中图法分类号:TP183　　**文献标识码**:A　　**文章编号**:1006-8961(2025)07-2325-18

论文引用格式:Qiu Z J, Zhang L, Yao Y, Feng X Q and Gao J T. 2025. Survey on semantic segmentation in 3D point cloud scenes. Journal of Image and Graphics, 30(7):2325-2342(仇志江, 张林, 姚垚, 冯小青, 高俊涛. 2025. 三维点云场景语义分割研究进展. 中国图象图形学报, 30(7):2325-2342)[DOI:10.11834/jig.240650]

三维点云场景语义分割研究进展

仇志江[1],张林[2],姚垚[2],冯小青[2*],高俊涛[1]

1. 东北石油大学计算机与信息技术学院,大庆　163318; 2. 浙江财经大学信息技术与人工智能学院,杭州　310018

摘　要:随着深度传感器技术的迅速发展和激光扫描仪的广泛应用,三维点云数据在自动驾驶、机器人技术、地理信息系统以及制造业等多个领域发挥着关键作用。本文从多个视角出发,对点云场景分割技术进行详尽的综述。1)介绍本领域的常用数据集、采样方法及点云特征提取技术;2)系统性地概述点云场景的语义分割研究方法,涵盖基于点的方法,分为基于MLP(multilayer perceptron)的方法、基于Transformer的方法、基于图卷积的方法和其他方法;基于体素的方法,基于视图的方法以及基于多模态融合方法。3)对上述各种方法的研究成果进行汇总与比较;4)指出当前面临的主要挑战,并探讨未来可能的研究方向。

关键词:三维点云;点云场景语义分割;深度学习;采样方法;特征提取

Survey on semantic segmentation in 3D point cloud scenes

Qiu Zhijiang[1], Zhang Lin[2], Yao Yao[2], Feng Xiaoqing[2*], Gao Juntao[1]

1. School of Computer and Information Technology, Northeast Petroleum University, Daqing　163318, China;

2. School of Information Technology and Artificial Intelligence, Zhejiang University of Finance and Economics, Hangzhou　310018, China

Abstract: With the continuous advancement of depth sensing technology, particularly the widespread application of laser scanners across diverse scenarios, 3D point cloud technology is playing a pivotal role in an increasing number of fields. These fields include, but are not limited to, autonomous driving, robotics, geographic information systems, manufacturing, building information modeling, cultural heritage preservation, virtual reality, and augmented reality. As a high-precision, dense, and detailed form of spatial data representation, 3D point clouds can accurately capture various types of information, including the geometric shape, spatial structure, surface texture, and environmental layout of objects. Consequently, the processing, analysis, and comprehension of point cloud data are particularly significant in these applications, particularly point cloud segmentation technology, which serves as the foundation for advanced tasks, such as object recognition, scene understanding, map construction, and dynamic environment monitoring. This study aims to conduct a comprehensive review and in-depth exploration of current mainstream 3D point cloud segmentation methods from multiple perspectives, dissecting the latest advancements in this research domain. In particular, it begins with a detailed analysis and discussion of the fundamentals of 3D point cloud segmentation, covering aspects such as datasets, performance evaluation metrics, sampling methods, and feature extraction techniques. This study summarizes currently publicly available mainstream point cloud datasets, including ShapeNet, Semantic3D, S3DIS, and ScanNet, meticulously dissecting the characteristics, annotation forms, application scenarios, and technical challenges associated with each dataset. In addition, it

收稿日期:2024-11-07;**修回日期**:2024-12-18;**预印本日期**:2024-12-25

*通信作者:冯小青　fenglinda@zufe.edu.cn

基金项目:教育部人文社会科学研究一般项目资助(20YJAZH028)

Supported by: Ministry of Education of Humanities and Social Science Project of China(20YJAZH028)

delves into commonly utilized performance evaluation metrics in the semantic segmentation of point cloud scenes, including overall accuracy, mean class accuracy, and mean intersection over union. These metrics provide effective means for quantifying and comparing model performance, facilitating comprehensive evaluations and improvements across different tasks and scenarios. In the data preprocessing phase, this study systematically summarizes prevalent point cloud sampling strategies. Given that 3D point cloud data typically possess large scales and irregular distributions, a suitable sampling method is essential for reducing computational costs and enhancing model training efficiency. This study introduces strategies, such as farthest point sampling, random sampling, and grid sampling, analyzing their application scenarios, advantages, disadvantages, and specific implementation methods in various tasks. Furthermore, it discusses feature extraction techniques for 3D point clouds, encompassing various methods, including global feature extraction, local feature extraction, and the fusion of global and local features. Through effective feature extraction, more discriminative representations can be provided for subsequent segmentation tasks, aiding the model in improved object recognition and scene understanding. Building on this foundation, this study systematically reviews 3D point cloud segmentation methods from four distinct perspectives: point-based, voxel-based, view-based, and multi-modal fusion methods. First, point-based methods directly process each point within the point cloud, maintaining the high resolution and density of the data while avoiding information loss. Point-based methods are further subdivided into multilayer perceptron (MLP)-based methods, Transformer-based models, graph convolutional network-based methods, and other related approaches. Each category exhibits unique advantages in different application scenarios. For example, MLP is effective for capturing the local features of point clouds, while Transformer-based models excel in handling long-range dependencies and global relationships. Despite the strong performance of point-based methods, their direct processing of a large number of 3D points results in high computational complexity and relatively low efficiency when managing large-scale point cloud scenes. Second, this study presents voxel-based methods, which process point cloud data by partitioning them into regular 3D grids (voxels), effectively reducing data size and simplifying subsequent computations. This approach structures point cloud data, providing relatively stable performance in large-scale scenes. It is particularly applicable to scenarios with large scene sizes but low resolution requirements. However, the inevitable information loss and reduction in spatial resolution during voxelization limit performance in handling fine-grained tasks. Third, the view-based approach processes a 3D point cloud by projecting it onto a 2D plane, leveraging mature 2D image processing techniques and convolutional neural networks. This method transforms the point cloud segmentation task into a traditional image segmentation problem, enhancing processing efficiency, particularly in scenarios where point cloud density is high and rapid processing is required. However, projecting 3D information onto 2D space may lead to the loss of spatial geometric information, resulting in potentially lower accuracy compared with methods that directly handle 3D point clouds in certain applications. Lastly, this study explores multimodal fusion methods, which combine various forms of data, such as point clouds, voxels, and views, fully utilizing the complementarity of different modalities in scene understanding to enhance the accuracy, robustness, and generalization ability of point cloud segmentation. Subsequently, this study conducts a detailed analysis and comparison of the experimental results from different methods. Based on various datasets and performance evaluation metrics, it reveals the strengths and weaknesses of each method in diverse application scenarios. For example, point-based methods excel in fine segmentation tasks and can capture subtle geometric information, while voxel-based and view-based methods offer higher processing efficiency when dealing with large-scale point cloud scenes. Through the comparative analysis of experimental results, this study provides valuable references for point cloud segmentation tasks across different application scenarios. Finally, this study summarizes the major challenges that are currently experienced in the field of 3D point cloud segmentation, including the sparsity and irregularity of point cloud data, the influence of noise and missing points, insufficient generalization ability across diverse scenes, and the demand for real-time processing. This study also anticipates future research directions and proposes measures, such as a deeper understanding of the complex semantic structures of point cloud data through the integration of large language models, the introduction of semi-supervised and unsupervised learning methods to reduce reliance on labeled data, and enhancements in real-time performance and computational efficiency to further advance point cloud segmentation technology. We hope that this comprehensive review can provide a systematic reference for researchers and industrial applications in the field of point cloud technology, facilitating the implementation and development of 3D point cloud technology in a

broader range of practical applications.

Key words: 3D point cloud; point cloud scene semantic segmentation; deep learning; sampling method; feature extraction

0 引　言

随着人工智能和计算机视觉技术的飞速进步，三维点云作为真实世界环境的高精度数字化表示，广泛应用于无人驾驶、无人机导航、虚拟现实、智慧城市和文化遗产保护等多个前沿领域。三维点云由激光扫描、结构光和雷达等多种传感器获取的大量空间坐标数据组成，直接反映物体的几何形态与表面特征。对这些数据进行高效精确的分析与处理，对于实现上述应用中的场景理解、目标识别与重建至关重要。三维点云场景语义分割作为预处理的关键步骤，旨在将复杂的点云数据划分为若干有意义的子集或对象，每个子集对应于场景中的独立物体或表面。此过程对后续的物体识别、追踪及环境建模等任务具有基础性作用。通过有效的分割，可以显著提升数据处理的效率和精度，为高级应用提供更加结构化的信息输入。

近年来，随着深度学习、图神经网络等先进算法的兴起，三维点云场景语义分割技术经历了从基于手工特征的方法到数据驱动方法的转变，取得了显著的进展。各种新颖的网络架构和优化策略不断涌现，推动了该领域的快速发展。然而，这些技术的适用范围、性能限制及实际应用中的挑战仍需深入探讨和总结。

本文系统梳理与分析近几年国内外的点云分割方法，深入探讨一些经典算法并进行对比分析，提出目前该领域面临的挑战，并讨论未来的研究方向。

1 背　景

1.1 任务与挑战

点云是由一系列点组成的具有显著几何数据表达结构的点集，但点云的无序性及非结构化特性，给基于点云的场景语义分割带来了诸多挑战。

三维点云（3D point cloud）是一种用于表示三维空间内物体或环境的数据形式，可以描述成一组有序三元组。如一个包含 n 个点的点云，可以表述为一个点集 $\boldsymbol{V}=\{v_1, v_2, \cdots, v_n\}$，其中，$\boldsymbol{v}_i=(x_i, y_i, z_i)$ 是第 i 点，x_i，y_i，z_i 分别是该点在笛卡儿坐标系中的 X、Y 和 Z 坐标值。除了位置信息外，一些点云还可能包含其他信息，如颜色、强度（如果是从激光雷达获得）、法线向量等。若每个点包括颜色信息，则每个点可以表示为一个 4 维向量 $\boldsymbol{v}_i=(x_i, y_i, z_i, c_i)$，其中 c_i 表示该点的颜色。

场景语义分割（scene semantic segmentation）是指将图像或三维点云中的每个像素或点分配给一个特定的类别标签的过程。该过程不仅关注于识别出图像或点云中的对象，而且还进一步细化到理解每个像素或点属于哪个具体的类别。场景语义分割的目标是使机器能够理解场景中的各个组成部分，并赋予它们语义意义。给定 $\boldsymbol{e}\in\mathbf{R}^3$，其中 $\boldsymbol{e}=\{e_1, e_2, e_3, \cdots, e_n\}$ 作为从输入的三维点云 $\boldsymbol{V}$ 中提取的一组点，场景语义分割是通过神经网络 N 将一个标签 l_i 从一组可能的语义标签集合 $\boldsymbol{L}=\{l_1, l_2, l_3, \cdots, l_m\}$ 分配给点云 $\boldsymbol{V}$ 中的每一个点 e_i 的过程（Adam 等，2023）。

三维点云场景语义分割的核心任务是为点云中的每一个点添加语义标签，以此识别并区分场景内的不同物体或区域，例如建筑物、道路、树木、行人和车辆等。任务的关键在于准确识别不同类别的点，尤其是处理边界区域和类别过渡区域。此外，语义分割还需要帮助理解整个场景的结构，通过提取多尺度特征，分割出不同的语义区域，同时应具备处理遮挡和噪声的能力。在自动驾驶和机器人导航等实时性要求较高的应用中，算法还需在有限的计算资源下快速运行，以确保动态环境中的精确感知。三维点云场景语义分割任务框架图如图 1 所示。

深度学习在自然语言处理和二维图像处理中的广泛应用推动了对三维点云数据处理的探索。然而，点云数据由离散的三维坐标点构成，缺乏固定的拓扑结构，这使得传统的卷积操作和序列模型在处理此类数据时面临重大挑战。为了有效捕捉点云中的几何特征和空间关系，构建高精度与高效率的点云三维场景分割策略，研究者提出多种基于点、体素、视图和多模态融合的点云场景语义分割方法。

1.2 常用数据集

标准的数据测试集和性能评价指标为评估和比

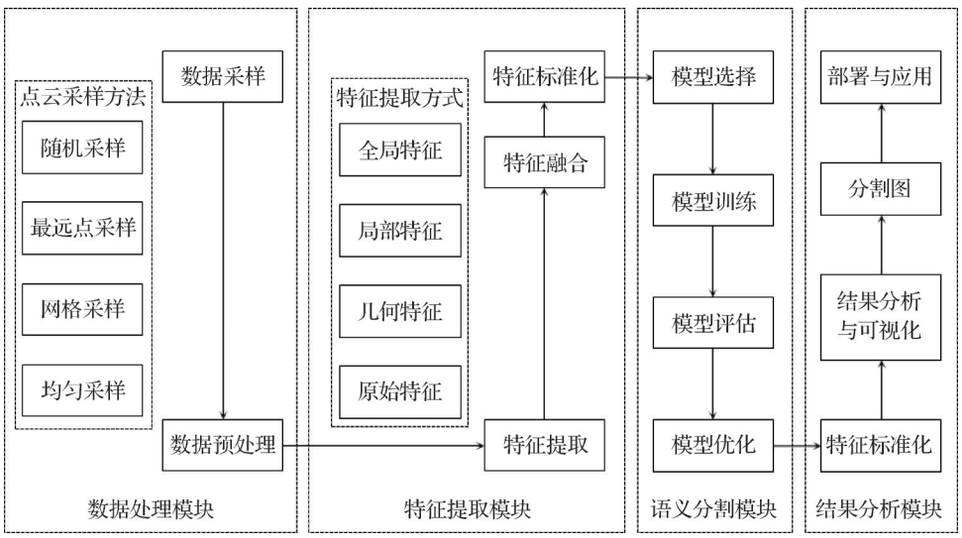


图1　三维点云场景语义分割框架图

Fig. 1　A framework for semantic segmentation of 3D point cloud scenes

较不同算法的有效性提供了基础，确保了研究成果的可重复性和可验证性。

ShapeNet（Chang等，2015）是一个广泛使用的、高质量的三维形状数据集，专为学术研究和计算机视觉任务设计。该数据集包含大量经过标注的三维模型，这些模型以CAD（computer aided design）格式呈现，涵盖了多种日常物体类别。

S3DIS（Stanford large-scale 3D indoor spaces dataset）（Armeni等，2016）包含6个不同建筑物内部的广泛场景，如办公室、会议室和走廊等，总计涵盖272个房间的数据，覆盖23 000 m^2的地面面积，包含超过5亿个点。数据集中每个点都包含了空间坐标、颜色和标签信息，并标注了13个详细的语言标签，如墙、地板、门和椅子等。

ScanNet（Dai等，2017）是一个由RGB-D传感器采集的大规模室内场景理解数据集，包含超过1 500个扫描场景，并为部分场景提供了详细注释，包括语义标签和实例标签等。

Semantic3D（Hackel等，2017）是一个大规模的室外三维点云数据集，包含数亿个带标签的点云数据，涵盖广泛的室外场景，如城市、森林、大学校园等。数据集中的点云被细致地标记了8个主要类别，包括人工结构、自然元素等。

SemanticKITTI（Behley等，2019）是一个面向自动驾驶和机器人技术的大型三维点云数据集，包含超过22 000帧的三维激光扫描数据，覆盖了城市和乡村环境。每个点云帧都经过详细的语义标注，涵盖19个不同的对象类别，如车辆、行人、建筑物、树木、地面等。该数据集以序列形式提供，不仅包含静态的三维点云，还包括车辆行驶过程中的连续帧。

ScribbleKITT（Unal等，2022）是首个为LiDAR（light detection and ranging）点云语义分割设计的手绘线标注数据集。该数据集基于SemanticKITTI创建，为训练集提供了1.89亿个标注点，但这些标注点仅占总点数的8.06%。通过使用线性手绘而非自由形式的涂鸦来注解LiDAR点云，显著加速标注过程，同时有效保留了几何结构清晰的类别关键信息，如：道路、建筑物和篱笆等。

2　点云数据采样

在处理点云数据时，采样是一个重要步骤，它可以有效降低算法所需的点云数据量，提高算法效率，同时保持点云的关键特征。常见的点云采样方法有最远点采样、随机采样和网格采样等。表1展示了多种点云采样方法的对比。

最远点采样（farthest point sampling，FPS）（Qi等，2017）是一种高效的点云数据处理算法，旨在从较大的点集中选择出一组具有代表性的点。这些被选取的点在空间中具有尽可能大的距离，以确保其在几何分布上具有良好的覆盖性。通过这一策略，FPS能够有效捕捉点云的结构特征，进而在保持信息完整性的同时，减少点云数据的规模。PointNet++（Qi等，2017）利用最远点采样选择代表性的点，从

表 1　多种点云采样方法的对比

Table 1　Comparison of point cloud sampling methods

采样方法	计算复杂度	特点	适用场景
最远点采样	O(N^2)	空间上均匀分布，有效保留点云数据结构。	对采样精度有严格要求且需要点云覆盖均匀。
随机采样	O(1)	实现简单，速度快，计算成本低。	大规模数据集且对点分布的均匀性要求不高。
网格采样	O(N)	在保持一定均匀性的同时实现较高的处理速度。	寻求平衡处理效率与分布均匀。

而确保从不同区域中选取的点具有最大的空间分散性。Point-BERT（Yu 等，2022）通过使用最远点采样获取中心点，再利用 k 近邻为中心点选择最近的邻居点，使得点云可以分割成多个局部区域，以便后续处理与特征提取。Yang 等人（2023）在进行类增量之前使用最远点采样从输入的点云中采样固定数量的关键点，在几何感知蒸馏模块中，FPS 采样的点用于提取点云的几何特征；在不确定性感知伪标签生成模块中，FPS 采样的点用于估计点云的不确定性。

随机采样（random sampling）（Hu 等，2020）是处理点云数据时一种基本且直观的采样方法，其直接从原始点云中随机抽取一定数量的点作为样本集。RandLA-Net（Hu 等，2020）在处理大规模的点云数据，采用随机采样策略降低计算复杂度，调高分割效率。RangeFormer（Kong 等，2023）通过随机采样平衡不同类别的点数，增加数据多样性。

网格采样（grid sampling）（Thomas 等，2019）是一种从点云数据中选取代表性点的技术，其通过将点云数据分割成若干个网格单元，并从每个单元中选择一个代表点来实现，这些代表点可以是单元内的中心点或是所有点的质心。KPConv（kernel point convolution）（Thomas 等，2019）中为了确保点采样位置的空间一致性，使用网格子采样，使得每一层的支持点（即携带特征位置的点）是通过选择所有非空网格单元内原始输入点的质心实现的。Point Cloud Mamba（Zhang 等，2024）通过将 3D 点云数据映射到 2D 网格，保留了点云的局部结构信息，使点云数据在网格中均匀分布，降低数据维度。

最远点采样方法能够确保采样后的点云在空间上的均匀分布，并且有效保留点云数据的主要结构特性。然而，该方法的计算复杂度较高，达到 O(N^2)，因此，适用对采样精度有严格要求并且需要保证点云均匀覆盖的场景。相比之下，随机采样因其简便性和较低的计算开销（复杂度为 O(1)）而得到广泛采用；然而，其显著缺点在于可能导致某些区域的点分布过于稠密或过于稀疏，从而影响最终的采样精度。但在需要处理大规模数据集的情况下，随机采样不失为一种有效的选择。网格采样则是通过将三维空间划分为多个单元格，从每个非空单元格中选择一个代表性点完成采样，计算复杂度约为 O(N)。该方法在保持一定均匀性的同时实现了较高的处理速度，但是网格边界的定义可能会影响采样结果。因此，在寻求平衡处理效率与分布均匀性的应用场合中，网格采样是一种合理的选择。

3　点云特征提取

由于三维点云的非结构化数据形式，点云数据处理需要依靠有效的特征提取来理解和分析点云中的信息。特征提取从原始点云数据中提取有用的特征信息，减少计算负担，以便于后续的分类、分割或识别任务。

3.1　全局特征提取

全局特征提取主要用于从整个点云中提取出代表整体形状或结构的特征。这种全局特征提供了对整个数据集的基本描述，有助于理解点云数据的大致形态和分布，且通常具备较高的计算效率。PointNet（Charles 等，2017）在特征提取阶段之后，每个点都拥有一个高维特征向量，通过最大池化操作，选取这些特征向量在各个维度上的最大值，从而形成一个全局特征向量。KPConv（Thomas 等，2019）采用核点卷积在任意密集度的点云上执行卷积操作，随后使用全局池化获取全局特征。PCT（point cloud Transformer）（Guo 等，2021）通过全局池化和平均池化相结合提取全局特征，以支持后续的分类和分割。

3.2　局部特征提取

局部特征提取主要用于提取点云中每个点或点团的特征表示，能够深入捕捉点云的局部几何细节，

在包含复杂形状或多变几何结构的点云中，局部特征能够更有效地处理物体的曲面、边界和尖角等信息，特别适合处理具有复杂拓扑结构的三维对象。与全局特征相比，局部特征更加灵活，能够应对几何形状细节的变化。如：PointNet++（Qi等，2017）采用球形查询机制来组织邻近点进行分组，并利用共享的多层感知机（multilayer perceptron，MLP）编码和最大池化方法捕获局部特征。此外，DGCSA（dynamic graph convolution spatial attention neural network）（宋巍 等，2021）通过训练学习到的空间转换矩阵对输入的点云数据进行坐标对齐，之后利用动态图卷积模块提取边缘特征，并利用空间注意力模块对局部邻域特征进行聚合。CL2M（contrastive learning label mining）（黄华 等，2025）通过引入结合自注意力机制和位置编码，局部特征的提取得到了显著提升，特别是在边界区域的分割精度上。其他研究如PointWeb（Zhao 等，2019）、DensePoint（Liu 等，2019）以及PAConv（Xu等，2021）分别提出不同的机制来优化局部特征提取过程，使得模型在不同场景下更高效地提取到有意义的局部特征。

3.3 全局和局部融合的特征提取

全局特征能够捕捉整个点云的整体结构和形状，而局部特征则能提取细节和局部几何信息。通过将全局特征与局部特征相融合，算法能够同时理解场景的整体布局和局部复杂细节，从而提升在复杂场景中的表现。例如：PVT（point-voxel Transformer）（Zhang等，2022a）通过体素分支和点分支分别聚合局部特征和全局特征，最后将两个分支的输出特征融合。Point-BERT（Yu等，2022）使用局部点嵌入和CLS（class token）令牌整合点云数据的局部几何细节以及全局理解。LGANet（local and global information attention network）（刘玉杰 等，2023）通过通道自注意力获取局部特征，并构建动态学习关键点以捕获全局信息，最终设计实现了基于空间注意力机制的特征融合模块。除此之外，PatchFormer（Zhang 等，2022b）、GrowSP（Zhang 等，2023b）以及LiM3D（less is more 3D）（Li 等，2023）也展示了如何在保留局部细节的同时，有效地整合全局上下文信息，以提升点云处理任务中的性能表现。

图2是三维点云场景语义分割的一些典型算法模型的时序发展图。在2015—2017年，研究主要聚焦于基础算法的开发，探索了从规则的三维体素化表示和二维投影图像到直接处理原始点云的不同方法，但这些方法主要依赖全局特征，并且在不同场景和数据集上的泛化能力较弱。2018—2020年，研究重心转向了提升点间关系信息的利用效率，通过引入图卷积网络等先进技术，有效提取了点云中的局部结构信息，提高了模型的表达能力，但全局上下文信息的捕捉仍显不足，影响了模型在复杂场景中的表现。进入2021年，PCT（Guo 等，2021）和PTv1（point Transformer）（Zhao 等，2021）开创性地将Transformer架构引入点云处理领域，通过自注意力机制有效捕捉全局上下文信息，为解决复杂场景下的语义分割问题提供了新思路。随后的研究将研究重心移至对全局和局部融合的特征提取上。

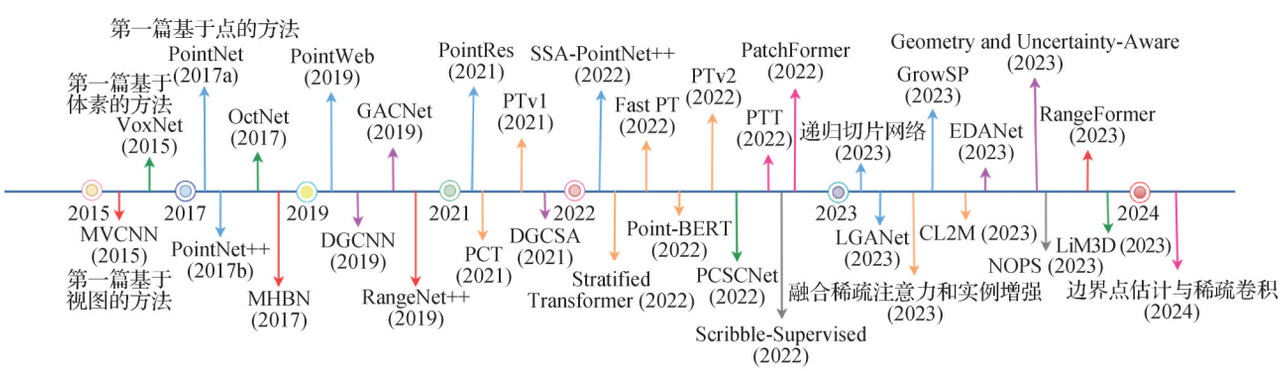


图2 三维点云场景语义分割方法时间顺序图

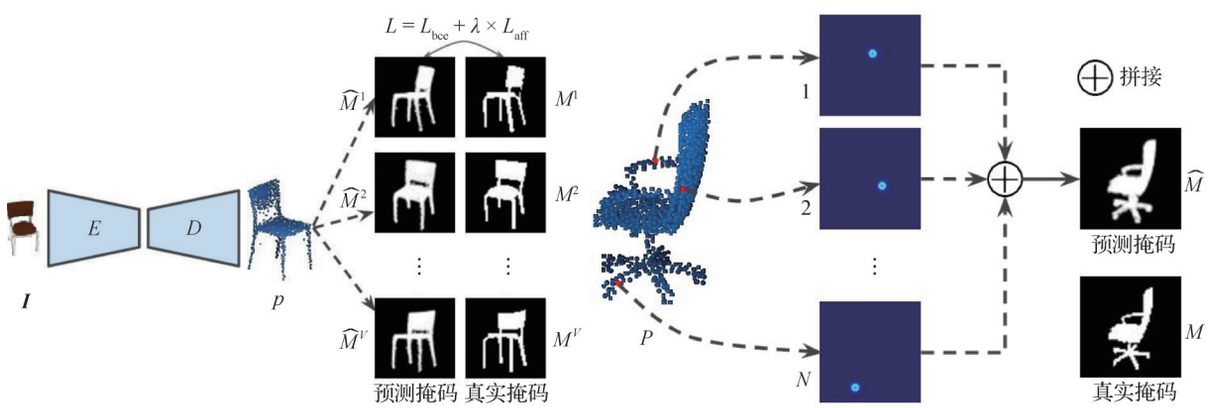
Fig. 2 Chronological diagram of semantic segmentation method for 3D point cloud scene

4　点云场景语义分割方法

点云场景语义分割是三维计算机视觉和机器学习领域的一个重要任务，其目标是从点云数据中识别并分离出不同的对象或区域。点云场景语义分割的方法主要分为基于视图的方法、基于体素的方法、基于点的方法和基于多模态融合的方法。基于视图的方法借助于先进的二维图像处理技术，能够高效地将三维点云映射至二维平面，简化了数据处理流程；基于体素的方法模拟了图像处理中的像素模式，促进了点云数据的有序化表示；由于将非结构化的点云数据转换成结构化的体素网格或视图会导致数据集异常膨胀，因此出现了直接处理点云数据的方法，以减少不必要的数据冗余。

4.1　基于视图的方法

基于视图的方法通过将三维点云投影到多个二维视图上，从而简化数据的复杂性，并利用成熟的二维图像处理技术进行分析和特征提取。这种方法能够有效降低计算复杂度，同时利用已有的图像处理工具提高对三维场景的理解能力。通过多视角的融合，可以更全面地捕捉三维结构的信息，为后续的分析和任务提供更为丰富的特征基础，示例图像如图 3 所示。



(a) 基于视图的点云及掩码预测图　　(b) 基于特征采样的点云掩码生成图

图 3　基于视图的三维物体识别示例图像

Fig. 3　Example image of view based 3D object recognition((a)image of view based point cloud and mask prediction; (b)image of point cloud mask generation based on feature sampling)

Su 等人(2015)提出的多视图卷积神经网络(multi-view convolutional neural network, MVCNN)是多视图方法中的开创性工作，能够从 3D 形状的多个渲染视图中学习并整合信息，形成单一且紧凑的形状描述符。该方法通过引入视图池化层以聚合来自不同视图的特征，然而，这一过程可能会导致信息的丢失。与之不同，Yu 等人(2018)提出的多视图协调双线性网络(multi-view harmonized bilinear network, MHBN)利用双线性池化处理局部卷积特征以有效表示 3D 对象，进而生成紧凑的全局描述符。MHBN 作为一个端到端可训练的框架，将协调双线性池化作为一个网络层，显著提升了模型的学习能力。

Milioto 等人(2019)提出一种基于 LiDAR 的语义分割方法 RangeNet++，该方法对 LiDAR 扫描数据进行球面投影，将点云转化为二维图像表示，类似于范围图像(range image)，从而充分利用旋转 LiDAR 传感器的检测模式，并引入一种有效的基于 GPU 的 K 最近邻(K-nearest neighbor, KNN)后处理方法，以解决直接在图像表示上操作可能引入的离散化误差和模糊的 CNN 输出的问题。Kong 等人(2023)提出一个包含网络架构、数据增强和后处理的网络 RangeFormer，该方法采用 Transformer 架构，将 LiDAR 点云的范围视图(range view)投影作为输入，通过多个连续的 Transformer 块提取多尺度特征，并最终通过多个解码头进行预测。该网络开发了一套轻量级的数据增强组合，直接作用于栅格化的范围视图图像，并提出一种监督方式的解决方案，通过创建“子点云”并为每个子集推断标签，从而直接减少信息损失并缓解“多对一”问题。此外，还设计了一种可扩展的范围视图训练(scalable training from range view, STR)策略，使得在任意低分辨率的二维范围图像上进行训练成为可能，同时仍能保持良好

的三维分割精度。基于视图的点云分割方法在数据集上的性能如表2所示。总体精度(overall accuracy,OA)用于评估分割方法中正确预测的样本数与样本总数之比;平均精度均值(mean accuracy,mACC)用于计算每个类别的精度平均值;平均交并比(mean intersection over union,mIoU),IoU用于衡量模型预测的区域和真实区域交集的面积与它们的并集之比,mIoU是所有类别IoU的平均值。

表2　基于视图的点云分割方法在数据集上的性能效果

Table 2　Performance of view based point cloud segmentation methods on datasets

/%

方法	年份	数据集	OA	mAcc	mIoU
MVCNN	2015	ModelNet40	90.10	–	–
MHBN	2017	ModelNet10	90.00	–	–
		ModelNet40	83.80	–	–
RangeNet++	2019	KITTI	–	–	52.20
RangeFormer	2023	SemanticKITTI	–	–	73.30
		nuScenes	–	–	80.10

注:"–"表示未采用对应指标。

虽然二维图像上的处理和分割技术、数据增强技术已经非常成熟并且在多种任务中表现出良好的效果,提升了模型的鲁棒性和泛化能力,但将三维点云投影到二维平面上会不可避免地丢失一些空间信息,尤其是深度信息。此外,选择适当的视角以最大化信息保留也是一项具有挑战性的任务。

4.2　基于体素的方法

基于视图的方法需要将三维点云转换为二维图像进行处理,基于体素的方法通过将连续的三维空间离散化为一系列小的体积单元(即体素),模仿图像处理中的像素概念,从而实现对点云数据的结构化表示,示例图像如图4所示。

Maturana和Scherer(2015)首次提出利用体素处理点云,引入了一种称为VoxNet的三维卷积神经网络,该方法将原始的点云数据转换为体素占用网格,并在这一结构化表示上执行三维卷积操作,解决了点云的非结构化问题。Riegler等人(2017)提出OctNet(octree network),该方法采用八叉树形式组织点云数据,通过数据密度动态分区3D空间,以适应性地分配内存和计算资源。

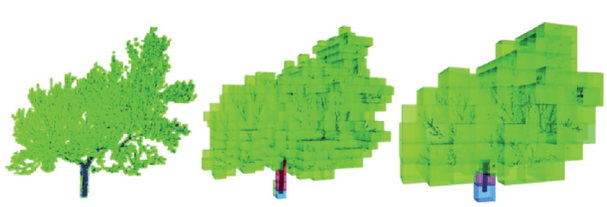


图4　基于体素的三维物体示例图像

Fig. 4　Example image of voxel based 3D object

Park等人(2023)设计了一种结合点卷积和三维稀疏卷积的方法PCSCNet(point convolution and sparse convolution network),该方法通过使用大尺寸体素对点云进行体素化,并利用点卷积提取每个体素的特征。为避免因大尺寸体素引起的离散化误差,该模型通过交叉熵损失和位置感知(position-aware, PA)损失进行优化。在特征提取之后,利用三维稀疏卷积快速扩大感受野,将体素级特征传播到邻近区域,从而加速特征传播过程。同年,LiM3D(Li等,2023)在其半监督三维激光雷达语义分割方法中,将不规则点云转换为规则的三维网格并设计了稀疏深度分离卷积(sparse depthwise separable convolution, SDSC)模块减少可训练参数,同时保持模型的整体表现。此外,使用时空冗余帧降采样(spatio-temporal redundant frame downsampling, ST-RFD)策略,通过分析传感器在环境中的运动信息,提取多样化的训练帧子集,去除时间上的冗余,提出一种基于激光雷达反射率的软伪标签方法,利用激光雷达的反射率反映场景中物体材质的特性,以提高有限标注数据的使用效率。基于体素的点云分割方法在数据集上的性能如表3所示。

表3　基于体素的点云分割方法在数据集上的性能效果

Table 3　Performance of voxel based point cloud segmentation methods on datasets

/%

方法	年份	数据集	OA	mAcc	mIoU
VoxNet	2015	ModelNet10	–	92.00	–
		ModelNet40	85.90	83.00	–
OctNet	2017	ModelNet10	90.00	–	–
		ModelNet40	83.80	–	–
PCSCNet	2022	nuScenes	–	–	72.00
		SemanticKITTI	–	–	62.70
LiM3D	2023	SemanticKITTI	–	–	59.50
		ScribbleKITTI	–	–	58.10

注:"–"表示未采用对应指标。

尽管体素化允许在不同尺度上进行操作，有助于捕捉不同大小的对象特征，并且体素化后的点云数据更容易进行数据增强，但在体素化过程中，每个体素通常只保留一个代表性的值，这可能会导致原始点云中的细节信息丢失。

4.3 基于点的方法

基于视图和基于体素的方法都需要先将点云转换为一些规则结构，该类方法使得卷积神经网络能成功应用于3D点云，但它也会导致点云模型信息的缺失、算法计算复杂高等问题，一些学者逐渐转向直接基于点的方法进行三维场景分割。基于点的方法是直接在无序的点云数据上进行操作，该操作可以避免点云转换到规则结构这一过程对3D模型信息造成影响。目前基于点的方法可以进一步分为基于MLP的方法、基于图卷积的方法、基于Transformer的方法以及其他方法。在MLP方法的初期应用中，主要侧重于提取全局特征，然而这种方法往往伴随着大量细节信息的损失。随后提出的图卷积网络方法有效捕捉了点云数据中的局部结构信息，但其在一定程度上忽视了全局上下文信息的价值。近年来，基于Transformer的模型因其在处理序列数据时展现出的卓越能力，被引入到点云分析领域，强调了全局上下文信息的重要性。这一转变不仅丰富了点云处理的理论框架，也显著提高了算法的准确性和鲁棒性。

4.3.1 基于MLP的方法

基于MLP的方法使用多个共享的多层感知机（MLPs）直接从每个点坐标以及其他属性中提取特征。2017年，Charles等人（2017）提出的PointNet是基于点的方法的开创性研究，这是一种全局特征聚合的方法。该方法能够直接处理无序的三维点云数据，每个点首先通过多层感知机（MLP）单独编码，然后通过最大池化操作整合所有点的特征，以此形成全局描述符，有效解决了点云的排列不变性和旋转不变性问题。尽管PointNet是基于深度学习的先驱，但由于每个点都独立学习特征，因此无法捕获点与点之间的局部结构信息。为了解决这一问题，Qi等人（2017）设计了层次网络PointNet++来捕获点云中的局部结构和几何特征。该网络通过多个阶段（称为集合抽象层（set abstraction layers））对点云进行分组和特征提取，集合抽象层是PointNet++的层次结构，主要由采样层、分组层和PointNet层构成。通过叠加多个集合抽象层级，PointNet++能够学习局部几何结构中的特征，并逐步对这些局部特征进行抽象处理，但多层次的采样和分组策略显著增加了计算负担。吴军等人（2022）在PointNet ++基础上引入自注意力机制构建点云语义分割网络SSA-PointNet++，如图5所示。该网络将点云邻域的自注意力机制分为中心自注意力和邻域自注意力，从而使网络更精细地学习采样点邻域的结构特征，并通过采用多种空间编码方法提升网络对点云局部拓扑结构的学习能力。最终，SSA-PointNet++通过构建基于自注意力机制的选择性特征聚合模块，将注意力池化和最大池化提取的特征进行差异性池化，并融合这些全局特征。

Zhao等人（2019）提出PointWeb算法。该算法在每个局部邻域内，通过密集连接各点，并依据该局

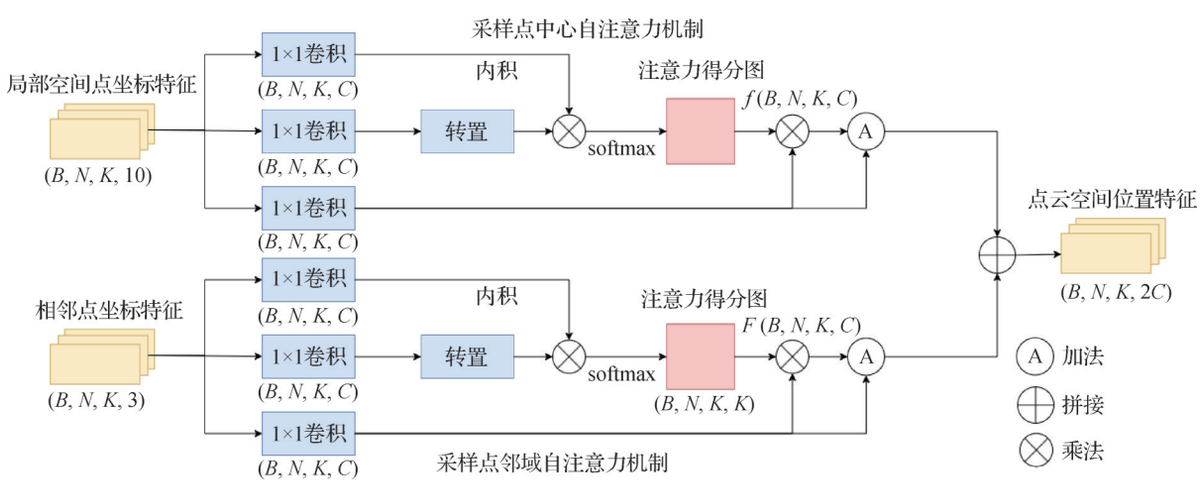


图5 SSA-PointNet++结构框架（吴军 等，2022）

Fig. 5 The architecture of the SSA-PointNet++ model（Wu et al. ，2022）

部区域的特征为每个点分配相应的特征，通过自适应特征调整模块，PointWeb能够学习点之间的相互作用并调整它们的特征。然而，由于PointWeb需要在局部区域内建立所有点对之间的连接，从而导致计算复杂度的增加，尤其在处理较大规模的点云数据时，可能遭遇性能瓶颈。Zhang等人（2023b）提出一种无监督的网络模型GrowSP（growing superpoint），该策略无需任何类型的人工标签或预训练模型，通过一个逐步增长超点的过程来发现3D语义元素。该方法利用特征提取器从输入点云中学习点特征，并通过超点构造函数将输入点云逐步分割为越来越大的超点，使得每个超点尽可能属于同一类别。最后，构建语义原语聚类模块将超点进一步聚类为语义元素，生成伪标签用于训练过程中的优化，从而实现点云的语义分割。

针对大规模点云切割破坏点云的几何特征的问题，杜静和蔡国榕（2021）设计的多特征融合与残差优化的点云语义分割方法PointRES（point residual network），创新性地将多特征与残差优化策略进行了融合。通过多特征提取模块整合点的几何与语义信息，加权得到特征集，应用注意力机制优化此集合并构建特征聚合模块，聚焦关键特征，最后在特征聚合模块中嵌入残差块。

针对当前点云网络只依靠局部邻域进行特征聚合导致特征提取能力不足的问题，刘玉杰等人（2023）提出一种基于注意力的局部信息和全局信息相结合的点云处理网络LGANet。通过引入通道自注意力机制优化局部特征的聚合过程，为了捕获点云中点的远程依赖信息，设计了一种动态学习方法来识别关键点，这些关键点有助于提取全局上下文信息，并且构建了一个特征融合模块，它利用空间注意力机制，使得每个点都能学习到全局上下文信息。该机制有助于提高网络对点云中复杂结构的理解能力，增强了对不同部分之间关系的辨识。刘苏毅等人（2023）引入一种基于递归切片网络的三维点云语义分割与实例分割技术，该方法对输入的点云进行切片，借助双向长短期记忆网络（bidirectional long short-term memory，BiLSTM）捕获每个点的局部信息并整合全局特征，形成一个编码特征矩阵并解码到两个分支上，进行多尺度融合，最后将语义特征和实力特征融合，提升分割性能。但滑窗块处理方法会因为切块的大小问题导致特征提取性能不佳或预测结果产生误差。基于MLP的点云分割方法在数据集上的性能如表4所示。

表4　基于MLP的点云分割方法在数据集上的性能效果

Table 4　Performance of MLP based point cloud segmentation methods on datasets

/%

方法	年份	数据集	OA	mAcc	mIoU
PointNet	2017	ShapeNetPart	–	–	83.70
		ScanNet	73.90	–	14.70
		S3DIS（Area 5）	–	49.00	41.10
		S3DIS（6-folds）	78.60	66.20	47.60
PointNet++	2017	ShapeNetPart	–	–	85.10
		ScanNet	84.50	–	34.30
		S3DIS（6-folds）	81.00	–	54.50
PointWeb	2019	ScanNet	92.30	89.40	–
		S3DIS（Area 5）	86.90	66.60	60.30
		S3DIS（6-folds）	87.30	76.20	66.70
PointRES	2021	S3DIS（6-folds）	87.20	81.70	69.20
		Semantic3D	93.50	–	74.00
SSA-PointNet++	2022	S3DIS（6-folds）	85.10	–	61.10
		Semantic3D	91.50	–	68.10
GrowSP	2023	S3DIS（Area 5）	78.40	57.20	44.50
		S3DIS（6-folds）	76.00	59.40	44.60
		ScanNet	57.30	44.20	25.40
		SemanticKITTI	38.30	19.70	13.20
LGANet	2023	ShapeNetPart	–	–	86.50
刘苏毅等人（2023）	2023	S3DIS（6-folds）	70.30	82.30	89.30
		ShapeNet	–	–	89.20

注：“–”表示未采用对应指标。

4.3.2　基于图卷积的方法

由于在MLP方法发展初期，主要注重提取全局特征，导致大量信息的丢失，基于图卷积的方法因此被提出，该方法将图卷积神经网络（graph convolutional networks，GCN）应用到点云数据的处理上，有效地学习点云中点的局部结构信息并利用它们之间的空间关系。

Wang等人（2019b）提出动态图卷积（dynamic graph CNN，DGCNN），如图6所示。该方法的核心是EdgeConv模块。在每一层网络中，该方法动态构建点云的图结构，并在图上执行卷积操作。考

虑点与其邻域内的其他点之间的关系，通过计算两点特征的函数更新每个点的特征表示。因此，DGCNN 通过堆叠多个 EdgeConv 层，可以学习到点云的全局形状属性，进而提升点云分割任务的性能。

宋巍等人（2021）引入一种结合动态图卷积和空间注意力的点云分类与分割方法 DGCSA，该方法利用动态图卷积构建 K 近邻图，并取边特征，以捕捉点云的局部几何信息。为应对聚合过程中可能的信息丢失，设计了基于点的空间注意力（spatial attention，SA）模块，该模块通过注意力机制自动聚焦于最重要的局部特征部分，学习到比仅使用最大值特征更具有代表性和区分度的特征表达，从而增强有用信息的强调并抑制无关信息，提高模型的分割精度。陈奔等人（2023）提出的基于动态图卷积的混合注意力点云特征学习网络 EDANet 中，设计了 EdgeDual-Attention 模块构建边缘特征集合，将通道注意力和局部区域注意力相结合获取全局和局部特征，并聚合采集到的特征信息进行逐层传递，提升了模型的表现能力。针对现有描述符方法依赖于对噪声敏感的手工特征或不具有旋转不变性等问题，赵宝等人（2025）提出一种基于动态图卷积和 PointNet 的三维局部特征描述符 DGCPNet（dynamic graph convolution and pointnet based network）。该方法首先利用与局部参考框架对齐的局部点云作为输入，通过动态图卷积捕获局部集合特征，同时借助 PointNet 提炼点特征，并且设计了一个由点自注意力（point self-attention，PSA）和局部空间注意力（local spatial-attention，LSA）组成的双重注意力机制层，以更好地融合由动态图卷积和 PointNet 提取的特征，进一步提升特征表示能力。最终，利用最大池化操作和共享的多层感知机将融合后的特征压缩成固定维度的紧凑特征描述符。

Wang 等人（2019a）提出图注意力卷积（graph attention convolution，GAC），该方法通过动态关注邻域特征，为每个邻近点分配依据动态学习特征确定的注意力权重，使卷积核能聚焦于最相关的邻域部分，并形成自适应卷积核形状，以适应不同结构物体，从而显著提升复杂结构理解能力。

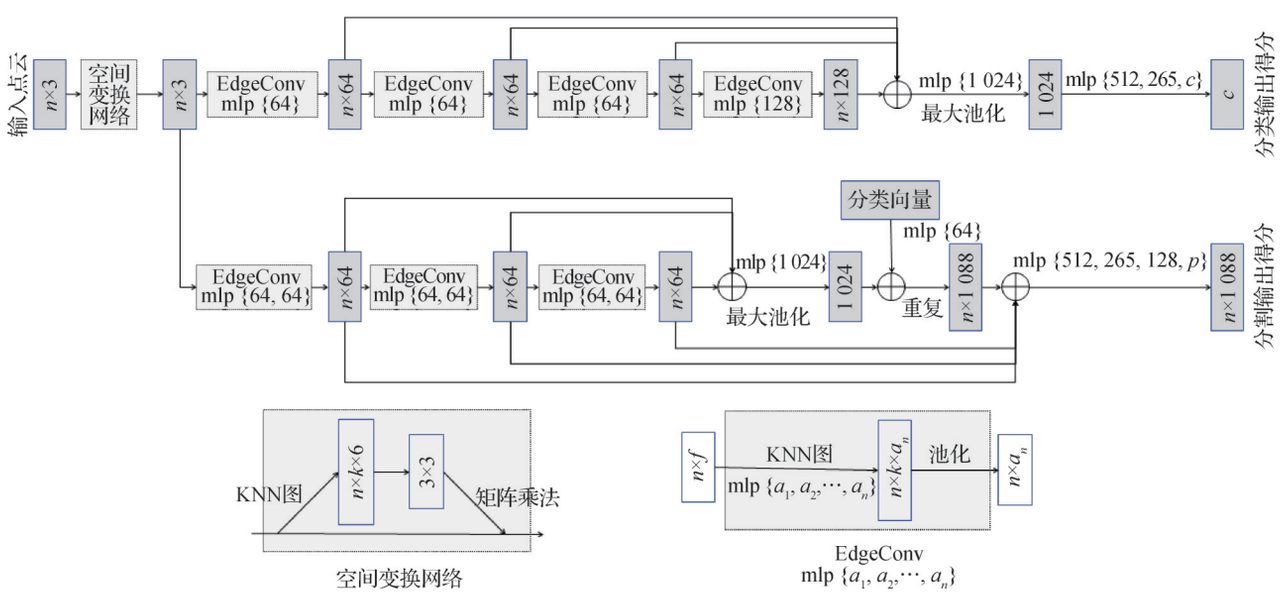


图 6　DGCNN 框架（Wang 等，2019b）

Fig. 6　DGCNN framework（Wang et al.，2019b）

针对目前方法在内存存储和再训练时的高成本问题，Yang 等人（2023）设计了几何和不确定感知三维点云类增量语义分割网络（geometry and uncertainty-aware 3D point cloud class-incremental semantic segmentation）。该方法在基模型训练中基于 DGCNN 获取特征，在新模型训练中使用几何感知特征关系转移模块（geometry-aware feature-relation transfer，GFT）计算每个点与其几何邻居之间的相对关系特征；此外，利用不确定性感知伪标签生成策略（uncertainty-aware pseudo-label generation，UPG），通

过在局部邻域内传播标签消除旧类别伪标签中的噪声，提高伪标签质量，为旧类提供有效的指导。基于图卷积的点云分割方法在数据集上的性能如表5所示。

表5　基于图卷积的点云分割方法在数据集上的性能效果

Table 5　Performance of graph convolutional point cloud segmentation methods on datasets

/%

方法	年份	数据集	OA	mAcc	mIoU
DGCNN	2019	ShapeNetPart	–	–	85.20
		S3DIS(Area 5)	–	–	47.10
		S3DIS(6-folds)	84.10	–	56.10
GACNet	2019	Semantic3D	91.90	–	73.20
		S3DIS(Area 5)	87.80	–	62.80
DGCSA	2021	ShapeNetPart	–	–	85.30
		S3DIS(Area 5)	–	–	50.10
		S3DIS(6-folds)	85.10	–	59.10
EDANet	2023	ShapeNet	–	–	86
Yang等人(2023)	2023	S3DIS(Area 5)	–	–	44.70
		ScanNet	–	–	28.90

注："–"表示未采用对应指标。

4.3.3　基于Transformer的方法

虽然基于图卷积的方法有效获取局部特征，但还是忽略了全局上下文信息，Transformer方法引入点云邻域，注重提取全局和局部特征，获取全局上下文信息。

Transformer是一种基于自我注意机制的新型深度学习架构，在自然语言处理(natural language processing，NLP)和计算机视觉(computer vision，CV)方面的重大成功启发了在点云处理中使用Transformer。一个标准的Transformer编码器一般由6个主要部件组成：输入嵌入、位置编码、自我注意机制、规范化、前馈网络和残差连接。Transformer的核心是自注意机制和前馈神经网络，不依赖于点的顺序，比CNN结构更适合于点云(Lu等，2022)。Guo等人(2021)首次将Transformer架构引入点云处理中，提出PCT，该方法利用FPS算法将原始点云下采样到一个样本点云，提取每个采样点的k个近邻点，计算输出特征，最后通过一个最大池化(max pooling，MP)操作和一个局部特征变换器(例如LBR(lightweight bottleneck residual)模块)处理这些特征。Transformer编码器结构图如图7所示。

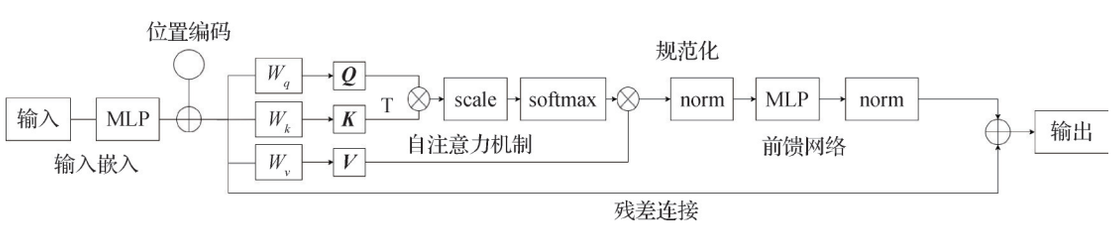


图7　Transformer编码器结构图

Fig. 7　The arthitecture of Transformer

PTv1(point Transformer)(Zhao等，2021)将自注意力层引入三维点云，将矢量注意力而非标量注意力与U-Net风格的编解码器框架相结合，增强了特征表达能力，使得网络能够对每个特征通道进行自适应调制。该方法基于自注意力算子，利用减法关系，将可训练的参数化位置编码添加到注意向量和变换特征中，并且引入位置编码使得自注意力机制能够适应数据中的局部结构信息。此外，以Point Transformer为核心构造残差Point Transformer块，促进局部特征向量之间的信息交换。然而，随着模型的深度和通道数量的增加，权值编码参数的数量也急剧增加，导致严重的过拟合，从而限制了模型的深度发展。为了解决这个问题，Wu等人(2022)提出PTv2(point Transformer V2)，该方法引入分组向量注意力机制，将向量注意力划分为多组，每组共享注意力权重，从而有效减少参数量。同时，PTv2引入了一种增强的位置编码机制，通过一个额外的位置编码乘子强化模型中的位置关系信息。此外，该模型还提出一种新的池化策略基于非重叠分区的池化，通过将点云空间划分为均匀格子，在每个格子内部聚合点特征，从而实现更高效且空间对齐更好的信息聚合，最终表现出比PTv1更优的结果，两者的结构图对照如图8所示(AGG表示aggregation)。

为了能让点云方法直接对长期依赖关系进行建

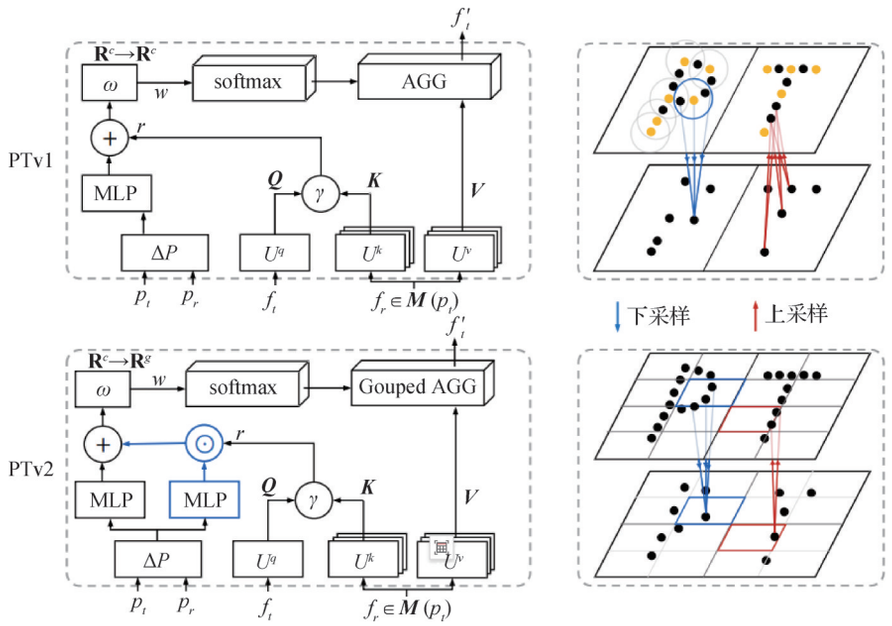


图 8　PTv1 和 PTv2 结构图对照（Wu 等，2022）

Fig. 8　Comparison of PTv1 and PTv2（Wu et al. ，2022）

模，Lai 等人（2022）设计了 Stratified Transformer。该方法创建了一种关键采样策略：分层模式采样点作为 keys，即：密集地采样附近点，稀疏地采样远处点；接着，利用首层点嵌入聚合局部信息，以应对不规则点排列的挑战，并通过上下文相关的相对位置编码自适应地获取位置信息；最后引入了一种提高内存效率的实现方法，通过预先计算所有查询点和键点之间的点积索引，并使用 CUDA（compute unified device architecture）内核直接在这些索引上执行操作，从而显著降低内存复杂度。随着 BERT（bidirectional encoder representaions from Transformer）的广泛应用，Yu 等人（2022）将 BERT 引入到点云中设计了 Point-BERT。该方法首先将点云分割为多个局部点块，并通过包含离散变分自编码器（discrete variational autoencoder，dVAE）的点云分词器生成携带有意义局部信息的离散点令牌。接下来，随机遮蔽输入点云的部分点块，并将这些经过遮蔽处理的点云送入 Transformer 骨干网络中。预训练的目标在于依据由分词器生成的点令牌，准确恢复被遮蔽区域的原始点令牌。通过这样的预训练策略，Point-BERT 能够同时捕获点云的局部几何结构和高层语义模式。

由于缺乏处理大规模的三维场景的有效方法，Park 等人（2022）提出了基于局部自注意的网络 Fast Point Transformer，通过轻量级的局部自注意力层，能够在不牺牲精度的前提下，高效地处理连续的三维坐标信息，利用体素哈希技术提升计算效率，通过构造非均匀的体素网格来近似点云，从而减少计算和存储开销，并且引入了一个新的评价指标一致性得分（coherence score，CScore），用于衡量模型在旋转和平移等变换下预测的一致性。

由于将不规则的点云投影成二维图像会导致几何信息的丢失，刘盛等人（2023）提出一种融合稀疏注意力和实例增强的雷达点云分割算法。该方法通过实例注入实现点云数据增强，引入 Transformer 模块以拓展感受野，使用基于稀疏卷积的空间注意力机制聚焦特征图关键信息，并且设计 TVloss 针对不同类别点云的边缘优化，从而获得良好的分割效果。不同于一些基于监督学习的算法，对比学习是一种无监督学习的方法，不需要标注数据来学习更加鲁棒和具有区分性的表示。针对已有基于对比学习的点云分割算法忽略点云无序和稀疏特性的问题，黄华等人（2025）提出基于对比学习和标签挖掘的点云分割算法（CL2M）。该方法采用自注意力机制，以捕捉点云的全局和局部信息，并结合对比学习与标签挖掘，设计了基于标签分布的对比边界学习模块。通过利用点云中不同实例的标签分布规律，CL2M 能够通过对比学习有效区分不同实例，从而限制提高分割的准确性，尤其是在边界区域的表现。基于 Transformer 的点云分割方法在数据集上的性能如表 6 所示。

表6 基于Transformer的点云分割方法在数据集上的性能

Table 6 Performance of Transformer based point cloud segmentation methods on datasets

/%

方法	年份	数据集	OA	mAcc	mIoU
PCT	2021	S3DIS(Area 5)	–	67.70	61.30
PTv1	2021	ShapeNetPart	–	–	86.60
		S3DIS(Area 5)	90.80	76.50	70.40
		S3DIS(6-folds)	90.20	81.90	73.50
Stratified Transformer	2022	ScanNetv2	–	–	74.30
		S3DIS(Area 5)	91.50	78.10	72.00
		ShapeNetPart	–	–	86.60
Fast PT	2022	S3DIS(Area 5)	–	77.30	70.10
		ScanNet	–	–	72.10
Point-BERT	2022	ShapeNetPart	–	–	85.60
PTv2	2022	S3DIS(Area 5)	91.10	77.90	71.60
		ScanNetv2	–	–	75.20
CL2M	2023	ShapeNetPart	–	–	87.00
		S3DIS(Area 5)	–	–	71.30
刘盛等人(2023)	2023	SemanticKITTI	–	–	64.60
		nuScenes	–	–	75.60

注:“–”表示未采用对应指标。

4.3.4 其他方法

针对密集地注释激光雷达点云过于昂贵和耗时,Unal等人(2022)提出一种基于标记(scribble)监督的LiDAR语义分割方法,并发布了一个名为ScribbleKITTI的数据集,这是首个为LiDAR点云提供标记注释的数据集,包含19个类别的189百万个标记点,占总点数的8.06%。该方法的教师—学生一致性损失在未标记点上引入弱形式的监督,通过一致性损失增强模型的泛化能力。针对类别分布不均衡和点密度随距离变化大的问题,设计了一种针对LiDAR数据特点的自我训练策略,提出一种多分辨率局部语义上下文描述符,增强每个点的语义先验信息,提高伪标签的质量,尤其是在不同分辨率下点的局部环境。

语义分割的新类发现(novel class discovery, NCD)任务旨在仅使用来自标记类的监督分割未标记类,但是这一方法尚未在三维点云数据上得到应用。为此,Riz等人(2023)提出一种在线聚类方法,称为NOPS(novel point segmentation),该方法能在3D点云数据处理中实时更新聚类中心,增强聚类效果。它采用过聚类(over-clustening)和队列策略解决类别缺失问题,并利用伪标签不确定性优化原型,生成高质伪标签。此外,使用Sinkhorn-Knopp算法保证跨视图一致性,并精选高置信度点扩充伪标签集,这些伪标签随后微调预训练模型,进而提升模型在不同视角下的类标签预测一致性,从而提高分割性能。

4.4 基于多模态融合的方法

基于多模态融合的方法结合了点、体素和视图等多种表示方式,以提升点云分割的效果。

Zhang等人(2022a)提出的PVT包含两条分支:基于体素的上分支,通过稀疏窗口注意(sparse window attention, SWA)处理体素的稀疏,从而聚合局部特征;基于点的下分支利用相对注意嵌入捕获全局特征。最终,将这两个分支的输出特征融合,以学习更丰富的语义表示。由于传统点云方法存在提取上下文特征不够丰富并且分割结果存在物体边界模糊的问题,杨军和张琛(2024)引入一种基于边界点估计与稀疏卷积神经网络的三维点云语义分割算法,分为体素分支和点分支:体素分支将点云体素化后,通过稀疏卷积提取上下文语义特征,再解体素化赋予每点初始标签,然后再将标签输入边界点估计(boundary point estimation, BPE)模块得到预测的边界点;点分支则利用改进的动态图卷积(EdgeConv++)获取局部几何特征,并通过空间与通道注意力增强后,与体素特征融合,提升点云特征多样性。

由于现有的point Transformers在生成大型注意图时计算成本高昂,PatchFormer(Zhang等,2022b)引入了PAT(patch attention),自适应地学习一个较小的基集来构建注意力图,该方法利用K-Means聚类既捕捉了全局形状上下文,又实现了计算复杂度的线性增长,还提出一个轻量级的多尺度注意力(multi-scale attention, MST)块将点云体素化为规则网格。然后,在不同尺度上通过深度可分离卷积(depth-width convolution, DWConv)抽取特征,并将这些特征拼接起来,再使用三线性插值将体素特征转换回点云特征,从而为模型提供多尺度特征。基于多模态融合的点云分割方法在数据集上的性能如表7所示。

多种数据表示形式提供不同的信息,结合起来

表 7　基于多模态融合的点云分割方法在数据集上的性能

Table 7　Performance of point cloud segmentation method based on multimodal fusion on datasets

/%

方法	年份	数据集	OA	mAcc	mIoU
PVT	2022	ShapeNet	–	–	86.60
		S3DIS(Area 5)	–	–	68.20
		S3DIS(6-folds)	88.30	76.20	69.20
		SemanticKITTI	–	–	86.50
PatchFormer	2022	ShapeNet	–	–	72.00
		S3DIS(Area 5)	–	–	68.10
杨军和张琛(2024)	2024	S3DIS(Area 5)	90.80	–	69.50
		SemanticKITTI	–	–	62.70

注："–"表示未采用对应指标。

能够更有效地识别细节特征，有助于模型学习更丰富的特征表示，提高分割精度。但是不同表示形式的数据需要在时间和空间上进行精确对齐，这对数据处理提出挑战，并且同时处理多种表示形式的数据会增加计算负担，可能影响模型的训练效率。

5　未来研究方向

大语言模型(large language models，LLM)的迅速发展显著推动了人工智能领域的进步，展现出前所未有的语言处理能力和应用潜力。随着技术的演进，大语言模型开始与视觉、语音等其他模态结合，形成多模态大语言模型，如 VLM(vision-language model)，这些模型在理解复杂现实世界场景表现出色，能够实现图像描述、视频内容生成以及跨模态问答等功能。最近的研究，如 3D-LLM(3D large language model)(Hong 等，2023)，进一步将 3D 物理世界的概念融入模型，使其能够处理点云数据，执行 3D 物体识别、场景描述和导航等相关任务，进一步扩展了大语言模型的应用范围。但是如何让大语言模型更加深刻地理解点云数据中的复杂语义结构，根据空间、几何和纹理等特征对点云进行分割，以及促进点云数据和大语言模型之间的双向交互学习仍面临挑战，尤其是考虑到点云数据的稀疏性和不规则性。

在点云场景语义分割领域，尽管全监督深度学习方法凭借其优异的性能仍占据主导地位，但半监督和无监督学习方法作为减少对昂贵标注数据依赖的有效途径，正逐渐成为研究热点。然而，这些方法在实际应用中面临着一系列严峻挑战：首先，数据的无监督特性可能导致学习信号的减弱，缺乏直接的分类或分割指导，使得模型难以准确捕捉点云中的复杂结构和细微特征。其次，伪标签的生成与选择问题在半监督学习框架中尤为突出，错误的伪标签不仅导致模型学习错误的特征表示，还会影响影响最终的分割精度。此外，在复杂场景下，尤其是在存在遮挡、形变以及多样对象共存的情况下，如何保持分割的准确性与鲁棒性仍然是一个亟待解决的难题。进一步而言，算法的计算效率与内存占用问题同样不容忽视，特别是在处理大规模无标签点云数据集时，如何有效平衡模型复杂度与运行效率是一个重要的研究考量。最后，设计更加高效和鲁棒的特征表示学习机制，以充分利用点云的几何、颜色等多模态信息，增强模型的泛化能力和对未见场景的适应性，仍是当前研究中亟待解决的关键问题。因此，尽管半监督和无监督点云分割方法展现出巨大潜力，但克服这些挑战，推动其在实际应用中的有效落地，仍然是一个长期且充满挑战的研究方向。

6　结　语

本文立足于场景语义分割领域，对点云场景语义分割研究进行分析总结，从任务及其面临的挑战对点云和场景语义分割进行介绍，并对深度学习点云算法的数据集、采样方法、特征提取、实验对比方法以及三维点云场景语义分割未来研究方向进行了总结。点云场景语义分割作为一个充满活力的研究领域，其持续发展将对智能系统感知周围环境的能力产生深远影响。本文希望可以为后续的研究提供一定的参考价值，并激发更多针对现有挑战的创新解决方案。

参考文献(References)

Adam J M, Liu W Q, Zang Y, Afzal M K, Bello S A, Muhammad A U, Wang C and Li J. 2023. Deep learning-based semantic segmentation of urban-scale 3D meshes in remote sensing: a survey. International Journal of Applied Earth Observation and Geoinformation, 121: #103365 [DOI: 10.1016/j.jag.2023.103365]

Armeni I, Sener O, Zamir A R, Jiang H, Brilakis I, Fischer M and Savarese S. 2016. 3D Semantic parsing of large-scale indoor spaces//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 1534-1543 [DOI: 10.1109/CVPR.2016.170]

Behley J, Garbade M, Milioto A, Quenzel J, Behnke S, Stachniss C and Gall J. 2019. SemanticKITTI: a dataset for semantic scene understanding of LiDAR sequences//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 9296-9306 [DOI: 10.1109/ICCV.2019.00939]

Chang A X, Funkhouser T, Guibas L, Hanrahan P, Huang Q, Li Z, Savarese S, Savva M, Song S, Su H, Xiao J X, Yi L and Yu F. 2015. ShapeNet: an information-rich 3D model repository [EB/OL]. [2024-11-07]. http://arxiv.org/pdf/1512.03012.pdf

Charles R Q, Su H, Kaichun M and Guibas L J. 2017. PointNet: deep learning on point sets for 3D classification and segmentation//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 77-85 [DOI: 10.1109/CVPR.2017.16]

Chen B, Xie X Y and Liu S. 2023. 3D point cloud feature learning network based on hybrid attention and dynamic graph convolution. Computer Technology and Development, 33(10): 157-162 (陈奔, 谢晓尧, 刘嵩. 2023. 基于动态图卷积的混合注意力点云特征学习网络. 计算机技术与发展, 33(10): 157-162) [DOI: 10.3969/j.issn.1673-629X.2023.10.024]

Dai A, Chang A X, Savva M, Halber M, Funkhouser T and Nießner M. 2017. ScanNet: richly-annotated 3D reconstructions of indoor scenes//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 2432-2443 [DOI: 10.1109/CVPR.2017.261]

Du J and Cai G R. 2021. Point cloud semantic segmentation method based on multi-feature fusion and residual optimization. Journal of Image and Graphics, 26(5): 1105-1116 (杜静, 蔡国榕. 2021. 多特征融合与残差优化的点云语义分割方法. 中国图象图形学报, 26(5): 1105-1116) [DOI: 10.11834/jig.200374]

Guo M H, Cai J X, Liu Z N, Mu T J, Martin R R and Hu S M. 2021. PCT: point cloud transformer. Computational Visual Media, 7(2): 187-199 [DOI: 10.1007/s41095-021-0229-5]

Hackel T, Savinov N, Ladicky L, Wegner J, Schindler K, Pollefeys M. 2017. Semantic3D.net: a new large-scale point cloud classification benchmark [EB/OL]. [2024-11-07]. http://arxiv.org/pdf/1704.03847.pdf

Hong Y N, Zhen H Y, Chen P H, Zheng S H, Du Y L, Chen Z F and Gan C. 2023. 3D-LLM: injecting the 3D world into large language models//Proceedings of the 37th International Conference on Neural Information Processing Systems. New Orleans, USA: Curran Associates Inc.: 20482-20494 [DOI: 10.5555/3666122.3667022]

Hu Q Y, Yang B, Xie L H, Rosa S, Guo Y L, Wang Z H, Trigoni N and Markham A. 2020. RandLA-Net: efficient semantic segmentation of large-scale point clouds//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 11105-11114 [DOI: 10.1109/CVPR42600.2020.01112]

Huang H, Bu Y F, Xu H L and Wang X R. 2025. Point cloud segmentation algorithm based on contrastive learning and label mining. Journal of Computer Research and Development, 62(1): 132-143 (黄华, 卜一凡, 许宏丽, 王晓荣. 2025. 基于对比学习和标签挖掘的点云分割算法. 计算机研究与发展, 62(1): 132-143) [DOI: 10.7544/issn1000-1239.202330491]

Kong L D, Liu Y Q, Chen R N, Ma Y X, Zhu X G, Li Y K, Hou Y N, Qiao Y and Liu Z W. 2023. Rethinking range view representation for LiDAR segmentation//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 228-240 [DOI: 10.1109/ICCV51070.2023.00028]

Lai X, Liu J H, Jiang L, Wang L W, Zhao H S, Liu S, Qi X J and Jia J Y. 2022. Stratified transformer for 3D point cloud segmentation//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 8490-8499 [DOI: 10.1109/CVPR52688.2022.00831]

Li L, Shum H P H and Breckon T P. 2023. Less is more: reducing task and model complexity for 3D point cloud semantic segmentation//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 9361-9371 [DOI: 10.1109/CVPR52729.2023.00903]

Liu S, Cao Y F, Huang W H and Li D D. 2023. LiDAR point cloud semantic segmentation combined with sparse attention and instance enhancement. Journal of Image and Graphics, 28(2): 483-494 (刘盛, 曹益烽, 黄文豪, 李丁达. 2023. 融合稀疏注意力和实例增强的雷达点云分割. 中国图象图形学报, 28(2): 483-494) [DOI: 10.11834/jig.210787]

Liu S Y, Chi J N, Wu C D and Xu F. 2023. Recurrent slice networks-based 3D point cloud-relevant integrated segmentation of semantic and instances. Journal of Image and Graphics, 28(7): 2135-2150 (刘苏毅, 迟剑宁, 吴成东, 徐方. 2023. 基于递归切片网络的三维点云语义分割与实例分割. 中国图象图形学报, 28(7): 2135-2150) [DOI: 10.11834/jig.220154]

Liu Y C, Fan B, Meng G F, Lu J W, Xiang S M and Pan C H. 2019. DensePoint: learning densely contextual representation for efficient point cloud processing//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South): IEEE: 5238-5247 [DOI: 10.1109/ICCV.2019.00534]

Liu Y J, Yuan Y F, Sun X R and Li Z M. 2023. A point cloud processing network combining global and local information. Journal of Zhejiang University (Science Edition), 50(6): 770-780 (刘玉杰, 原亚夫, 孙晓瑞, 李宗民. 2023. 局部信息和全局信息相结合的点云处理网络. 浙江大学学报(理学版), 50(6): 770-780) [DOI: 10.3785/j.issn.1008-9497.2023.06.012]

Lu D N, Xie Q, Wei M Q, Gao K, Xu L L and Li J. 2022. Transformers in 3D point clouds: a survey [EB/OL]. [2024-11-07].

https://arxiv.org/pdf/2205.07417.pdf

Maturana D and Scherer S. 2015. VoxNet: a 3D convolutional neural network for real-time object recognition//Proceedings of 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems. Hamburg, Germany: IEEE: 922-928 [DOI: 10.1109/IROS.2015.7353481]

Milioto A, Vizzo I, Behley J and Stachniss C. 2019. RangeNet++: fast and accurate LiDAR semantic segmentation//Proceedings of 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems. Macau, China: IEEE: 4213-4220 [DOI: 10.1109/IROS40897.2019.8967762]

Park C, Jeong Y, Cho M and Park J. 2022. Fast point transformer//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 16928-16937 [DOI: 10.1109/CVPR52688.2022.01644]

Park J, Kim C, Kim S and Jo K. 2023. PCSCNet: fast 3D semantic segmentation of LiDAR point cloud for autonomous car using point convolution and sparse convolution network. Expert Systems with Applications, 212: #118815 [DOI: 10.1016/j.eswa.2022.118815]

Qi C R, Yi L, Su H and Guibas L J. 2017. PointNet++: deep hierarchical feature learning on point sets in a metric space//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc.: 5105-5114 [DOI: 10.5555/3295222.3295263]

Riegler G, Ulusoy A O and Geiger A. 2017. OctNet: learning deep 3D representations at high resolutions//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 6620-6629 [DOI: 10.1109/CVPR.2017.701]

Riz L, Saltori C, Ricci E and Poiesi F. 2023. Novel class discovery for 3D point cloud semantic segmentation//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 9393-9402 [DOI: 10.1109/CVPR52729.2023.00906]

Song W, Cai W Y, He S Q and Li W J. 2021. Dynamic graph convolution with spatial attention for point cloud classification and segmentation. Journal of Image and Graphics, 26(11): 2691-2702 (宋巍, 蔡万源, 何盛琪, 李文俊. 2021. 结合动态图卷积和空间注意力的点云分类与分割. 中国图象图形学报, 26(11): 2691-2702) [DOI: 10.11834/jig.200550]

Su H, Maji S, Kalogerakis E and Learned-Miller E. 2015. Multi-view convolutional neural networks for 3D shape recognition//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE: 945-953 [DOI: 10.1109/ICCV.2015.114]

Thomas H, Qi C R, Deschaud J E, Marcotegui B, Goulette F and Guibas L. 2019. KPConv: flexible and deformable convolution for point clouds//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 6410-6419 [DOI: 10.1109/ICCV.2019.00651]

Unal O, Dai D X and Van Gool L. 2022. Scribble-supervised LiDAR semantic segmentation//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 2687-2697 [DOI: 10.1109/CVPR52688.2022.00272]

Wang L, Huang Y C, Hou Y L, Zhang S M and Shan J. 2019a. Graph attention convolution for point cloud semantic segmentation//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 10288-10297 [DOI: 10.1109/CVPR.2019.01054]

Wang Y, Sun Y B, Liu Z W, Sarma S E, Bronstein M M and Solomon J M. 2019b. Dynamic graph CNN for learning on point clouds. ACM Transactions on Graphics, 38(5): #146 [DOI: 10.1145/3326362]

Wu J, Cui Y, Zhao X M, Chen R X and Xu G. 2022. SSA-PointNet++: a space self-attention CNN for the semantic segmentation of 3D point cloud. Journal of Computer-Aided Design and Computer Graphics, 34(3): 437-448 (吴军, 崔玥, 赵雪梅, 陈睿星, 徐刚. 2022. SSA-PointNet++: 空间自注意力机制下的3D点云语义分割网络. 计算机辅助设计与图形学学报, 34(3): 437-448) [DOI: 10.3724/SP.J.1089.2022.18900]

Wu X Y, Lao Y X, Jiang L, Liu X H and Zhao H S. 2022. Point transformer V2: grouped vector attention and partition-based pooling//Proceedings of the 36th International Conference on Neural Information Processing Systems. New Orleans, USA: Curran Associates Inc.: 33330-33342 [DOI: 10.5555/3600270.3602685]

Xu M T, Ding R Y, Zhao H S and Qi X J. 2021. PAConv: position adaptive convolution with dynamic kernel assembling on point clouds//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE: 3172-3181 [DOI: 10.1109/CVPR46437.2021.00319]

Yang J and Zhang C. 2024. Semantic segmentation of 3D point cloud based on boundary point estimation and sparse convolution neural network. Journal of Zhejiang University (Engineering Science), 58(6): 1121-1132 (杨军, 张琛. 2024. 基于边界点估计与稀疏卷积神经网络的三维点云语义分割. 浙江大学学报(工学版), 58(6): 1121-1132) [DOI: 10.3785/j.issn.1008-973X.2024.06.003]

Yang Y W, Hayat M, Jin Z, Ren C and Lei Y J. 2023. Geometry and uncertainty-aware 3D point cloud class-incremental semantic segmentation//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 21759-21768 [DOI: 10.1109/CVPR52729.2023.02084]

Yu T, Meng J J and Yuan J S. 2018. Multi-view harmonized bilinear network for 3D object recognition//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 186-194 [DOI: 10.1109/CVPR.2018.00027]

Yu X M, Tang L L, Rao Y M, Huang T J, Zhou J and Lu J W. 2022. Point-BERT: pre-training 3D point cloud transformers with masked point modeling//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 19291-19300 [DOI: 10.1109/CVPR52688.2022.01871]

Zhang C, Wan H C, Shen X Y and Wu Z Z. 2022a. PVT: point-voxel

transformer for point cloud learning. International Journal of Intelligent Systems, 37(12): 11985-12008 [DOI: 10.1002/int.23073]

Zhang C, Wan H C, Shen X Y and Wu Z Z. 2022b. PatchFormer: an efficient point transformer with patch attention//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 11789-11798 [DOI: 10.1109/CVPR52688.2022.01150]

Zhang T, Yuan H B, Qi L, Zhang J N, Zhou Q Y, Ji S P, Yan S C and Li X T. 2024. Point cloud mamba: point cloud learning via state space model [EB/OL]. [2024-11-07]. https://arxiv.org/pdf/2403.00762.pdf

Zhang Z H, Yang B, Wang B and Li B. 2023b. GrowSP: unsupervised semantic segmentation of 3D point clouds//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 17619-17629 [DOI: 10.1109/CVPR52729.2023.01690]

Zhao B, Wang Z H, Jia Z H, Liang D and Liu Q. 2025. Three-dimensional local feature descriptor based on dynamic graph convolution and PointNet. Journal of Computer-Aided Design and Computer Graphics, 37(1): 89-99 (赵宝, 王梓涵, 贾兆红, 梁栋, 刘强. 2025. 基于动态图卷积和PointNet的三维局部特征描述符. 计算机辅助设计与图形学学报, 37(1): 89-99) [DOI: 10.3724/SP.J.1089.2023-00083]

Zhao H S, Jiang L, Fu C W and Jia J Y. 2019. PointWeb: enhancing local neighborhood features for point cloud processing//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 5560-5568 [DOI: 10.1109/CVPR.2019.00571]

Zhao H S, Jiang L, Jia J Y, Torr P and Koltun V. 2021. Point transformer//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 16239-16248 [DOI: 10.1109/ICCV48922.2021.01595]

作者简介

仇志江，男，硕士研究生，主要研究方向为计算机视觉、三维场景理解和机器学习。
E-mail: 238003071148@stu.nepu.edu.cn
冯小青，通信作者，女，副教授，主要研究方向为三维点云模型分割、数字孪生和机器学习。
E-mail: fenglinda@zufe.edu.cn
张林，女，副教授，主要研究方向为计算机视觉、数据挖掘、复杂网络和自然语言处理。E-mail: zhanglin_hz@163.com
姚垚，男，讲师，主要研究方向为深度学习和数字孪生。
E-mail: yaoyao@zufe.edu.cn
高俊涛，男，副教授，主要研究方向为大数据分析、数字孪生和软件建模。E-mail: gjt@nepu.edu.cn