地矿测绘
Geological and Mineral Surveying and Mapping
第 8 卷◆第 4 期◆版本 1.0◆2025 年
文章类型：论文|刊号（ISSN）：2630-4732 /（中图刊号）：561GL001

基于深度学习的点云语义分割在实景三维中的应用研究

赵晶晶
新疆维吾尔自治区测绘院
DOI:10.12238/gmsm.v8i4.2225

[摘　要] 实景三维作为数字中国建设的重要基础设施,其智能化处理与精细化分析已成为测绘地理信息领域的研究热点。点云数据作为实景三维的核心数据形式,其语义分割技术直接影响着三维场景的理解与应用深度。本文系统梳理了基于深度学习的点云语义分割技术发展脉络,重点探讨了PointNet系列、图卷积网络、Transformer等先进网络架构在点云处理中的创新应用。针对实景三维场景中点云数据规模大、类别不均衡、边界模糊等技术难点,本文提出了多尺度特征融合与注意力机制相结合的改进方案,并在城市建筑物提取、道路场景理解、文物保护等典型应用场景中进行了验证。实验结果表明,所提方法在大规模点云数据集上的平均交并比(mIoU)达到82.6%,相比传统方法有显著提升,有效提升了实景三维场景的智能化解译能力。本研究为实景三维数据的自动化处理提供了新的技术路径,对推动智慧城市、数字孪生等领域的发展具有重要的理论意义和应用价值。
[关键词] 深度学习；点云语义分割；实景三维；三维场景理解；特征提取
中图分类号：D422.63　文献标识码：A

Research on the application of point cloud semantic segmentation based on deep learning in real-world 3D

Jingjing Zhao
Institute of Surveying and Mapping, Xinjiang Uygur Autonomous Region

[Abstract] As an important infrastructure for the construction of Digital China, the intelligent processing and refined analysis of real-world 3D has become a research hotspot in the field of surveying and geoinformation. Point cloud data, as the core data form of real-world 3D, and its semantic segmentation technology directly affects the understanding and application depth of 3D scenes. This paper systematically reviews the development of point cloud semantic segmentation technology based on deep learning, focusing on the innovative applications of advanced network architectures such as PointNet series, graph convolutional networks, and Transformers in point cloud processing. The experimental results show that the mean Intersection over Union (mIoU) of the proposed method on a large-scale point cloud dataset reaches 82.6%, which is significantly improved compared to traditional methods and effectively enhances the intelligent interpretation capability of real-world 3D scenes. This study provides a new technical path for the automated processing of real-world 3D data, and has important theoretical significance and application value in promoting the development of smart cities, digital twins, and other fields.
[Key words] Deep Learning; Point Cloud Semantic Segmentation; Real-world 3D; 3D Scene Understanding; Feature Extraction

引言

随着激光扫描、倾斜摄影测量、移动测量等技术的快速发展,实景三维数据的获取效率和精度不断提升,已成为数字化表达真实世界的重要手段[1]。点云作为实景三维的基础数据形式,承载着丰富的几何、纹理和语义信息,在智慧城市、自动驾驶、文化遗产保护等领域发挥着越来越重要的作用[2]。然而,海量点云数据的智能化解译仍面临诸多挑战,其中语义分割作为点云理解的关键环节,直接决定了后续应用的效果和精度。

传统的点云语义分割方法主要依赖于手工设计的几何特征和机器学习分类器,如基于法向量、曲率、协方差矩阵等局部几何特征的提取与分类方法。这类方法虽然在特定场景下取得了一定效果,但面对复杂多变的实景三维场景,其特征表达能力有

文章类型：论文|刊号（ISSN）：2630-4732 /（中图刊号）：561GL001

限,泛化性能较差[3]。近年来,深度学习技术的突破性进展为点云语义分割带来了新的解决思路。特别是PointNet的提出,首次实现了对无序点云的端到端学习,开创了点云深度学习的新纪元[4]。此后,研究者们从网络架构、特征提取、损失函数等多个维度持续改进,涌现出PointNet++、DGCNN、Point Transformer等一系列优秀模型,使得点云语义分割的精度和效率得到显著提升[5]。

在实景三维应用场景中,点云数据呈现出与标准数据集不同的特点:一是数据规模庞大,单个场景往往包含数千万甚至上亿个点;二是类别分布极不均衡,建筑物、地面等类别占据主导,而路灯、交通标志等小目标类别样本稀少;三是目标边界模糊,不同类别之间常存在遮挡、黏连等现象[6-9]。这些特点对深度学习模型的设计提出了更高要求。本文针对实景三维场景的特殊性,系统研究了基于深度学习的点云语义分割方法,提出了融合多尺度特征和注意力机制的改进网络架构,并在典型应用场景中进行了验证与分析。

1 点云语义分割关键技术

1.1点云数据特性与预处理

点云数据本质上是三维空间中离散点的集合,每个点通常包含三维坐标(x,y,z)以及可选的强度、颜色、法向量等属性信息。与图像数据不同,点云具有无序性、非结构化、密度不均匀等固有特性,这使得传统卷积神经网络难以直接应用。在实景三维场景中,原始点云数据往往还包含大量噪声点、离群点,需要进行系统的预处理。预处理流程主要包括:噪声滤波、降采样、坐标归一化、法向量估计等步骤。统计滤波和半径滤波是常用的噪声去除方法,而体素网格下采样和最远点采样则是控制点云规模的有效手段。

1.2基于深度学习的特征提取

深度学习在点云处理中的核心挑战是如何有效提取特征并保持点云的排列不变性。PointNet通过共享的多层感知机(MLP)和对称函数(最大池化)实现了对点云的直接处理,其创新性的网络架构解决了排列不变性问题。PointNet++在此基础上引入了层次化特征学习机制,通过集合抽象层逐层提取局部到全局的多尺度特征,显著提升了对局部几何结构的感知能力。图卷积网络(GCN)从图论角度建模点云,通过动态构建k近邻图捕捉点与点之间的关联关系,EdgeConv操作能够自适应地学习边特征,增强了模型的表达能力。

近年来,Transformer架构在计算机视觉领域的成功引发了研究者对其在点云处理中应用的关注。Point Transformer将自注意力机制引入点云特征学习,通过位置编码和向量注意力机制,能够建模长距离的空间依赖关系,在复杂场景的语义分割任务中展现出优异性能。此外,基于稀疏卷积的方法通过在非空体素上进行卷积操作,大幅降低了计算复杂度,使得处理大规模点云场景成为可能。

1.3多尺度特征融合策略

实景三维场景中的目标尺度跨度大,从数厘米的交通标志到数十米的建筑物,单一尺度的特征难以兼顾不同大小的目标。多尺度特征融合通过整合不同感受野的特征表示,能够显著提升分割精度。常用的融合策略包括:特征金字塔结构,通过自顶向下和自底向上的路径融合不同层次的特征;多分支并行结构,采用不同采样率或卷积核提取多尺度特征后进行拼接;注意力加权融合,通过学习不同尺度特征的权重实现自适应融合。在实际应用中,需要根据场景特点和目标类别分布选择合适的融合方式。

2 面向实景三维的改进方法

2.1网络架构设计

针对实景三维场景的特点,本文设计了一种编码器-解码器架构的语义分割网络。编码器采用改进的PointNet++结构,通过四层集合抽象逐级提取特征,每层的采样率分别设置为原始点数的50%、25%、12.5%和6.25%,确保充分捕捉从局部到全局的几何信息。在特征提取过程中,引入了边缘增强模块,通过计算点与其邻域点的特征差异,强化目标边界的判别能力。解码器部分采用特征传播层,将编码器各层的特征通过跳跃连接与上采样特征融合,保留更多的细节信息。

2.2注意力机制优化

为解决类别不均衡问题,本文在网络中嵌入了双重注意力模块,包括通道注意力和空间注意力。通道注意力通过全局平均池化和全局最大池化提取通道统计信息,学习不同特征通道的重要性权重,抑制冗余特征的影响。空间注意力则通过计算每个点在特征空间中的相似度,动态调整不同位置特征的权重,使网络更关注小目标和边界区域。双重注意力机制的引入使模型能够自适应地聚焦于难分类的区域和类别,有效缓解了类别不均衡带来的负面影响。

2.3损失函数设计

本文采用组合损失函数来优化模型训练。主损失采用Focal Loss,通过引入调制因子降低易分类样本的权重,使模型更关注困难样本。针对边界模糊问题,增加了边界敏感损失项,对边界区域的预测误差施加更大惩罚。此外,引入了Lovász-Softmax损失优化mIoU指标,直接作用于评价指标的提升。组合损失函数表示为:$L_{total}=\alpha \cdot L_{focal}+\beta \cdot L_{boundary}+\gamma \cdot L_{lovász}$,其中$\alpha$、$\beta$、$\gamma$为权重系数,通过网格搜索确定最优值分别为0.6、0.2、0.2。

3 实景三维典型应用场景

3.1城市建筑物精细提取

城市建筑物的精细提取是实景三维应用的基础任务,对城市规划、不动产测量、三维建模等具有重要价值。本文方法在城市场景点云数据集上进行了测试,该数据集包含15个城市区域,总计约3.2亿个点,涵盖建筑物、道路、植被、车辆等8个类别。实验结果显示,建筑物类别的F1分数达到91.3%,相比基准方法提升了6.8个百分点。特别是在建筑物边界和复杂结构区域,如阳台、飞檐、幕墙等,分割精度有明显改善,平均边界误差从0.42米降低至0.21米。

文章类型：论文|刊号（ISSN）：2630-4732 /（中图刊号）：561GL001

3.2道路场景智能解析

道路场景理解是自动驾驶和智能交通系统的关键技术。本文在语义KITTI数据集和自采集的城市道路数据上进行了验证。道路场景包含路面、人行道、车辆、行人、交通标志、路灯等多种元素,类别间尺度差异显著。通过多尺度特征融合和注意力机制,所提方法对小目标的识别能力显著增强,交通标志和路灯的召回率分别达到78.5%和73.2%,较传统方法提升了22.1和18.6个百分点。在动态场景中,对移动目标如车辆和行人的分割准确率也保持在较高水平,整体mIoU达到79.4%。

3.3文化遗产数字化保护

文化遗产的数字化保护需要对古建筑、石窟、雕塑等进行精细的三维重建和语义标注。本文与某文物保护单位合作,对一处古建筑群进行了点云语义分割实验。该数据集包含殿堂、廊柱、石雕、彩绘等多种文物要素,几何结构复杂且存在大量残损。通过训练专门的分类模型,实现了对不同建筑构件和装饰元素的自动识别,总体分割准确率达到85.7%。语义分割结果可直接用于病害分析、修复规划、虚拟展示等应用,大幅提升了文物保护工作的效率和科学性。

4 实验结果与分析

4.1数据集与评价指标

本文在多个公开数据集和自采集数据上进行了全面实验,包括S3DIS室内场景数据集、SemanticKITTI道路数据集,以及自建的城市实景三维数据集。评价指标采用像素准确率(PA)、平均准确率(mAcc)、平均交并比(mIoU)和F1分数。实验平台配置为NVIDIA RTX 3090 GPU,采用PyTorch深度学习框架,优化器使用Adam,初始学习率设置为0.001,批大小为8,训练200个epoch。

4.2对比实验

将本文方法与多种先进算法进行对比,包括PointNet++、DGCNN、PointCNN、RandLA-Net和KPConv。在S3DIS数据集上,本文方法的mIoU达到68.9%,超过PointNet++的61.3%和DGCNN的64.2%。在SemanticKITTI数据集上,mIoU为59.7%,与KPConv的59.4%相当,但推理速度提升了35%。在自建的城市数据集上,本文方法的mIoU达到82.6%,显著优于其他方法,特别是在建筑物、道路等主要类别上保持了高精度,同时对小目标类别的识别能力也有明显优势。

4.3消融实验

为验证各模块的有效性,进行了详细的消融实验。移除边缘增强模块后,mIoU下降2.3个百分点,表明边缘信息对提升分割精度的重要性。去除双重注意力机制,小目标类别的F1分数平均下降5.7个百分点,证明注意力机制在处理类别不均衡问题上的作用。单独使用交叉熵损失替代组合损失,mIoU降低3.8个百分点。这些结果表明,所提出的各项改进措施均对模型性能有显著贡献。

4.4效率分析

在处理效率方面,本文方法在包含100万个点的场景中,单次前向推理时间约为1.2秒,能够满足大多数应用场景的实时性要求。与基于体素化的稀疏卷积方法相比,虽然推理速度略慢,但无需体素化预处理,保留了更完整的几何信息。通过模型剪枝和量化技术,可进一步将推理速度提升40%,同时精度损失控制在1%以内,为实际部署提供了可行方案。

5 结论与展望

本文系统研究了基于深度学习的点云语义分割技术在实景三维中的应用,提出了融合多尺度特征和双重注意力机制的改进网络架构,有效解决了实景三维场景中数据规模大、类别不均衡、边界模糊等难题。通过在多个典型应用场景的验证,证明了所提方法的有效性和实用性,为实景三维数据的智能化处理提供了新的技术手段。

未来的研究方向包括:一是探索更高效的网络架构,在保证精度的前提下进一步降低计算复杂度,实现真正的实时处理;二是研究弱监督和自监督学习方法,减少对标注数据的依赖,降低应用成本;三是融合多模态信息,将点云与影像、倾斜摄影模型等数据结合,提升语义理解的全面性;四是加强模型的泛化能力研究,使其能够适应不同地域、不同季节、不同传感器采集的数据;五是开展面向特定应用的定制化研究,如变化检测、目标提取、场景重建等,推动实景三维技术在更广泛领域的深度应用。

[参考文献]

[1]张祖勋,张剑清.实景三维中国建设的关键技术问题与发展趋势[J].测绘学报,2022,51(7):1203-1213.

[2]李德仁,李明.面向智慧城市的实景三维时空信息平台[J].武汉大学学报(信息科学版),2021,46(6):801-809.

[3]陈驰,王成,李清泉.基于深度学习的点云语义分割方法综述[J].测绘学报,2020,49(5):555-567.

[4]杨必胜,董震,梁福逊.基于PointNet++的机载LiDAR点云语义分割[J].测绘学报,2020,49(8):981-990.

[5]王明常,张文月,李琦.融合图卷积与Transformer的点云语义分割网络[J].中国图象图形学报,2023,28(4):1145-1156.

[6]程效军,章玉,陈动.大规模城市点云场景语义分割方法研究[J].测绘通报,2021,(10):45-51.

[7]史文中,陈云浩.实景三维在新型智慧城市中的应用[J].地球信息科学学报,2022,24(1):3-14.

[8]郭丙轩,李杏朝,马洪超.基于注意力机制的点云语义分割算法[J].计算机学报,2021,44(9):1812-1825.

[9]韩斌,吴波,王璐.面向不均衡数据的点云语义分割损失函数改进[J].中国科学:信息科学,2022,52(6):1089-1103.

作者简介:

赵晶晶(1979--),女,中国新疆乌鲁木齐人,本科,副高级工程师,从事地理信息系统,航空摄影测量研究。