3D Vision (3DV), September 16-19, Quebec City, QC, Canada. New York: IEEE Press, 2019: 47-56.
- [100] Boulch A. ConvPoint: continuous convolutions for point cloud processing[J]. Computers & Graphics, 2020, 88: 24-34.
- [101] Rensink R A. The dynamic representation of scenes [J]. Visual Cognition, 2000, 7(1/2/3): 17-42.
- [102] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems, December 4, 2017, Red Hook, NY, United States. New York: ACM, 2017: 5998-6008.
- [103] Paigwar A, Erkent O, Wolf C, et al. Attentional PointNet for 3D-object detection in point clouds[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June 16-17, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 1297-1306.
- [104] Hu Z Y, Zhang D B, Li S, et al. Attention-based relation and context modeling for point cloud semantic segmentation[J]. Computers & Graphics, 2020, 90: 126-134.
- [105] Yang B, Wang S, Markham A, et al. Robust attentional aggregation of deep feature sets for multi-view 3D reconstruction[J]. International Journal of Computer Vision, 2020, 128(1): 53-73.
- [106] Wang L, Huang Y C, Hou Y L, et al. Graph attention convolution for point cloud semantic segmentation[C] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 10288-10297.
- [107] Chen C, Fragonara L Z, Tsourdos A. GAPointNet: graph attention based point neural network for exploiting local feature of point cloud[J].

- Neurocomputing, 2021, 438: 122-132.
- [108] Yang J C, Zhang Q, Ni B B, et al. Modeling point clouds with self-attention and gumbel subset sampling [C] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 3318-3327.
- [109] Li H, Xiong P, An J, et al. Pyramid attention network for semantic segmentation[EB/OL]. (2018-05-25) [2020-10-15]. <https://arxiv.org/abs/1805.10180v1>.
- [110] Kang Z H, Li N. PyramNet: point cloud pyramid attention network and graph embedding module for classification and segmentation[EB/OL]. (2019-06-07) [2020-10-15]. <https://arxiv.org/abs/1906.03299v1>.
- [111] Chen L Z, Li X Y, Fan D P. LSA-Net: feature learning on point sets by local spatial attention[EB/OL]. (2019-05-14) [2020-10-15]. <https://arxiv.org/abs/1905.05442>.
- [112] Fu J, Liu J, Tian H J, et al. Dual attention network for scene segmentation [C] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 3141-3149.
- [113] Xie S N, Liu S N, Chen Z Y, et al. Attentional ShapeContextNet for point cloud recognition [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 4606-4615.
- [114] Zhao C X, Zhou W H, Lu L, et al. Pooling scores of neighboring points for improved 3D point cloud segmentation[C]//2019 IEEE International Conference on Image Processing (ICIP), September 22-25, 2019, Taipei, Taiwan, China. New York: IEEE Press, 2019: 1475-1479.
- [115] Feng M T, Zhang L, Lin X F, et al. Point attention network for semantic segmentation of 3D point clouds [J]. Pattern Recognition, 2020, 107: 107446.
- [116] Wang J J, Zhong Y F, Zheng Z, et al. RSNet: the search for remote sensing deep neural networks in recognition tasks [J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(3): 2520-2534.
- [117] Li R H, Li X Z, Heng P A, et al. PointAugment: an auto-augmentation framework for point cloud classification [C] // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 6377-6386.
- [118] Schmitt M, Zhu X X. Data fusion and remote sensing: an ever-growing relationship [J]. IEEE Geoscience and Remote Sensing Magazine, 2016, 4(4): 6-23.
- [119] Cheng X, Zheng Y H, Zhang J W, et al. Multitask multisource deep correlation filter for remote sensing data fusion [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020, 13: 3723-3734.
- [120] Wang Y Y, Zhu X X, Zeisl B, et al. Fusing meter-resolution 4-D InSAR point clouds and optical images for semantic urban infrastructure monitoring [J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(1): 14-26.