

面向三维点云的深度学习分类方法综述

程春旭, 任 爽

(北京交通大学 计算机科学与技术学院, 北京 100044)

摘 要: 随着三维点云在自动驾驶、地图测绘等领域的广泛应用以及深度学习的快速发展,点云分类已成为三维视觉领域的研究热点。从三维点云表示方式出发,对基于深度学习的点云分类方法进行全面的分类与综述。首先,介绍点云分类常用的数据集及评价指标;其次,着重论述了具有代表性的分类方法,并梳理了近年来的最新研究工作,并对不同分类方法的性能进行了分析与对比;最后,分析现存的问题和挑战,并探索了未来可能的研究方向。

关键词: 点云分类;深度学习;三维视觉;特征提取

DOI: 10.11907/rjdk.241831

中图分类号: TP18

文献标识码: A

文章编号: 1672-7800(2026)001-0191-11

扫描二维码阅读全文:



Overview of Deep Learning Classification Methods for 3D Point Clouds

CHENG Chunxu, REN Shuang

(School of Computer Science and Technology, Beijing Jiaotong University, Beijing 100044, China)

Abstract: With the widespread application of 3D point clouds in autonomous driving, mapping, and other domains, alongside the rapid development of deep learning, point cloud classification has become a research hotspot in the field of 3D vision. In order to promote future in-depth explorations and to inject new vigor and ideas into the 3D point cloud classification task, this paper conducts a comprehensive classification and review of deep learning-based point cloud classification methods, starting from the representation of three-dimensional point clouds. Firstly, it introduces commonly used datasets and evaluation metrics for point cloud classification. Secondly, it focuses on representative classification methods and summarizes the latest research works in recent years. Subsequently, it analyzes and compares the performance of these classification methods. Finally, it examines existing issues and challenges and explores potential future research directions.

Key Words: point cloud classification; deep learning; 3D vision; feature extraction

0 引言

随着3D扫描仪、激光雷达等三维传感器的快速普及,3D数据的采集和获取越来越便利,人们对三维视觉的关注程度也不断提高。点云作为最常用的数据表示形式,完整地保留了物体原始几何信息,其因具有丰富的形状和尺度细节成为很多研究者的首选。因此,点云在自动驾驶^[1]、三维重建^[2]、医学辅助^[3]等诸多领域得到了广泛应用。

三维点云是一组存在于三维坐标空间下的无序点的集合。点云分类是指将具有相同或相似属性的同一点云集按照相关特征定义进行分类,并用标签对分类的点云集

进行标记的过程。然而,由于点云数据具有不规则性、无序性和稀疏性等特点,以及从现实世界获取的点云存在噪声大、遮挡等因素,直接处理点云变得非常困难,点云分类仍然是一项极具挑战性的任务^[4]。

早期的点云分类大多采用“手工设计规则+分类器分配标签”的方法,通过将手动提取的点云特征与人工标记的语义信息作为分类学习规则进行大量训练。常见的代表性分类器有支持向量机^[5]、随机森林^[6]等。然而,这些分类器所提取的点云特征无法充分反映不同物体的特征属性,同时忽略了领域点的相关性,很容易导致分类准确率降低。近年来,深度学习技术在二维图像领域取得了显著成果,特别是随着卷积神经网络(Convolutional Neural Network, CNN)的广泛应用,研究者们逐渐尝试利用卷积处

收稿日期:2024-10-22

基金项目:国家自然科学基金项目(62072025)

作者简介:程春旭(2000-),女,北京交通大学计算机科学与技术学院硕士研究生,研究方向为点云分类;任爽(1981-),男,博士,北京交通大学计算机科学与技术学院副教授、博士生导师,研究方向为机器学习、计算机视觉、虚拟现实技术。本文通讯作者:任爽。

理三维点云分类问题。

2017年, Qi等^[7]开创性地提出PointNet网络模型, 直接将点云作为输入, 通过使用T-Net转换网络和最大池化(max-pooling)克服了点云排列不变性和旋转刚性的难题。此后, 越来越多学者在PointNet基础上进行深入研究, 点云分类任务取得了重大突破, 直接处理原始点云的深度学习方法也逐渐成为分类任务的主流^[8-11]。

目前国内外关于点云分类的综述很多, 但大多是对经典网络的详细描述, 分类体系尚不完善, 且未能及时涵盖近年来的新方法, 限制了读者对前沿动态的把握^[12-15]。为此, 本文在前人工作的基础上进行了归纳与完善。首先, 对现有点云分类方法进行了详细的分类和阐述, 包括基于多视图、基于体素、基于原始点云的三维点云分类方法, 其中又将基于原始点云的方法细化为基于逐点多层感知机(Multi Layer Perceptron, MLP)、基于卷积、基于图卷积和基于注意力4类; 其次, 整理了各类方法在不同数据集上的性能表现; 最后, 针对现存的问题和面临的挑战作进一步分析, 探讨了点云分类技术未来的发展方向。

1 三维点云数据介绍

1.1 常用数据集

点云数据是一种离散化的数据表示方式, 常见的获取方式有: ①通过三维激光扫描仪直接获取; ②利用摄影测量的方法对多张具有一定重叠度的影像进行立体匹配获取; ③使用RGB-D相机同时获取多帧彩色影像和深度图, 并结合相机内参将深度图反投至三维空间, 从而生成点云。随着三维视觉技术的发展, 越来越多研究机构开放了针对三维点云分类相关的数据集。常用的点云数据集如表1所示。

Table 1 Common datasets for point cloud classification

表1 点云分类常用数据集

数据集	年份	样本	类别	训练集	测试集	表示形式
ModelNet10	2015	4 899	10	3 991	605	Mesh
ModelNet40	2015	12 311	40	9 843	2 468	Mesh
ShapeNet	2015	51 300	3 135	-	-	Mesh
ScanObjectNN	2019	2 902	15	2 321	581	Point Cloud

(1) ModelNet。该数据集是2015年由普林斯顿大学视觉与机器人实验室发布的三维CAD模型数据集, 共提供了两个版本: ModelNet40和ModelNet10。ModelNet40包含40个类别和12 311个精细模型。ModelNet10为ModelNet40的子集, 包含10个类别和4 899个模型, 且所有模型都沿着重力轴方向进行了人工姿态对齐^[16]。

(2) ShapeNet。该数据集是由斯坦福大学、普林斯顿大学和美国芝加哥丰田技术研究所共同创建的大型三维CAD库, 包含3 135个语义类别和51 300个CAD模型。ShapeNetCore是ShapeNet的一个子集, 涵盖了55个常见的对象类别; ShapeNetSem是一个更小、注释更密集的子集,

包含270个类别和12 000个模型^[17]。

(3) ScanObjectNN。该数据集是由真实世界室内场景扫描提取的遮挡对象组成, 提供了约15 000个对象的真实样本, 覆盖了15种类别, 每类包含2 902个独立实例。数据不仅包括全局和局部坐标、法线、颜色属性和语义标签, 而且首次在真实数据上提供了部分标记, 拓宽了点云处理的应用范围^[18]。

1.2 评价指标

对于点云分类任务, 最常见的性能评价指标包括: 总体精度(Overall Accuracy, OA)和平均精度(Mean Accuracy, mAcc)。OA是指模型正确分类的样本数与正类预测样本数的比值, 计算公式如下:

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

其中, TP_i 表示正类判定为正类, FP_i 表示负类判定为正类, N 表示类别数。

mAcc是指正类预测样本数与总样本数的均值, 计算公式如下:

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

其中, FN_i 表示正类判定为负类, TN_i 表示负类判定为负类。

2 三维点云分类方法

相较于传统方法, 深度学习因具有无需人工参与设计和数据驱动学习的优势, 被广泛应用于点云分类任务中。如图1所示, 本文根据三维点云的表示方式, 将面向三维点云的深度学习分类方法划分为基于多视图、基于体素、基于原始点云的方法和其他方法(多态融合+大模型)四大类。图2给出了不同分类方法的发展时间线。

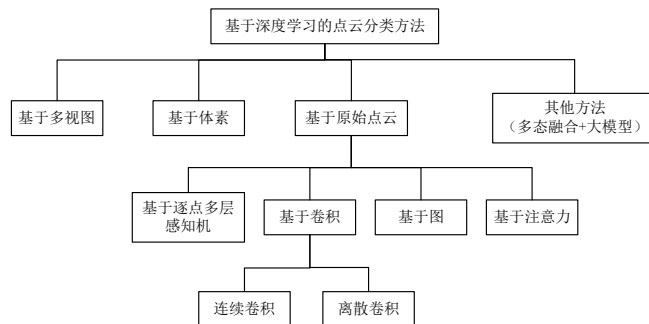


Fig. 1 Point cloud classification methods based on deep learning

图1 基于深度学习的点云分类方法

2.1 基于多视图的方法

在二维图像分析领域, 卷积神经网络(CNN)已成为一种广泛应用的强大工具, 尤其在图像分类任务上表现出色。这些卓越成果激励了学者们将类似技术扩展到三维数据处理中。然而, 不同于二维图像, 三维点云数据通常

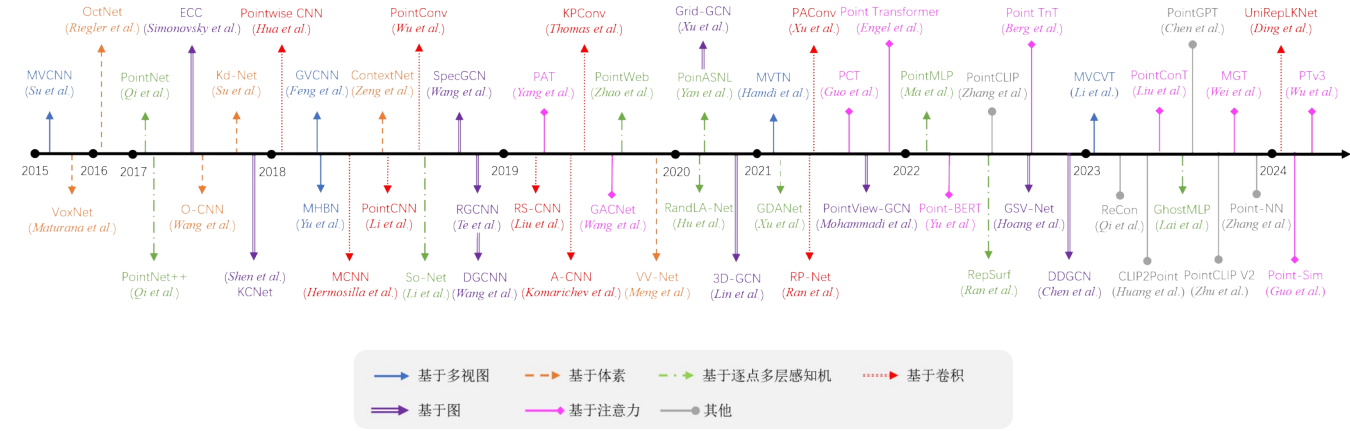


Fig. 2 Development timeline of point cloud classification methods based on deep learning
图2 基于深度学习的点云分类方法发展时间线

是非结构化的,从而使直接应用卷积操作变得非常困难。为了解决该问题,研究者们尝试将三维点云从不同方向映射到二维视图中,允许传统 CNN 提取每个视图的关键特征,以实现精确的形状分类。

2015 年, Su 等^[19]首次提出多视图卷积神经网络 (Multi-View CNN, MVCNN),该方法从不同视角捕获 3D 形状,生成多个 2D 视图,之后通过卷积层 (CNNs) 和池化层 (View pooling) 将这些视图的特征信息聚合到一个紧凑的三维形状描述符中,最后将聚合特征输入到全连接层 (Fully Connected Layers, FC) 进行分类。其原理架构如图 3 所示。然而,该方法投影检索过程的计算成本较高,且采用的最大池化方法只保留了特定视图中最显著的特征,导致大量信息丢失。针对这些问题, Bai 等^[20]提出一种实时三维形状检索引擎,利用 GPU 加速实现了高效的投影和视图特征提取。 Zeng 等^[21]在最大池化的基础上引入平均池化,通过训练最大池化层和平均池化层分配最优权重。 Wang 等^[22]引入循环聚类 and 基于优势集的池化层对 MVCNN 方法进行改进,提出了 RCPCNN (Recurrent Clus-

ter-Pooling CNN) 方法,大幅提高了多视图三维物体识别能力。

Feng 等^[23]在 MVCNN 基础上作进一步改进,提出组视图卷积神经网络 (Group-View CNN, GVCNN) 对不同视图的特征进行分组,生成组级描述符,再将每个组的特征进行池化得到形状级描述符以实现分类,充分考虑了多视图间的相关性。 Yu 等^[24]提出的 MHBN (Multi-view Harmonized Bilinear Networ) 采用协调双线性池化技术,将局部卷积特征融合为更紧凑、具有更强辨别能力的全局描述符。之后, Hamdi 等^[25]提出多视图变换网络 (Multi-View Transformation Network, MVTN),其使用可微分渲染器回归 3D 形状识别的最佳视点,并以端到端的方式训练下游任务特定网络,在没有额外训练监督的情况下性能获得了较大提升。

最近,受视觉变压器 (Vision Transformer, ViT) 成功应用于图像识别的启发, Li 等^[26]提出多视图卷积视觉变压器 (Multi-view Convolutional Vision Transformer, MVCVT),该方法使用 CNN 提取多尺度局部信息,利用 transformer 捕获

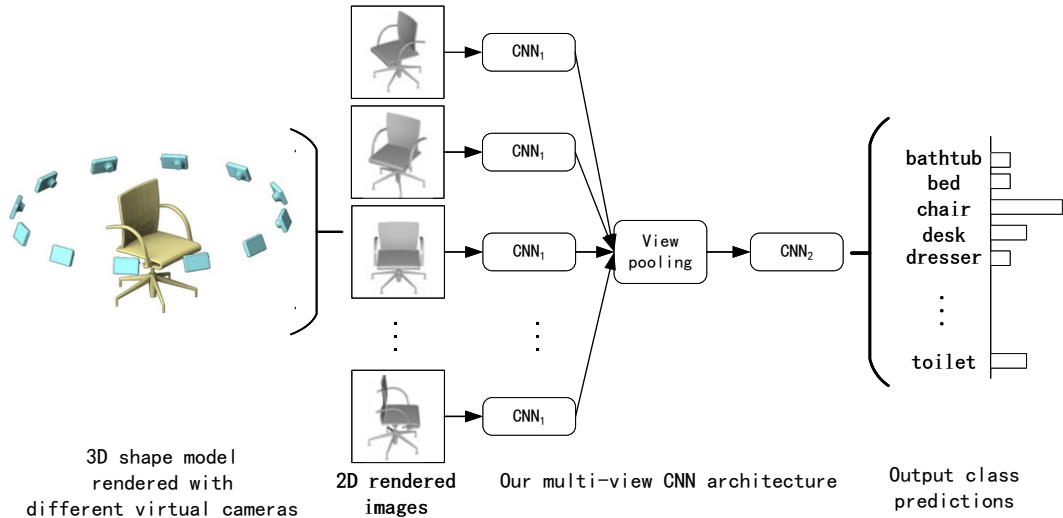


Fig. 3 Basic architecture of MVCNN
图3 MVCNN基本架构

多尺度信息在不同视图中的相关性,充分验证了Transformer在基于多视图的3D形状分类任务中的适应性和有效性。

综上所述,基于多视图的分类方法有效解决了手动提取特征导致的部分特征丢失问题,在小规模场景的点云分类中取得了不错的效果。但多视图容易造成物体的自遮挡和特征信息重叠,视点位置和角度的选择对分类结果影响很大。因此,该方法在信息充分利用以及大规模场景应用等方面仍存在挑战。

2.2 基于体素的方法

基于体素的点云分类方法通常是将点云体素化为最接近物体的3D体素网格,再利用3D CNN进行形状分类。2015年,Wu等^[16]提出卷积深度信念网络3D ShapeNets,将几何3D形状表示为3D体素网格上二元变量的概率分布,以数据驱动的方式学习所有3D体素的复杂联合分布。同年,Maturana等^[27]设计了一种名为VoxNet的体积占用网络,该网络由输入层、两个卷积层、一个最大池化层和两个全连接层组成,每一层都将传统的二维卷积扩展到三维,有效证明了体积卷积的可实现性。虽然上述方法都取得了突破性进展,但随着体素分辨率的增加,占用的体素变得更加稀疏,不可避免地带来大量不必要的计算。为解决这些问题,一些方法尝试使用其他索引结构,如八叉树^[28]、Kd树^[29-30]等。

2018年,Le等^[31]提出一种混合模型PointGrid,通过点和网格的集成,可以更好地表示局部几何形状的细节。其主要思想是:首先将点云数据固定在一定大小的立方体中,然后在每个立方体中随机选取k个点的特征作为对应立方体的特征,最后使用CNN提取几何细节。2019年,Meng等^[32]提出体素变分自编码器网络(Voxel VAE net, VV-Net),每个体素内的点分布由自编码器捕捉。该编码器采用径向基函数(Radial Basis Functions, RBF),既提供

了规则结构,又能获取详细的数据分布,可有效处理嘈杂的点云数据。

基于体素的点云分类方法有效解决了三维点云无序和非结构化问题,但点云固有的稀疏性容易导致有些体素分辨率过低。此外,体素大小的不同也会影响分类结果:分辨率过低会丢失物体细节,而分辨率过高又会增加计算量。尽管目前已有一些方法对这些问题进行了改进,但如何实现体素大小的自适应选择仍是提高分类准确性的关键研究方向。

2.3 基于原始点云的分类方法

基于多视图和基于体素的分类方法都需要提前对点云数据进行预处理,将不规则的点云数据转换成规则的数据,这两种方法在转换过程中容易丢失某些点云数据特性。为此,研究者们尝试直接从原始点云中学习特征信息。本文将基于原始点云的分类方法又细分为:基于逐点MLP、基于卷积、基于图和基于注意力的方法4类。

2.3.1 基于逐点MLP的方法

2017年,Qi等^[7]提出PointNet,开创了逐点特征提取的先河。该方法创新性地使用共享多层感知机提取每个点的特征,以及使用最大池化提取全局特征,消除了点云无序性的干扰,使得点云数据可直接用于训练。随后Qi等^[8]又在PointNet基础上提出优化网络PointNet++。如图4所示,该网络引入一种层次结构,包括采样层、分组层和特征提取层。采样层基于最远点采样(Farthest Point Sampling, FPS)迭代选出足够数量的中心点;分组层基于球查询方式构建中心点的固定半径局部区域;特征提取层利用PointNet获取局部特征向量。然而,PointNet++依旧未能考虑局部区域点集间的相互关系。为此,后续出现了很多基于PointNet或PointNet++框架的优化方法,如So-Net^[33]、PointWeb^[34]、PointMLP^[35]和PointNeXt^[9]等。

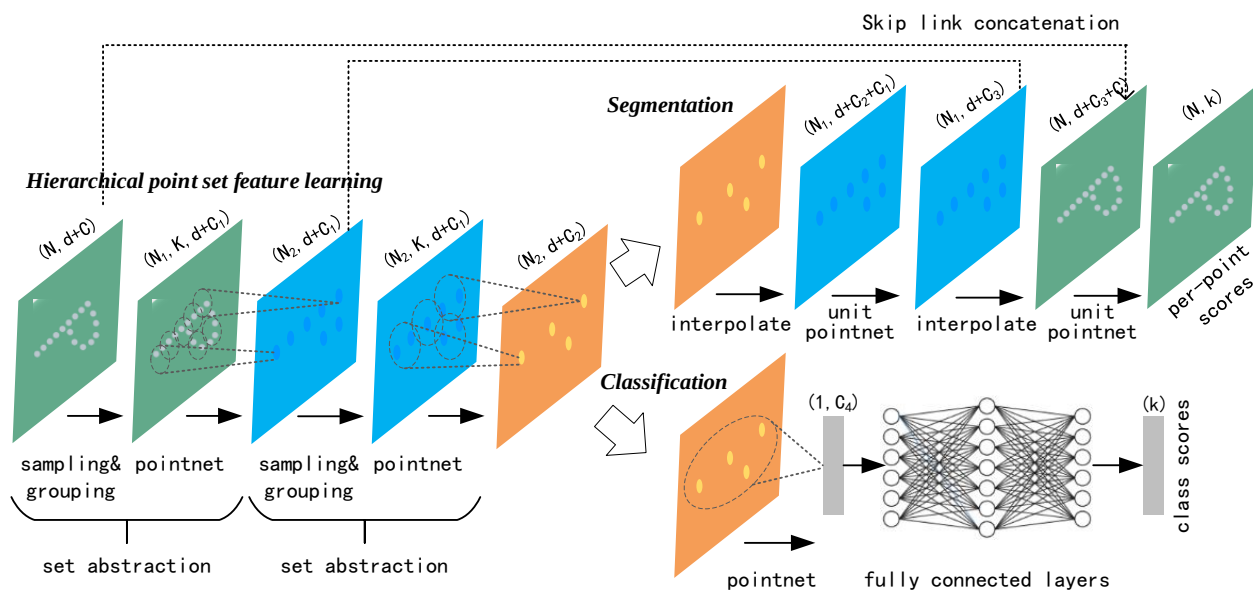


Fig. 4 The architecture of PointNet++

图4 PointNet++框架

考虑到网络性能和效率问题,许多研究开始重新思考轻量化网络设计和训练策略。2022 年, Ma 等^[35]提出的 PointMLP 引入轻量级局部几何仿射模块来自适应转换局部点特征,仅通过几个残差 MLP 分层提取局部特征,不仅提高了鲁棒性,而且计算更快、模型复杂度更低。其框架如图 5 所示。PointNeXt 则是在 PointNet++ 框架上引入倒置残差 MLP (Inverted Residual MLP, InvResMLP) 模块和可分

离 MLP 模块,从而实现了更少的模型参数和更快的运行速度^[9]。2023 年, Lai 等^[36]利用 GhostNet 基本组件替换 PointMLP 模型中的 MLP 层,提出轻量级高性能网络 GhostMLP,同样以较低的计算要求实现了卓越的分类性能。Yang 等^[37]利用点云栅格化模块 Gridding 和轮廓特征编码模块 SFE 块提取并聚合了点云的局部与轮廓特征,在轻量化重构的同时保证了网络的分类精度。

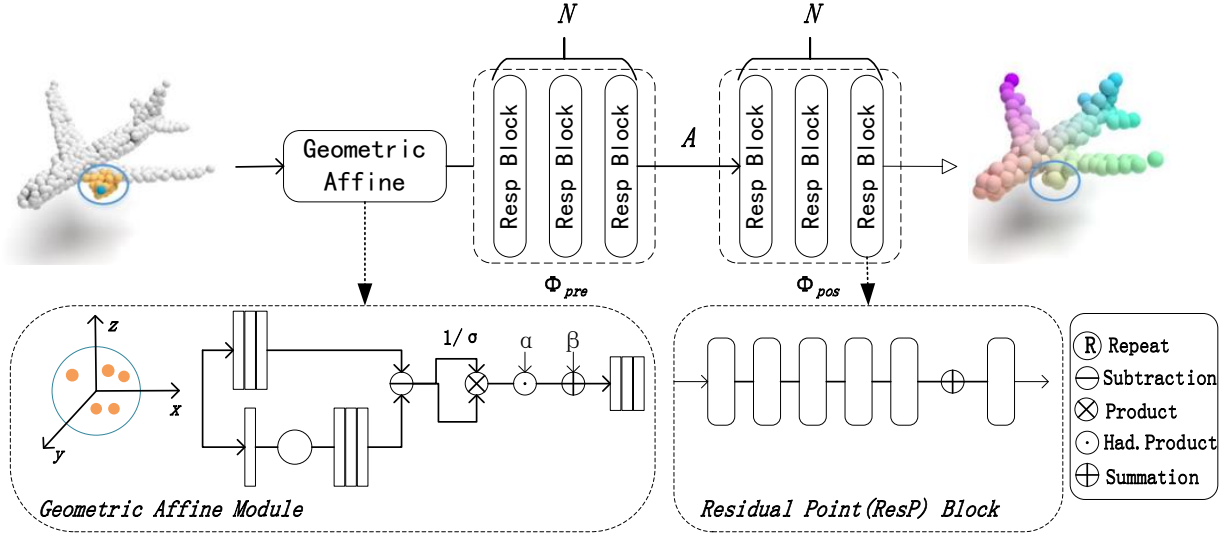


Fig. 5 The architecture of PointMLP

图 5 PointMLP 框架

综上所述,基于逐点 MLP 的方法解决了点云无序性和置换不变性问题,实现了较好的分类效果,在网络训练效率方面表现出色,目前已成为点云研究领域的主流方法。然而,由于捕获远程依赖关系和整体上下文的限制,在处理大规模和复杂点云时仍面临挑战。此外,研究者们大多在类似 ModelNet40 的干净数据集上进行训练,而忽略了点云缺失、重叠等情况,这些因素对网络的鲁棒性也有着潜在影响。

2.3.2 基于卷积的方法

根据权重函数的不同,可将基于点云的卷积分为离散卷积和连续卷积,如图 6 所示。离散卷积一般涉及在结构化点网格上滑动的固定大小的核,相邻点的权重分配与其相对于核中心点的偏移量有关。相反,连续卷积则定义为在连续空间中,权重函数直接依赖于邻域点相对于中心点的空间分布^[38]。本文将根据这两种卷积类型对基于卷积的点云分类方法进行归纳和总结,并按照时间顺序对代表性的卷积网络进行详细阐述。

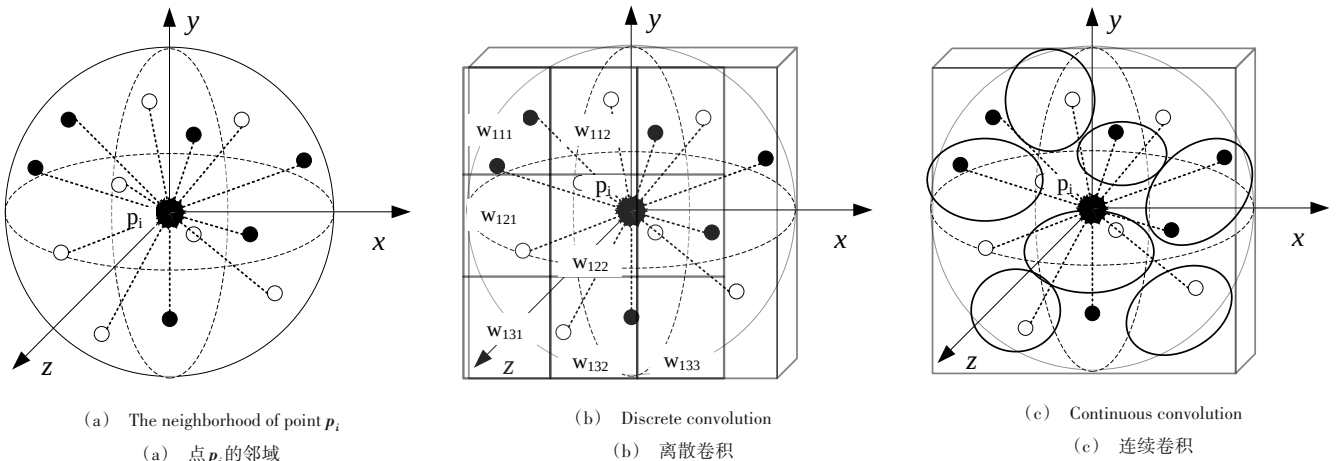


Fig. 6 The neighborhood centered on point p_i and its corresponding discrete convolution and continuous convolution

图 6 以点 p_i 为中心的领域及其对应的离散卷积和连续卷积

(1) 离散卷积。在离散卷积方面,一些研究致力于解决点云的排列不变性问题,另一些研究则考虑了点的空间几何特性。针对点云的排列不变性问题,Hua等^[39]将非均匀点云按照指定顺序进行排序,并输入到逐点卷积神经网络(Pointwise CNN)中,以每个点为中心得到指定数量的近邻点并将其划分到对应的核单元(Kernel Cell)。核的尺寸可以根据每个卷积层中不同数量的近邻点进行适当地调整。类似地,Li等^[40]考虑将变换应用于输入点坐标,定义了一种 χ -Conv算子,并提出点卷积神经网络PointCNN。 χ -Conv算子通过MLP学习点云的变换矩阵,之后通过变换矩阵对点云特征进行加权和置换,最后使用标准一维卷积进行处理得到高维特征。在点的空间几何特性方面,Lan等^[41]基于物体的几何信息隐含在点云坐标中这一假设,提出一种分层式特征提取框架(Geometric-induced CNN, Geo-CNN)。其分层提取特征,并沿层级保留了点的几何结构。Komarichev等^[42]提出一种环形卷积网络A-CNN,通过计算点的法向量生成正交切面,以实现点云的环形区域划分,并对环内的点依据余弦值进行排序,组合构建卷积序列。

(2) 连续卷积。连续卷积方法是在连续的空间中定义卷积算子,其中相邻点的权重与中心点的空间分布有关。连续卷积运算的关键在于其特征建模机制的设计,现有方法主要采用两种策略:一是增加一些虚拟化的辅助信息,二是对所使用的卷积核函数进行参数化^[38]。例如,

KPConv使用核点代替像素定义权值的应用区域,允许卷积核根据输入点云的局部几何结构动态调整形状,从而更好地捕捉空间信息^[43]。PAConv根据中心点及其近邻点之间的相对位置关系自适应生成权重矩阵,并通过组合存储在权重库中的基本权重矩阵与网络预测的位置自适应系数来构造动态卷积^[44]。SpiderCNN将卷积定义为分段权重函数与几何多项式基函数的加权求和,以确保卷积足够复杂以捕获复杂的局部几何变化^[45]。PointConv利用核密度估计生成密度尺度来重新加权学习的权重函数,实现了点云置换不变性和平移不变形^[46]。RS-CNN使用MLP学习从欧几里得距离和相对位置等低级关系到局部子集中点之间高级关系的映射,实现上下文形状感知学习^[47]。

综上所述,对于连续卷积方法,如PointConv^[46]和KPConv^[43],提高了模型对不同点云几何形状的适应性,然而忽略了点云分布。对于离散卷积方法,PointCNN^[40]可有效处理无序结构,Pointwise CNN^[39]在逐点特征学习方面表现出色。虽然这些方法都表现出良好的分类效果,但仍面临着计算成本高和性能不足等方面的问题。

2.3.3 基于图的方法

图作为处理非结构化数据的强大工具,在处理点云数据时具有显著优势。基于图的点云分类方法实质上是基于图卷积的方法。其将点云中的每个点作为图的顶点,通过点的邻居关系构建有向边,最后在空间域或谱域中进行特征学习。其典型架构如图7所示。

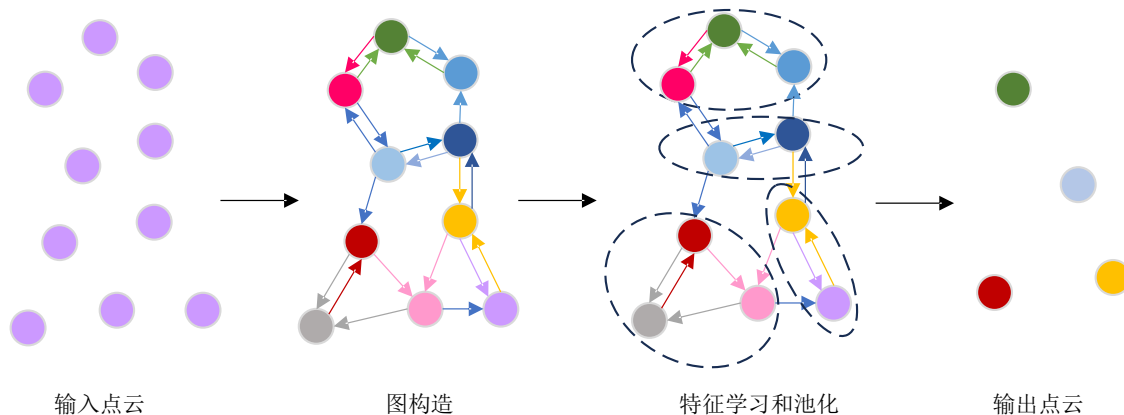


Fig. 7 Graph-based 3D point cloud network

图7 基于图的三维点云网络

Simonovsky等^[48]将图卷积引入到点云数据分类中,在空域上定义了利用边标签的局部图邻域卷积算子——边条件卷积(Edge-Conditioned Convolution, ECC),将点作为图的顶点,点之间的距离作为权重进行加权平均卷积操作,并使用最大采样聚合顶点信息,将点云分类转化为网络上的图分类任务。KCNet在PointNet架构上作进一步改进,采用两个不同的操作:一是将内核定义为可学习点的集合,然后确定内核与特定点邻域之间的从属关系;二是在三维位置计算的最近邻图上进行特征聚合^[49]。考虑到已有方法中点特征学习都是独立的,容易忽略相邻点的相

对布局,Wang等^[50]将局部光谱特征学习与递归聚类 and 池化策略相结合,提出一种新架构SpecGCN。Wang等^[51]提出的动态图卷积神经网络(Dynamic Graph CNN, DGCNN)同样解决了独立处理每个点的局限性问题,该网络使用EdgeConv层提取中心点特征以及中心点与K近邻点的边缘向量,在保持置换不变性的同时很好地捕获了局部几何结构。

上述方法充分保留了数据粒度,并且考虑了点之间的相互关系。然而,其在数据结构(如FPS和Ball Query)上花费了大量时间,限制了速度和可扩展性。为此,后续涌现

出很多优秀的采样方法^[52-53]。之后在二维卷积的驱动下, Lin 等^[54]提出三维图形卷积网络(3D Graph Convolution Networks, 3D-GCN),采用图最大池化机制定义可变形三维核,其形状和权重在训练阶段是可学习的。Zhou 等^[11]则提出自适应图卷积算子(Adaptive Graph Convolution, AdaptConv),根据动态学习的特征生成自适应卷积核,提高了点云卷积的灵活性。近年来,Chen 等^[55]提出融合距离和方向的分类网络,通过点云相似矩阵构建动态邻域图,并结合中心损失修改损失函数,使点云数据中的类内间距更加紧凑,增强了模型的判别能力。

总体而言,基于图的方法为处理不规则点云数据提供了另一种途径,但将这些方法(尤其是基于谱域图结构的方法)扩展到各种图配置中仍然很具有挑战性。

2.3.4 基于注意力的方法

基于注意力机制的点云分类方法是一种利用注意力机制来提高点云特征表示和分类能力的方法,其可以根据输入数据的不同维度分配不同权重,从而突出重要信息,抑制无关信息。近年来,随着注意力机制在自然语言处理和计算机视觉领域的不断发展,一些研究者将其应用于点云分割、识别、配准等任务中,为点云的特征提取和聚合方式提出新的设计范式。

2017年,Vaswani 等^[56]提出Transformer架构,将注意力机制推向了新高度。Transformer是一种解码器编码器结构,主要包含输入(单词)嵌入、位置(顺序)编码和自注意力等模块。其中,自注意力模块作为核心组件,首先将输

入嵌入和位置编码的总和作为输入,并通过训练好的线性层计算每个单词的3个向量:查询、键和值,然后通过匹配其查询和键向量(点积)来获得任意两个单词之间的注意力权重,最后将注意力特征定义为具有注意力权重的所有值向量的加权和^[10]。

Guo 等^[10]在Transformer原理基础上提出Point Cloud Transformer(PCT)模型,其采用的隐式拉普拉斯算子和归一化函数的offset-attention模块,与Transformer中的原始自注意模块相比,具有固有的置换不变性,更适合点云学习。具体架构如图8所示。Engel 等^[57]引入另一个称为Point Transformer的模型,该模型直接对无序和非结构化点集进行操作,使用局部-全局注意机制来捕获空间点关系和形状信息。Point Transformer V2(PTv2)是Point Transformer体系结构的增强版本,其引入了两个创新:分组向量注意力和基于分区的池化^[58]。分组向量注意力通过仅在点组内执行注意力来降低计算成本,从而在学习远程依赖时保持准确性;基于分区的池化通过将大点云划分为更小的非重叠分区和每个分区内的池化特征来提高大点云的准确性,从而实现计算负载减少的全局特征学习。Wu 等^[59]近年来提出的Point Transformer V3(PTv3)以其简单、高效和卓越性能在点云处理领域引起了广泛关注。PTv3用序列化点云定制的简化方法取代了更复杂的注意力补丁交互机制,同时取消了对相对位置编码的依赖,而是采用更简单的预正稀疏卷积层,实现了更快的点云处理速度。

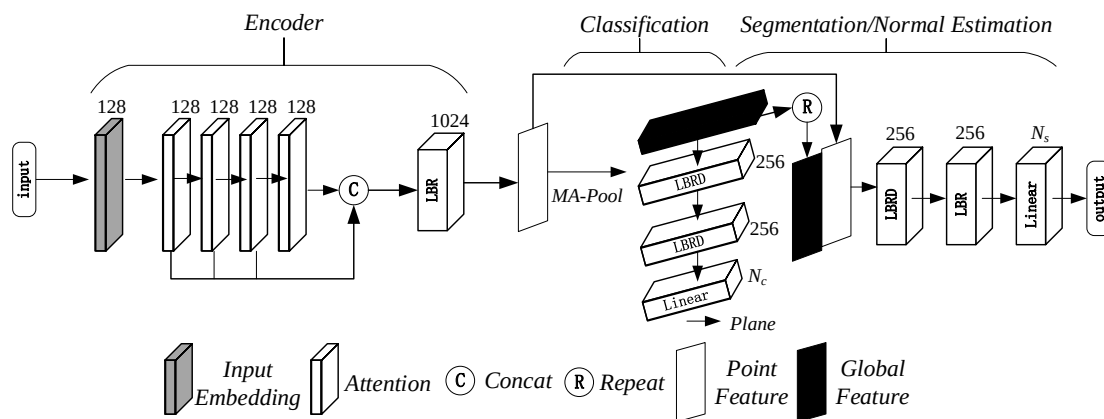


Fig. 8 The architecture of PCT

图8 PCT架构

尽管早期的点云网络尝试利用注意力机制,但由于数据规模和三维数据标定的限制,其性能未能得到充分发挥。为此,Yu 等^[60]将BERT概念应用于点云,提出Point-BERT网络,将点云局部领域视为点云词汇并进行token转换,增强了点云上下文关系的捕获能力。Pang 等^[61]则考虑到点云位置的不均匀信息密度和信息泄漏相关问题,进一步提出Point-MAE方法,在简化架构的同时提高了预训练效率。尽管这些方法利用外部数据辅助点云模型提高了性能,但并没有完全解决点云结构表示的问题。为此,

Wang 等^[62]创新地基于全局感知和局部结构拟合Transformer,能够以很高的精度从点云中学习详细的形状信息。此外,点云处理领域还不断涌现出诸如PointConT^[63]、Point-TnT^[64]等新颖且富有成效的方法,展现了Transformer技术的强大能力。

为了更好地提升点云分类网络的性能、节约计算资源,一些研究开始尝试使用非参数方法来提取特征,如Point-NN^[65]。其结合非参数模块构建非参数网络,非参数组件包括三角嵌入、最远点采样(FPS)、k近邻(k-NN)和池

化等操作。但 Point-NN 在特征提取时利用三角函数进行特征嵌入存在一定的盲目性。为了尽可能消除这种盲目性,Guo 等^[66]提出基于非参数能量函数的注意力机制,并构建了非参数网络 Point-Sim。该网络通过能量函数计算特征的能量来增强嵌入特征,从而实现在不向原始网络添加任何参数的情况下增强特征。

2.4 其他点云分类方法

近年来,随着多模态大模型的不断发展以及业界对模型轻量化的要求越来越高,三维点云分类方法除基于逐点 MLP、卷积、图和注意力 4 类方法外,越来越多研究开始探索多模态融合方法^[67-68]以及大模型背景下的分类方法^[69-70]。

2022 年,Zhang 等^[67]提出的 PointCLIP 在预训练的 CLIP(Contrastive Language-Image Pre-training)基础上对点云进行学习,在不渲染的情况下将点云投影到多视图深度图中来编码点云,并对视图上的零点预测进行聚合,从而实现从 2D 到 3D 的知识迁移。2023 年,Huang 等^[68]认为 PointCLIP 中渲染的深度图与 CLIP 训练数据集的图像域有很大不同,点云数据的稀疏性和无序性也会导致多个视图的深度分布不同,从而混淆了 CLIP 的聚合,致使 PointCLIP 结果仍具有局限性。为此,其提出了 CLIP2Point,该方法提供了一种具有两类学习机制的预训练方案:①RGB 图像与深度图的跨模态对比学习;②深度模态内的自监督学习,以增强深特征聚合的不变性。Chen 等^[69]提出的新型 GPT 模型 PointGPT 通过自回归生成任务的预训练,避免了位置信息泄漏问题,提高了模型的泛化能力。此外,也有研究使用知识蒸馏和对比学习的方法,在有限的实现高效的 3D 表征提取,如 ReCon^[71]。

综上所述,自从 PointNet++ 起,后续分类方法为了提升性能,一方面设计了更复杂的局部空间算子,另一方面增加了网络的可学习参数量。然而,除不断更新的可学习模块外,其基本都沿用了同一套潜在的多尺度网络框架。目前,已有工作(如文献[65])表明,非参数组件在处理非结构化点云数据时具有独特优势,但相关研究非常少,因此未来工作应继续探索非参数组件的潜力。

3 点云分类算法性能对比

本文列举并分析了各点云分类网络在不同数据集上的性能表现,包括 ModelNet40 数据集和 ScanObjectNN 数据集。如表 2 所示,采用的评价指标为 OA 和 mAcc。

通过对比各个网络的 OA 值可以发现,相对较优的分类网络是 MVCVT,其利用 transformer 捕获不同视图下特征之间的相关性,很好地保证了特征提取的灵活性^[26]。仅次于其的方法是 PointGPT^[69],以及 PointMLP^[35]。PointGPT 成功的关键在于点云数据的序列化处理,PointMLP 则在于没有集成复杂的局部几何提取器,而只是重复堆叠几个

Table 2 Performance comparison of point cloud classification networks on ModelNet40 and ScanObjectNN datasets

表 2 点云分类网络在 ModelNet40 和 ScanObjectNN 数据集上的性能

类别		年份	比较 (%)			
			ModelNet40		ScanObjectNN	
			OA	mAcc	OA	mAcc
基于多视图	MVCNN ^[19]	2015	90.10	78.90	-	-
	GVCNN ^[23]	2018	93.10	79.70	-	-
	RCPCNN ^[22]	2019	93.80	-	-	-
	MVTN ^[25]	2021	92.00	93.80	82.80	-
	MVCVT ^[26]	2023	95.4	-	-	-
	VoxNet ^[27]	2015	85.90	83.00	-	-
	OctNet ^[28]	2016	86.50	-	-	-
基于体素	Kd-Net ^[29]	2017	91.80	88.50	-	-
	3DContextNet ^[30]	2018	90.20	-	-	-
	PointGrid ^[31]	2018	92.00	88.90	-	-
	PointNet ^[7]	2017	89.20	86.20	-	-
基于逐点 MLP	PointNet++ ^[8]	2017	91.90	-	77.90	75.40
	So-Net ^[33]	2018	90.80	87.30	-	-
	PointWeb ^[34]	2019	92.30	89.40	-	-
	PointMLP ^[35]	2022	94.50	91.40	85.40	83.90
	GhostMLP ^[36]	2023	93.70	91.60	88.40	87.20
	PointCNN ^[40]	2018	92.20	88.10	87.90	-
	Geo-CNN ^[41]	2019	93.90	91.60	-	-
基于卷积	A-CNN ^[42]	2019	92.60	90.30	-	-
	SpiderCNN ^[45]	2018	92.40	-	-	-
	RS-CNN ^[47]	2019	93.60	-	-	-
	ScoreNet ^[44]	2021	93.90	-	-	-
基于原始点云	UniRepLKNNet ^[72]	2024	93.20	90.3	-	-
	KCNet ^[49]	2017	91.00	-	-	-
	DGCNN ^[51]	2018	82.20	90.20	86.20	-
基于图	Grid-GCN ^[52]	2020	93.10	91.30	-	-
	AdaptConv ^[11]	2021	93.40	90.70	-	-
	DDGCN ^[55]	2022	92.70	90.40	-	-
	PAT ^[59]	2019	91.70	-	-	-
	GAPNet ^[61]	2021	92.40	89.70	-	-
	PCT ^[10]	2021	93.20	-	-	-
	PTv2 ^[58]	2022	94.20	91.60	-	-
基于注意力	Point-BERT ^[60]	2022	93.80	-	83.07	-
	Point-MAE ^[61]	2023	93.80	-	85.18	-
	Point-Sim ^[66]	2024	93.30	-	86.60	-
	GPSFormer ^[62]	2024	94.20	92.20	95.40	93.80
其他方法 (多态融合+大模型)	PointCLIP ^[67]	2022	94.05	-	-	-
	ReCon ^[71]	2023	94.20	-	90.35	-
	PointGPT ^[69]	2023	94.90	-	93.40	-

MLP 模块。

此外,从整体上看,大多数方法并没有在 ScanObjectNN 数据集上进行实验,或者在 ScanObjectNN 数据集上的性能表现几乎都低于 ModelNet40。这是因为该数据集是从真实世界室内场景扫描获取的,其包含复杂的背景干扰、严重的数据缺失以及几何形变等问题,因而极具挑战性。但这也间接证明,面向现实世界室内真实数据和大规模室外场景数据的点云分类可能是未来亟需探索与研究的方向。

4 总结与展望

随着各种传感设备的普及以及计算机技术水平的不断提高,深度学习背景下的三维点云分类方法发展迅速,取得了很多应用成果。本文介绍了三维点云分类任务常用数据集与评价指标,以及近年来基于深度学习的点云分类方法研究进展,并分析和对比了这些方法在不同数据集上的性能表现。现对当前分类方法存在的问题进行总结和归纳,并对未来的研究方向进行展望。

4.1 问题与挑战

(1)网络结构设计复杂,不够轻量化。一方面,复杂的网络结构需要大量计算资源,包括高性能的GPU、大容量的内存等硬件设备,这对于一些资源有限的场景(如嵌入式设备、移动设备)是很难满足的。另一方面,复杂的网络结构虽然能够更有效地提取点云特征,实现高精度分类,但这也意味着网络在训练和推理过程中需要花费大量计算时间,在实际生产应用中是不适用的,即便是在高性能的计算设备上也无法达到实时应用的要求。

(2)点云数据存在噪点、离群点问题。受采集设备、环境、遮挡物等因素影响,所获取到的点云数据往往存在稀疏、噪声点、离群点以及自遮挡现象,从而干扰点云的特征提取,降低分类精度和模型的鲁棒性。因此,如何有效地去除噪声和离群点,同时保留点云的重要特征是一项难题。

(3)点云数据不平衡。在实际应用中,点云数据的不同类别之间通常存在严重的不平衡问题。例如在城市点云中,建筑物和道路的点云数量通常多于树木的点云数量。这种不平衡问题会导致分类算法对少数类别的分类准确率较低,从而影响整个点云分类任务的性能。

(3)复杂真实场景。现实真实场景下扫描得到的点云数据通常是大规模的,可能包括道路、建筑、植被等多种复杂的目标对象,对这些数据进行预处理和分类非常耗时,且对硬件要求较高。因此,大规模场景下的点云分类任务是一个具有挑战性的问题。

4.2 未来展望

(1)模型轻量化。当前大多数方法都旨在提高模型分类精度,往往会设计较复杂的网络结构,其参数多、运算量大,缺乏一定的灵活性和自适应性,难以应对实时性强、更复杂的场景。因此,未来应注重网络架构的简化,降低计算复杂度和内存占用率,并开发应用软件和硬件加速技术,以实现3D点云任务在精度与效率上的最大平衡。

(2)多模态融合与大模型应用。从近两年提出的方法看,多模态融合和点云大模型将是未来的研究趋势。由于点云本身往往是无序、无纹理以及稀疏的存在,基于单模态的点云分析正逐渐走向瓶颈,许多研究开始探索多模态表征学习,如PointCLIP^[67]。同时,随着GPT大模型的快速

发展,点云与GPT的结合也成为发展趋势,如PointGPT等^[69]。因此,在未来的研究中应继续探索其他模态数据的融合对点云分类的影响。

(3)数据增强方法应用。当前面向三维点云的深度学习分类方法在网络架构设计方面已进行了广泛研究,而针对网络架构中数据处理部分(数据增强)的研究还处于初步阶段。以往方法大都参考并复用经典方法中的简单数据增强方法,如旋转变换、高度附加、随机抖动等,并未作进一步探索,但实际上多种数据增强方法的协同组合能够很好地提高模型的泛化能力和鲁棒性。目前在目标检测领域已有一些先进的增强技术,如Real3D-AUG通过向原始数据中插入新物体来模拟遮挡数据,AdaMix通过重采样和数据混合策略以提高数据的多样性^[73]。未来工作可尝试将这些三维技术应用于点云分类中。

(4)现实真实场景下的点云数据获取。针对大规模复杂场景问题,未来应探索更高效的数据压缩技术。同时,由于扫描视角限制和遮挡问题,获取到的点云可能存在缺失和不完整现象,或者包含噪声和离群点,从而干扰点云分类的准确性。因此,在未来研究中需要对数据补全技术和去噪算法进行更深入地探索。

参考文献:

- [1] DAS S, PATIBANDLA H, BHATTACHARYA S, et al. TMCOS: thresholded multi-criteria online subset selection for data-efficient autonomous driving[C]// International Conference on Computer Vision, 2021: 6341-6350.
- [2] SUN H X, LUO J X, ZHANG Y Y, et al. A review of typical techniques of image-based 3D reconstruction[J]. Software Guide, 2023, 22(9): 9-24.
孙海迅, 罗健欣, 张艳艳, 等. 基于图像的三维重建典型技术综述[J]. 软件导刊, 2023, 22(9): 9-24.
- [3] LI M J, YU Z K, LIU X, et al. Progress of point cloud algorithm in medical field[J]. Journal of Image and Graphics, 2020, 25(10): 2013-2023.
李美佳, 于泽宽, 刘晓, 等. 点云算法在医学领域的研究进展[J]. 中国图象图形学报, 2020, 25(10): 2013-2023.
- [4] LEE D, LEE J, LEE J, et al. Regularization strategy for point cloud via rigidly mixed sample[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 15900-15909.
- [5] WANG Z Q, CHEN T Y, WANG S F, et al. Research on road cone recognition based on improved DBSCAN clustering algorithm and support vector machines classification[J]. Laser & Infrared, 2022, 52(12): 1796-1803.
王兆权, 陈天炎, 王水发, 等. DBSCAN聚类改进算法与支持向量机结合的道路路锥识别研究[J]. 激光与红外, 2022, 52(12): 1796-1803.
- [6] WU D, YAN W D, WANG J L. Random forest point cloud classification algorithm based on feature importance weighting[J]. Electronic Measurement Technology, 2023, 46(20): 120-127.
吴冬, 阎卫东, 王井利. 基于特征重要性加权的随机森林点云分类研究[J]. 电子测量技术, 2023, 46(20): 120-127.
- [7] QI C R, SU H, MO K, et al. PointNet: deep learning on point sets for 3D classification and segmentation[C]//Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, 2017: 652-660.

- [8] QI C R, YI L, SU H, et al. PointNet++: deep hierarchical feature learning on point sets in a metric space [C]//Proceedings of the 31st Neural Information Processing Systems, 2017: 5105–5114.
- [9] QIAN G C, LI Y C, PENG H W, et al. PointNetXt: revisiting PointNet++ with improved training and scaling strategies [C]//Proceedings of Neural Information Processing Systems, 2022: 23192–23204.
- [10] GUO M H, CAI J X, LIU Z N, et al. PCT: point cloud transformer [DB/OL]. <https://arxiv.org/abs/2012.09688>.
- [11] ZHOU H R, FENG Y D, FANG M S, et al. Adaptive graph convolution for point cloud analysis [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 4965–4974.
- [12] RANI A, ARROYO D O, DURDEVIC P. Advancements in point cloud-based 3D defect detection and classification for industrial systems: a comprehensive survey [DB/OL]. <https://arxiv.org/abs/2402.12923>.
- [13] ZHANG H, WANG C S, TIAN S W, et al. Deep learning-based 3D point cloud classification: a systematic survey and outlook [J]. Displays, 2023, 79: 102456.
- [14] LI M W, ZHAO Y M, LI S B, et al. A survey of point cloud classification [J]. Video Engineering, 2022, 46(5): 1–8.
利满雯, 赵艳明, 李绍彬, 等. 点云分类方法综述 [J]. 电视技术, 2022, 46(5): 1–8.
- [15] WEI T Q, ZHENG X S. Research of deep learning-based classification methods for 3D point cloud [J]. Application Research of Computers, 2022, 39(5): 1289–1296.
魏天琪, 郑雄胜. 基于深度学习的三维点云分类方法研究 [J]. 计算机应用研究, 2022, 39(5): 1289–1296.
- [16] WU Z R, SONG S R, KHOSLA A, et al. 3D ShapeNets: a deep representation for volumetric shapes [C]//Proceedings of the 28th IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1912–1920.
- [17] CHANG A X, FUNKHOUSER T, GUIBAS L, et al. ShapeNet: an information-rich 3D model repository [DB/OL]. <https://arxiv.org/abs/1512.03012>.
- [18] UY M A, PHAM Q H, HUA B S, et al. Revisiting point cloud classification: a new benchmark dataset and classification model on real-world data [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 1588–1597.
- [19] SU H, MAJI S, KALOGERAKIS E, et al. Multi-view convolutional neural networks for 3D shape recognition [C]//Proceedings of the IEEE International Conference on Computer Vision, 2015: 945–953.
- [20] BAI S, BAI X, ZHOU Z, et al. GIFT: a real-time and scalable 3D shape search engine [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 5023–5032.
- [21] ZENG H, WANG Q, LI C, et al. Learning-based multiple pooling fusion in multi-view convolutional neural network for 3D model classification and retrieval [J]. Journal of Information Processing Systems, 2019, 15(5): 1179–1191.
- [22] WANG C, PELILLO M, SIDDIQI K. Dominant set clustering and pooling for multi-view 3D object recognition [DB/OL]. <https://arxiv.org/abs/1906.01592>.
- [23] FENG Y F, ZHANG Z H, ZHAO X B, et al. GVCNN: group-view convolutional neural networks for 3D shape recognition [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 264–272.
- [24] YU T, MENG J J, YUAN J S. Multi-view harmonized bilinear network for 3D object recognition [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 186–194.
- [25] HAMDI A, GIANCOLA S, GHANEM B. MVTN: multi-view transformation network for 3D shape recognition [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 1–11.
- [26] LI J, LIU Z, LI L, et al. Multi-view convolutional vision transformer for 3D object recognition [J]. Journal of Visual Communication and Image Representation, 2023, 95: 103906.
- [27] MATURANA D, SCHERER S. VoxNet: a 3D convolutional neural network for real-time object recognition [C]//Proceedings of the 28th IEEE International Conference on Intelligent Robots and Systems, 2015: 922–928.
- [28] RIEGLER G, OSMAN U A, GEIGER A. OctNet: learning deep 3D representations at high resolutions [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 3577–3586.
- [29] KLOKOV R, LEMPITSKY V. Escape from cells: deep Kd-Networks for the recognition of 3D point cloud models [C]//Proceedings of the 16th IEEE International Conference on Computer Vision, 2017: 863–872.
- [30] ZENG W, GEVERS T. 3DContextNet: K-d tree guided hierarchical learning of point clouds using local and global contextual cues [C]//Proceedings of the European Conference on Computer Vision (ECCV) Workshops, 2018: 314–330.
- [31] LE T, DUAN Y. PointGrid: a deep network for 3D shape understanding [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 9204–9214.
- [32] MENG H Y, GAO L, LAI Y K, et al. VV-Net: voxel VAE net with group convolutions for point cloud segmentation [C]//Proceedings of the 17th IEEE International Conference on Computer Vision, 2019: 8500–8508.
- [33] LI J X, CHEN B M, LEE G H. So-Net: self-organizing network for point cloud analysis [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 9397–9406.
- [34] ZHAO H S, JIANG L, FU C W, et al. PointWeb: enhancing local neighborhood features for point cloud processing [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 5565–5573.
- [35] MA X, QIN C, YOU H X, et al. Rethinking network design and local geometry in point cloud: a simple residual MLP framework [DB/OL]. <https://arxiv.org/abs/2202.07123>.
- [36] LAI H, LAW K E. Efficient point cloud object classifications with Ghost-MLP [J]. Remote Sensing, 2023, 15: 2254.
- [37] YANG G X, ZHOU H, LIU H, et al. Lightweight network for point cloud classification based on gridding and surface features encoder [J]. Software Guide, 2024, 23(5): 9–16.
杨官学, 周昊, 刘慧, 等. 融合栅格与表面特征编码的轻量级点云分类网络 [J]. 软件导刊, 2024, 23(5): 9–16.
- [38] HAN B, ZHANG X Y, REN S. Survey of convolution operations based on 3D point clouds [J]. Journal of Computer Research and Development, 2023, 60(4): 873–902.
韩冰, 张鑫云, 任爽. 基于三维点云的卷积运算综述 [J]. 计算机研究与发展, 2023, 60(4): 873–902.
- [39] HUA B S, TRAN M K, YEUNG S K. Pointwise convolutional neural networks [C]//Proceedings of the 31st IEEE Conference on Computer Vision and Pattern Recognition, 2018: 984–993.
- [40] LI Y Y, BU R, SUN M C, et al. PointCNN: convolution on χ -Transformed points [C]//Proceedings of the 32nd Neural Information Processing Systems, 2018: 828–838.
- [41] LAN S Y, YU R C, YU G, et al. Modeling local geometric structure of 3D point clouds using Geo-CNN [C]//Proceedings of the 32nd IEEE Confer-

- ence on Computer Vision and Pattern Recognition, 2019:998–1008.
- [42] KOMARICHEV A, ZHONG Z C, HUA J. A-CNN: annularly convolutional neural networks on point clouds [C]//Proceedings of the 32nd IEEE Conference on Computer Vision and Pattern Recognition, 2019: 7413–7422.
- [43] THOMAS H, QI C R, DESCHAUD J E, et al. KPConv: flexible and deformable convolution for point clouds [C]//Proceedings of the 17th IEEE Int Conference on Computer Vision, 2019: 6410–6419.
- [44] XU M T, DING R Y, ZHAO H S, et al. PAConv: position adaptive convolution with dynamic kernel assembling on point clouds [C]//Proceedings of the 34th IEEE Conference on Computer Vision and Pattern Recognition, 2021:3173–3182.
- [45] XU Y F, FAN T Q, XU M Y, et al. SpiderCNN: deep learning on point sets with parameterized convolutional filters [C]//Proceedings of the 16th European Conference on Computer Vision, 2018:90–105.
- [46] WU W X, QI Z G, LI F X. PointConv: deep convolutional networks on 3D point clouds [C]//Proceedings of the 32nd IEEE Conference on Computer Vision and Pattern Recognition, 2019: 9613–9622.
- [47] LIU Y C, FAN B, XIANG S M, et al. Relation-shape convolutional neural network for point cloud analysis [C]//Proceedings of the 32nd IEEE Conference on Computer Vision and Pattern Recognition, 2019: 8887–8896.
- [48] SIMONOVSKY M, KOMODAKIS N. Dynamic edge-conditioned filters in convolutional neural networks on graphs [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 3693–3702.
- [49] SHEN Y R, FENG C, YANG Y Q, et al. Mining point cloud local structures by kernel correlation and graph pooling [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4548–4557.
- [50] WANG C, SAMARI B, SIDDIQI K. Local spectral graph convolution for point set feature learning [C]//Proceedings of the European Conference on Computer Vision, 2018: 52–66.
- [51] 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.
- [52] XU Q G, SUN X D, WU C Y, et al. Grid-GCN for fast and scalable point cloud learning [C]//Proceedings of the 33rd IEEE Conference on Computer Vision and Pattern Recognition, 2020: 5660–5669.
- [53] NEZHADARYA E, TAGHAVI E, RAZANI R, et al. Adaptive hierarchical down-sampling for point cloud classification [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020:12956–12964.
- [54] LIN Z H, HUANG S Y, WANG Y C. Convolution in the cloud: learning deformable kernels in 3D graph convolution networks for point cloud analysis [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020:1800–1809.
- [55] CHEN L F, ZHANG Q. DDGCN: graph convolution network based on direction and distance for point cloud learning [J]. The Visual Computer, 2023, 39(3): 863–873.
- [56] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]//Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017:6000–6010.
- [57] ENGEL N, BELAGIANNIS V, DIETMAYER K. Point transformer [J]. IEEE Access, 2021, 9:134826–134840.
- [58] WU X Y, LAO Y X, JIANG L, et al. Point Transformer V2: grouped vector attention and partition-based pooling [C]//Proceedings of Neural Information Processing Systems, 2022: 33330–33342.
- [59] WU X Y, LAO Y X, WANG P S, et al. Point Transformer V3: simpler faster stronger [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024:4840–4851.
- [60] YU X M, TANG L L, RAO Y M, et al. Point-BERT: pre-training 3D point cloud transformers with masked point modeling [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 19313–19322.
- [61] PANG Y T, WANG W X, TAY F E, et al. Masked autoencoders for point cloud self-supervised learning [C]//Proceedings of the European Conference on Computer Vision, 2022: 604–621.
- [62] WANG C S, WU M Q, LAM S K, et al. GPSFormer: a global perception and local structure fitting-based transformer for point cloud understanding [C]//Proceedings of the European Conference on Computer Vision, 2024: 75–92.
- [63] LIU Y H, TIAN B, LV Y S, et al. Point cloud classification using content-based transformer via clustering in feature space [J]. IEEE/CAA Journal of Automatica Sinica, 2023, 11(1): 231–239.
- [64] BERG A, OSKARSSON M, O’CONNOR M. Points to patches: enabling the use of self-attention for 3D shape recognition [C]//Proceedings of the 26th International Conference on Pattern Recognition, 2022: 528–534.
- [65] ZHANG R R, WANG L H, GUO Z Y, et al. Parameter is not all you need: starting from non-parametric networks for 3D point cloud analysis [DB/OL]. <https://arxiv.org/abs/2303.08134>.
- [66] GUO J C, LUO W J. Point-Sim: a lightweight network for 3D point cloud classification [J]. Algorithms, 2024, 17(4):158.
- [67] ZHANG R R, GUO Z Y, ZHANG W, et al. PointCLIP: point cloud understanding by clip [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 8552–8562.
- [68] HUANG T Y, DONG B W, YANG Y H, et al. Clip2Point: transfer clip to point cloud classification with image-depth pre-training [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023:22157–22167.
- [69] CHEN G Y, WANG M L, YANG Y, et al. PointGPT: auto-regressively generative pre-training from point clouds [C]//Proceedings of Neural Information Processing Systems, 2024:29667–29679.
- [70] QI Z Y, FANG Y, SUN Z Y, et al. GPT4Point: a unified framework for point-language understanding and generation [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024:26417–26427.
- [71] QI Z K, DONG R P, FAN G F, et al. Contrast with reconstruct: contrastive 3D representation learning guided by generative pretraining [C]//Proceedings of International Conference on Machine Learning, 2023: 28223–28243.
- [72] DING X H, ZHANG Y Y, GE Y X, et al. UniRePLKnet: a universal perception large-kernel ConvNet for audio video point cloud time-series and image recognition [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024:5513–5524.
- [73] ZHAI R F. Research on data augmentation and autoencoder optimization in 3D point cloud perception [D]. Changchun: Jilin University, 2024. 翟睿峰. 三维点云感知中的数据增强与自编码器优化研究 [D]. 长春: 吉林大学, 2024.