

基于深度学习的大规模室外场景 Scan2BIM 研究综述

莫洋¹⁾, 王晶^{1)*}, 蒋海峰¹⁾, 张业星¹⁾, 曲柄宇¹⁾, 王雷¹⁾, 冯卓航^{1,2)}

¹⁾(中国交通信息科技集团有限公司 北京 101399)

²⁾(中国科学院大学 北京 100049)

(wangjing110@ccccltd.cn)

摘要: 当前,大规模室外基础设施的数字化需求持续扩大,基于深度学习的自动扫描到建筑信息模型(scanning to building information modeling, Scan2BIM)通过卓越的特征学习能力和自动化流程显著提升了建模精度和构建速度,在结构复杂的室外场景重建中发挥了关键作用.文中介绍了 Scan2BIM 的 4 大核心模块及其相关研究进展.其中,针对 3D 点云获取模块,从采集设备与采集来源 2 个维度概括了 3D 点云数据采集的技术发展,并着重梳理了代表性 3D 点云数据集;根据学习方式的不同,将大规模点云对齐算法划分为基于优化和深度学习 2 大类,并从精准度、计算效率、鲁棒性等多维度对比分析了相关工作;在点云分割模块中,分别对点云全景分割和点云实例分割算法通过统一的评估指标进行了整理归纳;对于 BIM 自动化建模,简述了 BIM 核心互操作标准体系,并分类总结了多种几何实体建模与关系建模算法.最后,通过深入分析和前瞻性探讨,指出了现阶段大规模室外场景建模的高效性、精准性、泛化性与统一性的无法有效结合的问题;未来将重点围绕多源数据融合建模、精度与鲁棒性协同优化、端到端 Scan2BIM 通用框架构建以及大模型应用与探索等方向展开.

关键词: Scan2BIM; 深度学习; 3D 点云; 逆向建模; 大规模室外场景

中图分类号: TP391.41

DOI: 10.3724/SP.J.1089.2024-00265

Deep Learning-Based Scan2BIM for Large-Scale Scenes: A Comprehensive Review

Mo Yang¹⁾, Wang Jing^{1)*}, Jiang Haifeng¹⁾, Zhang Yexing¹⁾, Qu Bingyu¹⁾, Wang Lei¹⁾, and Feng Zhuohang^{1,2)}

¹⁾(China Communications Information & Technology Group Co., Ltd., Beijing 101399)

²⁾(University of Chinese Academy of Sciences, Beijing 100049)

Abstract: With the increasing demand for digitizing large-scale outdoor infrastructure, the method of deep learning-based Scan2BIM (scanning to building information modeling) substantially improves modeling precision and construction speed by leveraging its powerful feature learning capabilities and automated workflows, which play a crucial role in the 3D reconstruction of structurally complex outdoor scenes. This paper introduces the four main modules of Scan2BIM and their corresponding research advancements. Firstly, the development of 3D point cloud data acquisition is summarized from the perspectives of both acquisition devices and sources, with dedicated emphasis on the systematic organization and analytical evaluation of representative 3D point cloud datasets. Secondly, large-scale point cloud registration algorithms are systematically categorized into two principal classes, according to learning approaches, which are optimization-driven and deep learning-based, followed by the multidimensional comparative analysis of existing methodologies through critical performance metrics including alignment precision, computational efficiency, and operational robustness. Thirdly, the algo-

收稿日期: 2024-05-22; 修回日期: 2025-04-23. 莫洋(1991—), 女, 硕士, 工程师, 主要研究方向为人工智能; 王晶(1987—), 男, 博士, 高级工程师, 论文通信作者, 主要研究方向为人工智能; 蒋海峰(1980—), 男, 硕士, 高级工程师, 主要研究方向为建筑信息模型; 张业星(1983—), 男, 硕士, 高级工程师, 主要研究方向为建筑信息模型; 曲柄宇(1991—), 男, 硕士, 工程师, 主要研究方向为建筑信息模型; 王雷(1992—), 男, 硕士, 工程师, 主要研究方向为建筑信息模型; 冯卓航(2000—), 男, 硕士, 主要研究方向为人工智能.

rithms of point cloud segmentation for both point cloud panoramic segmentation and point cloud instance segmentation are analyzed based on standardized evaluation metrics. Fourthly, the core BIM interoperability standard framework is briefly outlined, and various geometric entity and relational modeling algorithms are categorized in the automated BIM modeling module. Finally, through in-depth analysis and prospective discussion, this paper pointed out the challenges in achieving a balanced integration of efficiency, accuracy, generalization, and uniformity for large-scale outdoor scene modeling. Future work will primarily focus on multi-source data fusion modeling, the synergistic optimization of accuracy and robustness, the construction of end-to-end general Scan2BIM frameworks, and the application and exploration of large models.

Key words: Scan2BIM; deep learning; 3D point cloud; reverse modeling; large-scale outdoor scenes

建筑信息模型(building information modeling, BIM)涵盖了建筑项目从设计到拆除的全过程,在建筑、工程和施工行业得到了普遍应用。BIM 结合模型构建、数据集成和平台优势,不仅有助于改进和扩展建筑设施,还能降低运营风险并减缓各类灾害影响,为建筑设施从规划到运营的全周期带来革新。

随着建筑行业对精确度和效率的不断追求,将现有建筑的精确数字表示与新设计相结合变得尤为重要,因此,将现实世界的复杂性设施无缝转化为精确 BIM 的需求应运而生。在已建基础设施维护中,常遇到的难题包括设计图纸遗失、设计与实际施工不一致,以及建筑设施的逐渐老化。例如,在退役核电厂拆除成本和工作量的评估中^[1],由于设施信息管理不系统、设计图纸过时以及竣工图纸不完善等问题,传统的估计算法估算拆除量时误差巨大,估算工作量误差范围从 10%到超过 10 倍,拆除成本误差超过 20%。Shin 等^[1]采用自动扫描到建筑信息模型(scanning to building information modeling, Scan2BIM)对核电厂进行拆除量估计,克服传统算法的限制,提高了拆除工程成本估算和方案制定的可靠性。Scan2BIM 被广泛地应用于多种类型设施的逆向建模,基于扫描设备捕捉建筑物的 3D 点云信息后构建 BIM,精确捕捉建筑物或结构的几何形状,还原物体真实结构;该项技术大幅提升了建模效率,成为设计验证和质量控制的有力工具,能够更全面地理解设施现状,从而更好地支持修复和扩建等决策过程。

传统的 Scan2BIM^[2-3]对激光扫描数据处理的复杂性和高成本,一度限制了 BIM 在建筑行业的广泛应用。随着深度学习算法的迅速发展,基于深度学习的 Scan2BIM^[4-7]的研究引起了业界的广泛兴趣,但大多数工作聚焦于室内场景、机电和管道

等方向。近年来,室外建筑设施研究正加速向数字化与智能化方向深化发展,Scan2BIM 的应用逐步由室内场景扩展到交通基础设施等大规模复杂室外场景^[3,8-9],如道路、铁路和桥梁等。因此,室外大规模复杂场景的深度学习 Scan2BIM 逐渐成为研究热点,并得到了较为广泛的研究。

当前,已有一些与 Scan2BIM 相关的综述文献^[10-16]。Rashdi 等^[10]较完整地对比 Scan2BIM 的系统性研究进展进行综述,囊括了点云扫描、采样、配准和语义分割、BIM 开放标准等多个模块,但其仅从整体上归纳和概述相关工作,对室外场景的研究成果较少,也没有针对深度学习进行深入分析和归纳,缺乏 BIM 自动化参数提取和构建过程研究。Czerniawski 等^[11]总结了针对建筑物的 Scan2BIM 整体算法框架和流程,但是,其研究重点为建筑物,且只对 2020 年之前的自动化数字建模的研究成果进行综述,因而不足以对室外大规模场景的 Scan2BIM 的最新研究成果进行全面总结,尤其是基于深度学习的相关研究。文献^[12-13]针对 3D 场景理解和重建技术进行综述,主要关注 3D 场景理解、语义分割和点云补全,缺少对 BIM 重建的研究。还有一些工作^[14-16]只对 Scan2BIM 系统中的某一模块进行综述, Mirzaei 等^[14]的重点是 3D 点云处理算法, Monji-Azad 等^[15]关注 3D 点云配准相关研究, Salzano 等^[16]只关注 BIM 互操作。总体而言,现有的工作未系统地基于深度学习的室外大规模场景自动化 BIM 建模的最新进展进行综述。

1 本文框架

与上述综述工作不同,本文重点关注面向室外大规模场景的深度学习 Scan2BIM,从 3D 点云获取、大规模场景点云对齐、点云分割和自动化

BIM 建模的 Scan2BIM 全链条模块进行系统性阐述. 进一步, 全面地回顾和总结了基于深度学习的 Scan2BIM 各个关键模块的研究进展: (1) 按照采集设备和采集来源 2 个维度概括 3D 点云数据采集技术的最新进展, 并归纳了大规模复杂室外场景下的 3D 点云数据集成果, 包括纯采集数据集、3D 点云补充数据集和 3D 点云合成数据集; (2) 系统性梳理点云配准算法的相关工作; (3) 分析和总结点云实例分割和点云全景分割最新研究成果; (4) 从 BIM 互操作标准和 BIM 几何建模 2 个方面对自动化 BIM 建模进行归纳和总结. 本文为研究面向大规模室外场景的深度学习 Scan2BIM 提供了研究和实践基础参考.

Scan2BIM 的整体算法框架包含 3D 点云获取、

大规模场景点云对齐、点云分割、自动化 BIM 建模 4 大关键模块, 如图 1 所示.

本文框架如图 2 所示. 首先对 3D 点云数据集的通用采集技术进行综述, 并介绍大规模室外场景常用的点云数据集, 这是后续研究的基础; 然后深入探讨大规模室外场景下点云预处理和点云配准算法, 帮助理解点云数据的预处理和融合过程; 再探讨基于深度学习的 3D 实例分割算法, 对于大规模复杂室外场景中 3D 点云场景理解至关重要; 接着, 讨论自动化 BIM 建模过程中的 BIM 互操作标准和 BIM 几何建模, 分析了从点云实例中提取几何参数并构建精准的 BIM 工作; 最后指出目前基于深度学习的大规模 Scan2BIM 的局限性, 并展望了该领域未来的发展趋势.

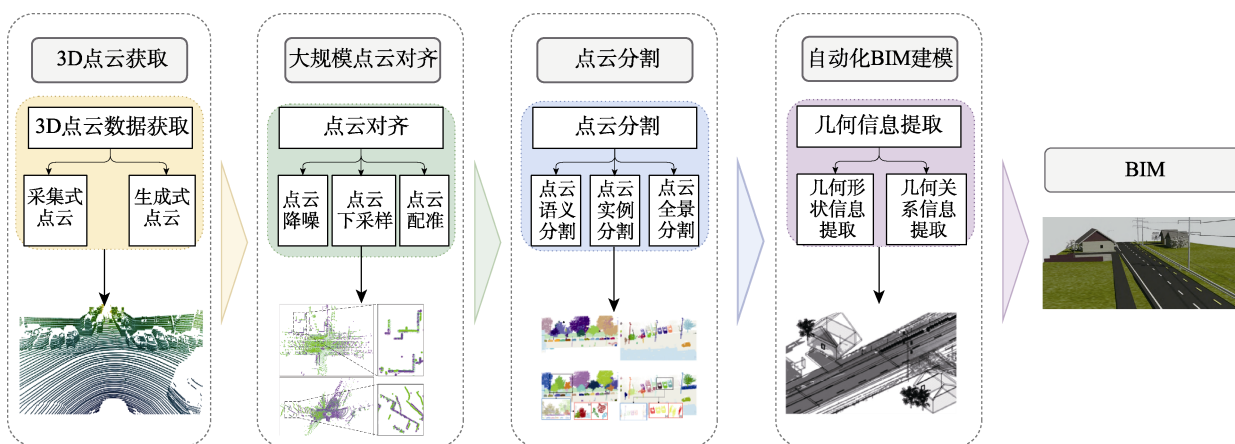


图 1 Scan2BIM 的整体算法框架^[17-20]

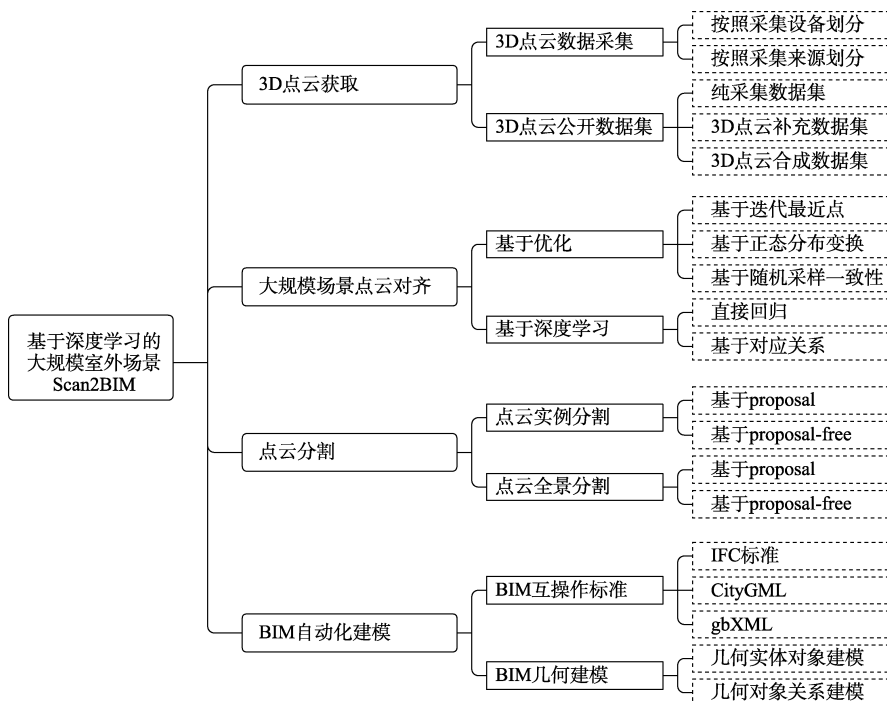


图 2 本文框架

2 3D 点云获取

3D 点云是一组由 3D 空间中的点组成的数据集,通常用于表达真实物体或者场景形状.每个点都包含其在 3D 空间中的坐标值(X, Y, Z),还可能包含额外属性信息,如颜色、强度、每个点的法线方向或全球卫星定位系统(global positioning system, GPS)等^[10].3D 点云是将现实世界映射到数字世界的重要介质,是 Scan2BIM 后续流程中必不可少的基础,为其提供强大的数据支持.

通过精密的激光扫描技术,能够捕捉并映射出物体表面的复杂几何结构,记录数以万计的点的精确空间坐标,生成 3D 点云数据,该过程涉及多样化的 3D 点云数据采集设备和采集方式.因此,首先概述 3D 点云数据采集和生成相关工作,然后整理并分析大规模场景中常见的 3D 点云公开数据集.

2.1 3D 点云数据采集

由于深度学习算法对数据集的规模和多样性要求较高,因此需要收集包含丰富多样场景的高质量点云数据集.然而,采集、处理和分析大规模室外复杂场景的 3D 点云数据面临诸多挑战,包括数据采集的难度、处理大量数据的计算负担、数据多样性和不平衡等问题.本节将对 3D 数据集采集过程中涉及的数据来源、设备和技术展开分析,帮助面向大规模室外复杂场景的 Scan2BIM 方向的研究人员选择 3D 点云数据集采集方式、采集设备等.

2.1.1 按照采集设备划分

3D 点云作为一种基础的 3D 数据表示方式,记录了真实世界中对象的空间和物理信息,能够直接从具有测距能力的传感器获取,或从立体或多视角图像生成,包括但不限于激光扫描仪、立体视觉相机、结构光扫描仪和光学时间飞行相机^[21]等,不同采集设备和技术的数据特征和应用范围有很大的差异.按照采集设备,3D 点云数据来源可以分为直接采集设备、间接采集设备和点云增强设备 3 种类型.

直接采集设备由于采集过程中自带距离或深度信息,因此可以直接生产点云数据,包括激光探测及测距系统(light detection and ranging, LiDAR)扫描仪和 RGB-D 相机 2 大类.

(1) LiDAR 扫描仪一般产出 3D 点云集合、反射强度和ación等.按照安装平台,可以将该扫描仪分为空中激光扫描仪、无人机激光扫描、地面激光扫描仪和移动激光扫描仪^[10,22];根据测量原理,该扫描仪可以分为相位差 LiDAR、时间飞行

LiDAR、多光谱 LiDAR、脉冲宽度调制 LiDAR、多普勒 LiDAR、固态 LiDAR 和闪光 LiDAR^[10,22]等.

(2) RGB-D 相机则产出 3D 点云集合、RGB 图像、深度图像和时间戳等信息.该相机通常包含至少一个彩色传感器和一个深度传感器,可以通过不同的技术实现深度测量,如结构光、时间飞行或立体匹配,且成本远低于 LiDAR 系统.

间接采集设备在采集过程中一般产出图像作为中间结果,并不包含距离信息,但可以通过后续相关软件,根据摄影测量、立体视觉等技术手段间接获取距离信息,间接生成 3D 点云数据.常见的采集设备包括光学相机和干涉合成孔径雷达(interferometric synthetic aperture radar, InSAR).

(1) 光学相机捕捉的是物体表面反射或辐射的光谱信息,产生的是人眼能够看到的图像.光学相机一般通过电光系统获取立体或多视角 2D 图像,然后借助摄影测量学或计算机视觉,间接地确定每个像素点在立体空间中的定位,生成对应的 3D 点云数据^[21].按照图像生成点云算法,可分为传统摄影测量、密集匹配、立体匹配、多视角立体视觉匹配、运动结构匹配等;按照应用场景,可以分为卫星相机、无人机相机、工业级相机、专业测绘相机等.

(2) InSAR 是一种遥感技术,用于测量地表变形和制作数字高程模型.InSAR 不直接生成点云数据,而是间接推算地形的高程信息,辅助生成点云,如层析 SAR 和持久散射体干涉测量技术^[23].

除了主流的点云生成设备,其他设备虽然不能直接用于生成点云数据,但是可以用于提高点云精度和质量、丰富属性特征等.例如,结合高光光谱成像的激光扫描能够在生态监测中辨识不同植物,而热成像技术可在建筑效能评估中揭示热损失,这类辅助技术让点云数据不仅更精确,还能提供有关环境和物体的详细特性.

在选择 Scan2BIM 的点云采集设备时,需要平衡精度、扫描范围、分辨率、速度、便携性、成本,以及数据处理能力等因素,各点云采集设备区别如表 1 所示.例如,对于高精度需求的逆向工程,可考虑使用高端激光扫描仪;大型地形或建筑则可选择精度较低的设备;扫描范围要与项目大小相适应,室内建模适合短距离扫描器,而室外或大型结构则需要远距离扫描器;对于大面积数据快速采集,如城市规划或大型基础设施建模,可考虑空中雷达、无人机雷达或移动雷达传感器;对于大范围的间接数据采集,可以使用卫星相机、无人机

表 1 点云采集设备对比

采集设备	应用场景	优势	劣势
LiDAR 扫描仪	高精度需求的逆向工程	(1) 可以在更远的距离上提供高精度的深度测量 (2) 不受光照条件影响, 具备全天候环境适应能力	(1) 设备价格昂贵 (2) 设备一般较笨重, 不易集成 (3) 只提供深度信息, 无颜色信息
RGB-D 相机	低成本的点云采集需求	(1) 价格低廉 (2) 体积较小, 易于集成	(1) 有效测距范围有限 (2) 对环境光照条件敏感
光学相机	大范围的间接数据采集	(1) 分辨率高的彩色图像 (2) 实时性强 (3) 体积较小, 易于集成	(1) 环境光照依赖 (2) 深度信息缺失 (3) 易被遮挡
InSAR	全天候的数据采集	(1) 在各种天气条件下工作 (2) 在大范围内成像 (3) 可以穿透某些材料	(1) 成本较高 (2) SAR 数据的处理和解读相对复杂 (3) SAR 提供强度和相位信息, 缺乏颜色信息

相机、测绘相机或 SAR 等。实际项目中, 通常结合多种传感器收集不同数据, 如图像、点云、GPS、惯性计量单元(inertial measurement unit, IMU)数据, 不仅效率高、信息丰富, 还能提升点云数据的质量。

2.1.2 按照采集来源划分

传统的点云数据采集方式普遍成本高、质量难保证, 如何以低成本制作高质量数据集成为一个热点研究方向。近年来, 3D 模拟器仿真和人工智能生成内容(artificial intelligence generated content, AIGC)为 3D 点云样本获取引入了突破性的解决方案。因此, 按照采集来源进行划分, 3D 点云还可以分为采集式点云和生成式点云 2 种类型。实际采集的点云虽然真实多样, 但可能存在噪声和数据不完整;

相比之下, 生成式点云非常干净、完整, 但可能缺乏真实复杂性, 可以作为采集式点云数据的补充。

大规模室外复杂场景的模拟器可以通过高级图形软件, 如 Unreal Engine 或 Unity 3D 的物理引擎构建逼真的 3D 环境, 能够大幅地降低获取点云标注数据的成本, 提供多样化、可控制的环境和天气条件, 显著地提高了数据集的采集效率和质量, 丰富样本多样性; 但可能缺乏某些真实世界数据的细节, 而且仿真环境和传感器的真实性依赖于软件的精确度。具体的仿真场景自动生成算法及相关进展可参考邓伟文等^[24]的综述, 不再赘述。表 2 总结了可用于模拟 3D 数据采集过程的国内外常用开源仿真模拟器。

表 2 国内外开源仿真模拟器总结

仿真器	引擎	开发者	开源	优势	劣势
AirSim ^[25]	Unreal Engine	Microsoft	是	开源, 支持多种传感器, 多平台, 高还原度的物理引擎和逼真的视觉场景	需要高性能硬件
LGSVL ^[26]	Unity 3D	LG Electronics Silicon Valley Lab	是	可集成多个自动驾驶框架, 丰富的应用程序接口、传感器和环境	需要高性能硬件, 入门成本高
CARLA ^[27]	Unreal Engine	Intel Labs, CVC, TRI	是	多用户支持, 传感器、环境和天气条件丰富	需要高性能硬件
Apollo ^①	Unity 3D+	百度	是	开放平台, 强大的社区和企业支持	需要与 Apollo 平台绑定
MetaDrive ^[28]	Panda3D	个人	是	python 库、轻量级、可自定义场景	相对较新、社区和文档较少
VIREs VTD ^②	VTD	VIREs Simulation-technologie GmbH	否	高度逼真的模拟, 广泛的行业接受度	商业软件且成本较高
Cognata ^③	Omniverse	Cognata	否	利用 AI 生成详细的仿真环境	商业软件, 依赖于 AI 模型准确性
NVIDIA Drive Sim ^④	自研	NVIDIA	否	高度集成的硬件和软件解决方案	专用硬件、商业软件、成本较高
Metamoto ^⑤	自研	Metamoto, Inc.	否	可扩展性、参数化、对象级仿真、支持云/本地部署	商业软件、成本较高

① <https://developer.apollo.auto/developer.html>

② <https://hexagon.com/products/virtual-test-drive>

③ <https://www.cognata.com>

④ <https://developer.nvidia.com/drive/simulation>

⑤ <https://velodynelidar.com/automated-with-velodyne/metamoto>

(1) AIGC 生成点云. 在新兴的大规模语言和视觉模型背景下, 创建各类大规模场景 3D 点云数据集面临挑战和机会, AIGC 的出现则提供了新的可能性. 近期的研究分为合成虚拟场景再生成点云数据和直接合成点云数据 2 个方向.

(2) 虚拟场景 AIGC 合成. 研究如何创建包含多种大场景元素、复杂场景标签、变化场景的逼真虚拟环境, 以低成本收集特殊情境和罕见场景的数据, 推动研究进展. Swerdlow 等^[29]提出一种基于条件生成框架的算法 BEVGen, 能从鸟瞰图(bird's-eye view, BEV)布局中合成实景街道图的模型, 该算法通过创新的视角转换和空间关注机制精确地呈现道路和车道线, 适应不同时间情况和天气情况交通场景, 实现了为 BEV 分割模型创建合成训练数据, 简化数据生成流程; Gao 等^[30]提出 MagicDrive 框架, 能合成细节丰富的高质量街景图, 并允许用户精确地操控 3D 对象位置, 同时提供无限街景图像生成的可能性; Yang 等^[31]提出了 Holodeck 系统, 借助 GPT 模型和大量 3D 对象库实现了完全自动创造与用户指令相符的多样 3D 场景. 然而, 现有的合成场景仍然局限于室内小场景, 如何扩展到大规模室外场景仍需继续探索和研究. 综上所述, 由于 AIGC 生成的仿真场景

与现实世界数据仍然存在不小的差距, 其数据质量和实用性还需进一步研究和验证.

(3) AIGC 合成点云数据. 基于有限的真实点云数据样本直接高效地合成真实多样的点云样本数据, 尤其适合实际数据采集成本极高的应用场景. 通过 AIGC 合成点云数据, 不但训练机器学习模型在数据稀缺的情况下可以补充实际数据, 还可以模拟不同的测试环境和条件, 进行模型评估和验证等. 然而, 此类算法生成的点云数据需要确保其足够逼真且具有实际应用价值, 也需要大量的计算资源和专业知识对其开发和训练.

2.2 3D 点云公开数据集

大规模复杂室外场景的 3D 点云数据已经成为计算机视觉、机器学习、自动驾驶、城市规划等多个领域研究的重要基础和数字资产. 大规模复杂室外场景 3D 点云公开数据集一方面能为研究人员提供宝贵的资源, 用以训练和测试; 另一方面, 也是推动相关领域的技术进步的基石和内驱力. 下面按照相关工作的出版年, 总结大规模复杂室外场景下常用公开数据集^[32-44], 并梳理各个数据集的基本属性、采集设备、数据分布、标注内容、支持任务类型和官方链接等, 如表 3 所示.

表 3 大规模复杂室外场景 3D 点云的公开数据集汇总

数据集	出版年	国家	采集设备	数据分布	标注类别数	支持任务类型	官方链接
KITTI ^[32]	2012	德国	移动激光扫描仪	22 个场景、4.3 万帧、约 22 h	28	3D 目标检测	https://www.cvlibs.net/datasets/kitti
Semantic3D ^[33]	2017	瑞士	地面激光扫描仪	约 40 亿个点	8	3D 点云分类、3D 语义分割	http://semantic3d.net
Argoverse ^[34]	2019	美国	移动激光扫描仪	25 万个场景(截至 2023 年)	30	3D 目标检测	https://www.argoverse.org
SemanticKITTI ^[35]	2019	德国	移动激光扫描仪	39.2 km, 22 个场景、4.3 万帧、约 22 h	28	3D 点云语义分割、3D 实例分割	http://semantic-kitti.org
nuScenes ^[36]	2020	美国和新加坡	移动激光扫描仪	1 000 个场景、39 万帧、140 万幅图像、约 55 h	23	3D 目标检测、3D 点云语义分割	https://www.nuscenes.org
Waymo ^[37]	2020	美国	移动激光扫描仪	约 3 000 个场景、60 万帧、16.7 h(截至 2023 年)	4	3D 目标检测	https://waymo.com/open
A*3D Dataset ^[38]	2020	新加坡	移动激光扫描仪	约 3.9 万帧	7	3D 目标检测	https://github.com/l2rdl2/A-STAR-3D
Toronto-3D ^[39]	2020	加拿大	移动激光扫描仪	1 km, 7 830 个点	8	3D 语义分割	https://github.com/WeikaiTan/Toronto-3D
PandaSet ^[40]	2021	美国	移动激光扫描仪	超 100 个场景、超 1.6 万帧、超 4.8 万幅图像	37	3D 目标检测、3D 语义分割	https://pandaset.org
Lyft Level 5 ^[41]	2021	美国	移动激光扫描仪	17 万个场景、约 1 180 h	9	3D 目标检测	https://www.lyft.com/
NuPlan ^[42]	2021	美国和新加坡	移动激光扫描仪	1 200 h, 1 亿帧、5 亿幅图像	7	3D 目标检测	https://www.nuscenes.org/nuplan
SmanticSTF ^[43]	2023	瑞典、丹麦、挪威	移动激光扫描仪	恶劣天气条件下 2 076 帧、超 2 万个点	24	3D 点云语义分割	https://github.com/xiaoaoran/SemanticSTF
ONCE ^[44]	2024	中国	移动激光扫描仪	200 km ² , 100 万帧, 144 h, 700 万幅图像	5	3D 目标检测	https://once-for-auto-driving.github.io/index.html

表 3 中基于大规模复杂室外场景 3D 点云的公开数据集属于纯采集数据集, 是该研究领域内使用最广泛的数据集类型. 这类数据集都是研究人员通过特定采集设备, 按照一定的采集方式获取的现实世界真实 3D 点云和其他相关数据, 并通过一定的预处理和标注方式标注; 然后整理所有数据后公开发布. 此外, 还有一些 3D 点云补充数据集和 3D 点云合成数据集.

3D 点云补充数据集通常是对一个或者多个公开数据集扩充其标注标签来适应新任务而发布的新数据集, 如 OpenScene^①, MonoScene^[45], Occ3D^[46], Occupancy-for-nuScenes^②, SurroundOcc^[47], OpenOccupancy^[48]和 SSCBench^[49]. 针对最新的 3D 语义占用预测任务, 之前的 3D 点云纯采集数据集中一般都没有该任务对应的标签. 因此, 研究人员在原始 3D 点云公开数据集之上补充了 3D 语义占用预测标签, 通常包括是否占用和占用类别这 2 类标签. 具体而言, Occupancy-for-nuScenes, OpenOccupancy 和 SurroundOcc 都是在 nuScenes 数据集的基础上补充了 3D 语义占用标签; OpenScene 则补充了 nuPlan 数据集; Occ3D 融合了 nuScenes 和 Waymo 数据集; SSCBench 对 nuScenes, KITTI 和 Waymo 数据集进行融合和补充.

通常, 3D 点云合成数据集是基于 3D 虚拟仿真模拟器进行渲染后, 通过配置的仿真传感器设备模拟采集过程而生成和标注的 3D 点云相关公开数据集. 常见的 3D 点云合成数据集包括 SynLiDAR^[50], PreSIL^[51], OPV2V^[52], SYNTHIA^[53], SHIFT^[54], UrbanScene3D^[55], HoliCity^[56], Livox^③和 51WORLD^④等.

如果公开大规模复杂室外场景 3D 点云数据集仍然不能满足需求, 一般通过自行采集 3D 数据集或直接对公开 3D 数据集进行自动化^[57]或人工标注. 常用的 3D 标注工具有 Sagemaker Ground Truth^⑤ (amazon, 商业软件), Semantic Segmentation Editor^⑥ (开源), Point Cloud Labeling Tool^⑦ (开源), Point

Cloud Annotation Tool^⑧ (开源), Multi-sensor Labeling Platform^⑨ (Segments.ai, 开源) 和 Data Engine^⑩ (scale, 商业软件) 等. 点云学习中, 在计算资源不足和训练任务要求严格的情况下, 还可以利用有限标注的算法提高点云处理的效率. Xiao 等^[58]系统地总结了 4 种当前的有限标注学习算法, 包括数据增强、域转换学习、弱监督学习和预训练基础模型.

上述大规模复杂室外场景的 3D 点云的数据集的分析 and 比较, 对于后续标准化评估和性能基准的设立具有重要价值. 通过综合评估这些数据集, 研究人员可以更好地理解现有数据的优点和局限性, 选择适合自己研究课题的数据集, 并且启发未来数据集的创建和改进. 本节综述为学术界和工业界提供了有价值的基础参考和指导意义, 帮助把握大规模 3D 点云数据集的发展趋势, 以及如何更有效地利用这些数据资源推动大规模复杂室外场景点云处理算法和应用的创新.

3 大规模场景点云对齐

在 3D 点云数据采集过程中, 需要从不同位置和角度进行多次扫描以捕捉大规模场景中的每个细节. 这些来自不同视角的 3D 点云扫描帧必须先合并, 才能用于后续现实场景物体的完整 BIM 重建, 该过程依赖于大规模场景的 3D 点云全局对齐算法. 3D 点云对齐指将 3D 点云放置于一个共同的坐标系下的过程, 确保 3D 点云没有重叠或遗漏.

大规模场景的 3D 点云全局对齐包括点云预处理、点云配准、闭环检测和全局估计等模块^[59], 在机器人和自动驾驶导航等领域, 该过程被称为构建全局一致性 3D 点云, 属于即时定位与地图构建 (simultaneous localization and mapping, SLAM) 范畴. 考虑点云配准是必不可少的核心模块, 本节集中讨论大规模复杂室外场景下 3D 点云配准算法, 回顾、分析并比较了近 5 年内基于深度学习的先进

① <https://github.com/OpenDriveLab/OpenScene>

② <https://github.com/FANG-MING/occupancy-for-nuscenes>

③ <https://www.livoxtech.com/simu-dataset>

④ <https://github.com/51WORLD/SyntheticDataset>

⑤ <https://aws.amazon.com/cn/sagemaker/data-labeling>

⑥ <https://github.com/Hitachi-Automotive-And-Industry-Lab/semantic-segmentation-editor>

⑦ https://github.com/jbehley/point_labeler

⑧ <https://github.com/springzfx/point-cloud-annotation-tool>

⑨ <https://segments.ai/data-labeling/multi-sensor>

⑩ <https://scale.com/data-engine>

研究成果,为后续研究提供参考.点云配准又称为局部姿态估计,指从不同视角或时间点采集并存在部分重叠的点云数据对齐到一个共同坐标系的过程,通过估计的变换矩阵将每帧点云都在局部范围内与前一帧精确对齐,确保点云数据的连续性和相对一致性.由于 3D 点云的稀疏、无序、密度不一致、非结构性和不完整等特殊性质,点云配准一直是一个具有挑战性的热点研究问题^[60],可分为基于优化和基于深度学习的点云匹配算法.

3.1 基于优化的点云配准算法

基于优化的点云配准算法包括粗配准和精配准 2 个环节.粗配准找到一个大致的对齐,这个过程不需要非常精确,为接下来的精配准提供了一个良好的起点.精配准从粗配准的结果出发,通过迭代的优化算法逐步提高配准的精度,直至达到某个停止准则.基于优化的点云配准算法大致分为 3 类:基于迭代最近点(iterative closest point, ICP)系列算法^[61-62]、基于正态分布变换(normalized distribution transform, NDT)系列算法^[63]和基于随机采样一致性(random sample consensus, RANSAC)系列算法^[64].ICP^[61]通过交替地将源云中的点分配给目标云中的最近邻,并通过最小化 2 个点云之间的欧几里得距离(或其他相似性度量)来不断地迭代计算出最优空间变换矩阵来改进估计的姿态;该算法存在 3 个问题:(1)经常陷入次优或局部最小值,所以需要一个好的初始值;(2)对异常值敏感,因此鲁棒性不强;(3)常需要高密度数据才能获得准确的结果,计算效率低下.NDT^[63]是一种基于局部概率密度函数的精细点云配准算法,对初始值要求较低,与 ICP 系列算法相比,该算法更稳定,精确度更高并且速度更快;缺点是也存在局部最优、计算量大和效率较低等问题,后续的变体算法对其进行了有针对性的研究.RANSAC^[64]是估计点对应关系下的粗略变换矩阵的算法,常被用来从初始对应关系中去除异常值,该算法虽然具有较高的计算效率,但由于其随机性质也无法给出全局最优解,并且计算耗时且对噪声敏感.当前,研究人员从空间兼容性角度和提升计算效率等方向提出了很多优化算法,如 4PCS 系列算法^[65]等.

近年来,深度学习取得了突破性的进展,其在点云配准任务中的应用潜力和研究价值得到了日益广泛的认可.多次扫描之间的完全自动化点云配准是困难的,基于优化的配准算法虽然也适用于大规模点云,但其效率低下.相比之下,深度学习算法在配准精度与计算效率方面呈现显著优势,并且在

实时应用中取得了良好的结果.基于深度学习的点云匹配分为直接回归算法和基于对应关系算法 2 类.直接回归算法是一种数据驱动的 AI 算法,根据 2 个输入点云的全局特征之间的差异直接回归刚性运动参数,点云通过学习软对应或无明确对应进行对齐,如 DeepCLR^[66], DCP^[67], PointNetLK^[68], IDAM^[69], DeepGMR^[70]和 FMR^[71]等.基于对应关系的算法中,首先利用深度学习网络检测关键点和强大的判别描述符来确定点云对应关系,然后应用基于学习的异常值拒绝模块进行对应关系过滤,最后使用求解器或鲁棒估计器计算变换矩阵.其中,提取正确的对应关系是基于对应关系的算法中最具挑战性的部分.如经典的 3DMatch^[72], PPFNet^[73], 3DSmoothNet^[74]和 DGR^[75]等都采用了深度学习,结合适配的网络架构或损失函数学习独特特征来估计稳健的对应关系.

3.2 基于深度学习的点云配准算法

尽管点云配准算法在室内场景中取得了显著的进展^[15,60],但这些算法的计算复杂度和全局搜索限制了它们直接应用于大规模点云场景中.室外大规模的 Scan2BIM 必须应对更多挑战,如扫描仪快速移动导致的点云位移、点云稀疏性以及大型复杂环境.下面重点对基于深度学习框架的点云配准算法的研究成果进行综述,旨在揭示关键技术难点,探讨在大规模场景下如何利用深度学习提高点云配准的精度和效率.其中,大规模场景基于深度学习的点云配准算法同样可以分为直接回归算法和基于对应关系算法.

3.2.1 直接回归算法

大规模场景下的深度学习点云配准算法普遍采用端到端的直接回归算法,在单次前向传递中同时进行特征提取和姿态估计.Lu 等^[76]提出一种 DeepVCP 算法,以端到端的方式结合了局部相似性和全局几何约束,尽管它在户外基准测试中进行了评估,但其关键点匹配仍受限于局部空间;Lee 等^[77]提出一种新的、高效和鲁棒的框架,用于大规模、真实世界的 3D 扫描的准确和稳健的配准,该算法基于深度几何上下文局部特征提取和 Hough 投票直接回归变换参数,无需集成密集的异常值过滤算法,在室外场景场景仍具有鲁棒性,但在复杂几何结构场景中易受异常值干扰;RegFormer^[18]不需要寻找关键点或者建立明确的点对应关系,只需要利用投影感知的分层变换器进行全局点特征提取,并通过捕获长距离的依赖关系识别和过滤异常值,由于其具有线性复杂度,

即使在大规模场景中也能保持高效率, 整个模型无需 RANSAC, 具有高精度和极高的效率. 直接回归算法对于 2 个点云的分布和对应关系做出了严格的假设, 在大规模复杂场景中可能无法达到期望的结果.

3.2.2 基于对应关系算法

大规模场景下, 基于对应关系的深度学习点云配准算法遵循一个由特征提取和鲁棒特征匹配组成的框架, 大致分为 3 种方案: 基于密集对密集匹配、基于稀疏对稀疏匹配和基于稀疏到密集两阶段匹配. 其中, 在 3D 点云特征提取方面, Guo 等^[78]将研究算法归纳为多视图、体素和点 3 类表示形式.

在密集对密集匹配算法方面, Yew 等^[79]提出一种利用弱监督学习 3D 特征检测器和描述子进行点云匹配的 3DFeatureNet 框架, 不需要手动编码点云数据, 而是利用对齐和注意机制, 从 GPS 标记的 3D 点云中学习特征对应关系, 但该算法容易受到外部环境噪声影响. Choy 等^[80]从稀疏张量表示中学习紧凑的全卷积几何特征 (fully-convolutional geometric features, FCGF), 并在室外点云配准任务上进行验证, 结果表明计算耗时较少. SpinNet^[81]用于提取旋转不变的局部特征, 以实现精确的 3D 点云配准, 包括空间点变换器和神经特征提取器 2 个模块, 前者将输入的局部表面映射到一个精心设计的圆柱形空间中, 后者利用强大的点和 3D 圆柱形卷积神经层提取紧凑且具有代表性的描述符, 该算法在室外数据集上表现出色, 并且在不同的传感器模态下具有很好的泛化性.

为了高效地提取特征并有效地进行刚性变换估计, 基于稀疏对稀疏匹配的点云匹配算法从大量数据中选取少量但重要的关键点, 既能有效地避免从原始点云集合中随机采样噪点和密度变化等问题, 也能保证有效完成点云匹配任务. USIP^[82]采用自监督学习算法生成虚拟关键点并提取描述符特征. D3Feat^[83]中提出一种密度不变关键点选择策略, 将空间和通道维度上具有最大分数的点视为 3D 关键点, 并通过最近邻搜索找到关键点的对应关系. Liu 等^[84]通过实验, 证明了基于 D3Feat 的关键点检测算法的大规模点云配准的有效性和稳定性. BUFFER^[85]中首先引入逐点学习器预测关键点, 并通过估计点方向提高特征的表征能力, 然后利用轻量级局部特征学习器 Patch-wise Embedder 提取高效和通用的补丁特征, 并提出一种结合了简单神经层和一般特征的 Inliers Generator 来搜索内部对应关系, 在精度、效率与泛化能力间实现了

均衡提升. Wu 等^[86]提出一种创新的局部到局部的配准网络 RORNet, 其由重叠点估计模块和表示生成模块组成, 解决了 3D 点云部分对部分配准中存在重叠点云问题时, 如何找到少量关键点作为可靠的重叠表示来进行配准, 以减少重叠估计误差对配准过程的影响; 该算法在室内室外数据集均表现良好, 噪声鲁棒性强, 能够高效地处理重叠区域估计和配准任务. Liu 等^[87]提出一种新颖的深度学习语义图匹配网络 (deep semantic graph matching, DeepSGM), 将大规模室外点云配准问题建模为语义实例匹配和配准任务, 使用语义实例匹配替换关键点匹配, 通过结合语义信息和几何信息改善点云配准效果, 可高效地实现大规模户外点云配准, 并且准确性和鲁棒性方面超越了现有算法, 但整体流程较长且复杂, 训练难度较大, 计算耗时长. 基于稀疏对稀疏匹配的点云关键点检测并不总是完美的, 有限的关键点可重复性导致某个源点丢失对应的目标点的风险较高; 此外, 关键点检测误差还可能导致 2 个匹配关键点之间的位置偏差, 降低配准精度.

基于稀疏到密集匹配两阶段算法是一种无关键点检测算法, 首先将输入点云降采样为虚拟点, 并通过检查它们的局部邻域是否重叠进行匹配; 然后通过虚拟点稀疏匹配后传播到各个点, 产生密集的点匹配关系. Yu 等^[88]提出 CoFiNet 算法, 采用粗到细的机制, 在粗略层次中, 使用一种基于重要性的权重规则匹配点云中的节点; 在细粒度层次上, 使用一种基于密度适应的匹配模块提高对应性. GeoTransformer^[89]基于几何变换器学习点云配准中的稳健粗到细的对应关系, 通过编码超点之间的成对距离和三元角度, 捕捉具有变换不变性的点云之间的几何一致性, 该算法使用一种无需 RANSAC 的方式实现了快速而准确的配准. Lu 等^[90]提出一种适用于大规模室外 LiDAR 点云匹配的高效分层网络 HRegNet, 利用邻域信息从稀疏的超点匹配传播到稠密点匹配, 实现了精确的点云配准, 然而, 由于关键点描述符中存在错误和使用点到补丁匹配策略, 会引入错误的对应关系. 之后, Lu 等^[91]又提出了改进版 HRegNet 算法, 其中引入双向共识和邻域共识进行关键点匹配, 并设计新颖的相似性特征, 结合了更深层的可靠特征和浅层中的精确位置信息, 显著地提高了点云匹配的稳健性和精确性. RDMNet^[92]通过粗到细的方式高效地获得稠密点对应关系, 并以此改善最终的姿态估计, 通过 3D-RoFormer 机制结合点云的上下文和几何

信息, 生成了稳定的超点对应关系, 并进一步利用邻域信息从稀疏的超点匹配传播到稠密点匹配, 实现了精确的点云配准; 在多个室外数据集上进行评估, 证明 RDMNet 在所有测试集上都超越了对比算法, 并展现出强大的泛化能力. Xue 等^[93]提出一种新颖的层次化神经网络 HDMNet, 通过引入双软匹配网络和特征一致性相似性, 建立了可靠的关键点对应关系, 并利用深层的对应置信度优化浅层的关键点, 通过粗到细的方式找到密集的点云对应关系, 并基于这些可靠的对应关系改进最终位姿估计; 该算法为户外 LiDAR 点云配准提供一种新的解决方案, 在保持高计算效率的同时实现了大规模户外点云配准的最先进性能.

以上大部分点云配准工作都在 KITTI 数据集上进行了算法评测, 测试结果和对比分析结果如表 4 所示, 表中配准召回率指同时满足相对旋转误差小于 5° 且相对平移误差小于 2 m 的情况下成功配准比率. 此外, 图 3 展示了点云配准相关算法的完整的发展时间线.

4 点云分割

尽管 3D 点云数据详尽地记录了物体的几何坐标, 但无法直接用于 BIM 建模, 其中最重要的是为每个点云分配精准的语义标签, 并识别出独立的结构实体, 此过程被称为点云分割, 该算法对于构建高精度和高质量的 BIM 发挥着决定性的作用. 点云分割分为语义分割、实例分割和全景分割, 其任务结果示意图如图 4 所示.

语义分割将点归类到语义标签, 如道路或树木, 以实现整个场景语义理解. 实例分割则进一步将同类物体区分为独立个体, 这对于建立每个物体的独立 BIM 实体模型尤为关键, 确保 BIM 的精确性和细节. 鉴于已有诸多研究者对结合深度学习的点云语义分割算法进行了广泛的综述, 本节不对该方向既有文献进行赘述, 只着重分析点云实例分割和点云全景分割.

4.1 点云实例分割

按照发展线路, 点云实例分割算法分为 2 种: 基于传统点云实例分割算法往往需要专家经验和复杂的参数设置, 且泛化能力差; 与此相反, 基于深度学习算法通过自动学习点云的多维特征, 不仅提高了分割精度, 还具备更好的泛化性, 因此在处理大规模复杂室外场景等复杂任务时更有效, 已被广泛地应用于点云分割任务. 为了阐述室外

场景的实例分割算法相对于当前室内场景算法研究的区别和挑战, 对室内场景的点云实例分割算法进行归纳和分析. 基于深度学习的室内 3D 点云实例分割主流算法分为 2 大类: 基于 proposal-free 的算法和基于 proposal 的算法^[94].

基于 proposal 的算法先识别目标要素的候选区域, 再通过提取各个区域特征来识别各候选区域实例标签. 该类算法通常包括 3D 对象检测、实例掩模预测和非最大抑制等处理过程; 近年来, 它在室内场景分割任务中的研究相对较多, 如 SGPN^[95], 3D-SIS^[96], PanopticFusion^[97], 3D-MPA^[98], SSTNet^[99], SPIB^[100]和 GaPro^[101]等. 虽然这些算法在室内场景点云实例分割上达到了出色的性能, 但无法直接扩展到室外大规模场景. 该类检测算法虽然具有简单直观、准确率较高的优势, 但通常需要经过大量多阶段的复杂训练过程, 并且高度依赖于密集的候选框生成, 当目标对象的边界框定位不够精确时, 算法难以准确地刻画实例对象. 此外, 整个分割预测流程较为烦琐, 不仅耗时较长且计算资源消耗大, 对硬件设备的性能要求也相对较高.

基于 proposal-free 的算法不再需要候选框, 通常将语义分割和聚簇相结合实现实例分割, 大致分为基于语义分割和聚簇两阶段级联学习算法和联合语义和实例分割多任务并行学习算法.

基于语义分割和聚簇两阶段级联学习算法首先为每个点学习一个语义判别特征或者语义判别标签, 然后通过聚类算法把同一实例的点云聚合成单独的点云簇, 以分割 3D 点云中任意形状(即对象). 以室内场景作为研究对象^[102-106]的经典算法包括 MASC^[102], JSIS3D^[103], CLOI-NET^[104], Softgroup^[105]和 DyCo3D^[106]等. 其中, 点云聚类算法可概括为 4 种类型^[107]: (1) 基于欧几里得距离的 3D 空间聚类. 首先在整个点云上构建 kD 树, 然后将所有半径阈值内的邻居点分组为一个实例. (2) 监督体素聚簇. 将点云转为超体素或超点后, 采用 k -means、层次、Mean-Shift 和 DBSCAN 等聚类算法. (3) 深度聚簇. 在距离图像上修改一次遍历的连通分量标记. (4) 扫描线运行簇. 该算法是一种基于有序点云或距离图像的逐行快速扫描算法, 是 LiDAR 距离图像上基于图像的 2 遍连通分量标记进行聚类的算法.

上述经典算法主要针对密集且均匀的室内点云设计. 然而, 由于大规模室外场景点云的非均匀密度和实例大小的变化, 中心回归无法为聚类提供

表 4 基于 KITTI 数据集的大规模点云配准算法对比

算法	精准度			计算 时间/s	鲁棒性	优缺点	应用场景
	配准 召回率/%	相对旋转 误差/(°)	相对平移 误差/m				
Reg-Former ^[18]	99.8	0.23	0.084	0.10	有异常值消除能力和错配拒绝策略	整个模型无 RANSAC, 精度高, 效率极高	市区、乡村和高速公路等室外场景; 城市环境中的自动驾驶场景
DeepHV ^[77]	99.1*	0.29	0.198	0.83	对异常值高度鲁棒	高效、准确、稳健、高精度、低内存开销	家具和房间结构等室内场景; 市区、乡村和高速公路等室外场景
3DFeat-Net ^[79]	96.0	0.57	0.259	3.40	无法在过于嘈杂的点云上进行良好的训练	鲁棒性差、对输入点云的最大值有限制	多种天气、行人和交通情况下的连续道路场景; 市区、乡村和高速公路等室外场景
FCGF ^[80]	96.6	0.30	0.095	3.40		准确、快速, 但不是端到端的	家具和房间结构等室内场景; 市区、乡村和高速公路等室外场景
SpinNet ^[81]	99.1	0.47	0.099	60.60		精度较好、泛化能力强	家具和房间结构等室内场景; 市区、乡村和高速公路等室外场景
USIP ^[82]	98.6	0.42	0.210	0.01	对噪声和下采样都具有鲁棒性	在可重复性、独特性和计算效率方面明显优于现有的检测器	物体模型; 市区、乡村和高速公路等室外场景和室内 RGB-D 扫描场景
D3Feat ^[83]	99.8	0.24	0.069	3.10	在不同误差阈值下具有鲁棒性	基于独特描述符网络实现高精度检测	家具和房间结构等室内场景; 市区、乡村和高速公路等室外场景
BUFFER ^[85]	97.7	0.22	0.054	0.30	可实现数据驱动的深度等变特征鲁棒学习	在准确性、效率和泛化方面实现了最佳权衡; 但未研究重叠估计和异常值拒绝	家具和房间结构等室内场景; 市区、乡村和高速公路等室外场景
DeepSGM ^[87]	90.7	0.23	0.066		通过探索语义实例之间的空间拓扑关系增强学习特征的鲁棒性	算法新颖, 但召回率指标较差	市区、乡村和高速公路等室外场景
CoFiNet ^[88]	99.8	0.41	0.085	1.90	通过密度自适应匹配模块, 在不规则点上表现出更强的鲁棒性	参数更少, 消除了重复采样的副作用, 并能够处理密度变化点	家具和房间结构等室内场景; 市区、乡村和高速公路等室外场景
GeoTransformer ^[89]	99.8	0.24	0.068	1.60	学习几何特征, 以实现鲁棒的超点匹配	快速、准确, 在多个场景都表现良好; 但可能存在内存占用和计算成本高的问题	家具和房间结构等室内场景; 市区、乡村和高速公路等室外场景; 常见物体的 3DCAD 模型; 起居室和办公室场景
HRegNet (2021) ^[90]	99.7	0.29	0.12	0.11	分层结构利用底层的鲁棒特征	关键点少, 具有高效性和高精度	市区、乡村和高速公路等室外场景; 城市环境中的自动驾驶场景
HRegNet (2023) ^[91]	100.0	0.147	0.047	0.14	引入空间一致性过滤, 以提高鲁棒性	精度更高, 鲁棒性更好; 但配准低重叠度的点云对时性能下降	市区、乡村和高速公路等室外场景; 城市环境中的自动驾驶场景; 高速公路、城市街道、住宅区等多种场景
RDMNet ^[92]	99.8	0.18	0.053		在所有噪声水平上都比其他算法更准确、更稳健	快速、轻量级、配准效果好, 泛化能力强; 但 Mulran 数据集上效果不佳	机器人与自动驾驶领域, 市区、乡村和高速公路等室外场景; 市中心、高速公路、校园和住宅区等多种城市和郊区环境; 高速公路、城市街道、住宅区等多种场景
HDMNet ^[93]	99.9	0.159	0.050	0.12	有效地利用最深层的鲁棒关键点, 实现全局配准	高精度、高效率	市区、乡村和高速公路等室外场景; 城市环境中的自动驾驶场景

注. *表示满足相对旋转误差小于 5°且相对平移误差小于 0.6 m.

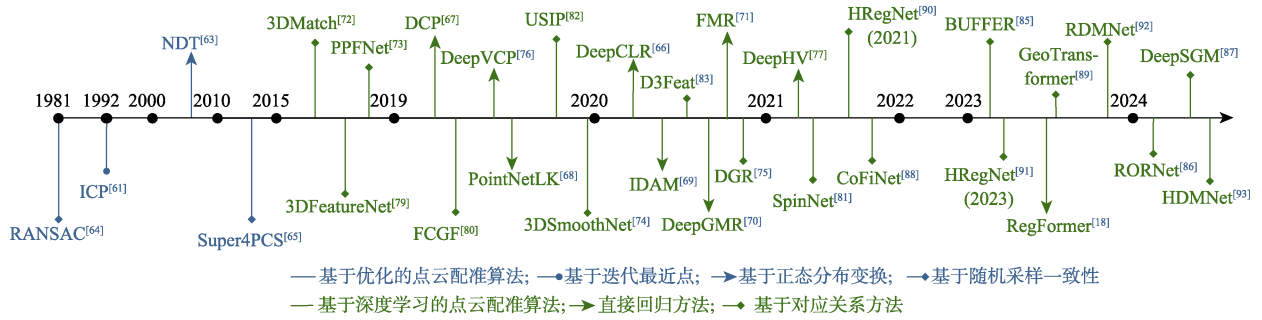
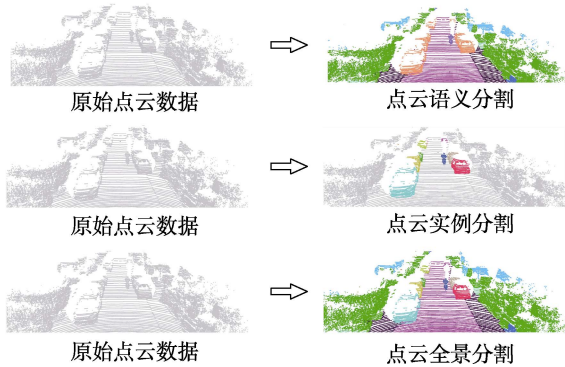


图 3 点云配准时间线图

图 4 3 种不同点云分割任务结果示意图^[94]

理想的点分布. 一方面, 回归的中心通常形成密度和大小不一的嘈杂条带分布, 导致这些广泛使用的几种启发式聚类算法很难直接扩展到室外大规模场景; 另一方面, 3D 场景空间中, 直接进行点聚类获取实例的算法, 对于紧凑、互相分离的物体(如道路上的车辆)可能效果很好, 但一旦物体靠得很近, 如紧密停放的汽车, 甚至是接触在一起的附近树冠, 往往会无法区分. 在这种情况下, 需要关于物体形状和布局的先验知识^[108], 而这些知识也适用于语义分割, 因此可以从相同的网络隐式特征中共同学习, 以减少计算开销.

为了解决上述技术挑战, 另一个研究思路就是从联合语义和实例分割任务出发, 进行多任务学习来提高点云实例分割效果. 这是由于两者存在一些相关性, 可以辅助区分实例对象和语义类别, 从而联合增强 2 个任务各自的分割性能. 针对室内场景, 该方向已取得了较多的研究成果, 如 PointGroup^[108], ASIS^[109], JSNet^[110], MTML^[111], HAIS^[112], ABJNet^[113] 和 MaskGroup^[114] 等. 然而, 这些算法通常仅适用于点云数量较少的情况, 只能够为大规模场景研究提供一定参考价值, 难以直接扩展和复用. 一方面, 对于室外场景, 边界框检测器往往只对少数几类物体有效, 而且经常漏检像路障这样的小物体以及像树木这种形状和长

宽比差异很大的物体; 另一方面, 将室外大规模点云数据分割成房间这样的自然实体是不可行的, 相反, 它们必须通过滑动窗口或随机采样进行任意的、计算上可管理的分块. 因此, 大尺寸物体常被分割为多个局部区域, 且各分割区域仅保留部分可见特征, 导致依赖全局形状特征的检测器难以有效地识别完整目标对象.

大规模的室外复杂点云结构对点云实例分割提出了特殊的挑战, 除了点云处理中常见的遮挡、尺度大和点密度等, 一个重要问题是缺乏原始处理单元, 与室内扫描可以按照房间划分处理单元不同, 室外场景无法划分为独立处理单元. 因此, 研究人员针对如何获取合适的候选集、提高计算效率、寻找合适点云表示变换和解决遮挡问题等方向进行了研究和探索. 类似地, 大规模的室外点云的实例分割算法也可分为基于 proposal 的算法和基于 proposal-free 的算法这 2 类.

(1) 基于 proposal 的算法. OSIS^[115] 是一种新颖的开放集点云实例分割网络, 可分割已知和未知类别的对象, 其整体流程包括 2 个阶段. 首先进行封闭集感知, 通过学习到的嵌入空间将点与已知事物和事物的原型相关联; 然后进行开放集感知, 将关联不确定的点分类为未知, 并通过实例感知嵌入和 3D 坐标作为特征将它们聚类为实例, 实现了在已知或未知类别中点云实例分割. 考虑专门用于 LiDAR 点云实例分割研究的公开数据集极少, Zhang 等^[116] 构建了一个包含精确的 3D 边界框和点云实例分割标签的全新大规模公开数据集, 可以用于训练鲁棒的实例分割模型; 并提出专门为大规模户外 LiDAR 点云实例分割设计的鲁棒基线算法 LiDARseg, 包括一个用于不同层次特征的密集连接堆叠骨干结构, 以及一种无需区域提议网络的实例预测头, 通过预测物体的水平中心和高度限制对点云进行分组, 进一步移除离群点. Shin 等^[117] 提出一种 Spherical mask 算法, 结合球坐标系下的粗略检测

和细粒度优化, 能够更精确地描述实例形状和大小, 提高分割精度; 通过学习每个点的径向偏移量纠正误分类, 减少误差积累, 最终生成精确的实例掩模.

(2) 基于 proposal-free 的算法. Wu 等^[118]提出一种基于卷积神经网络(convolutional neural network, CNN)的 SqueezeSeg 的端到端处理流程, 首先通过 CNN 架构对预处理后的点云数据实施端到端的标签预测, 然后通过基于循环层实现的条件随机场进行细化; 根据精细的标签划分不同语义类别掩模后, 采用传统聚类算法在 XYZ 坐标上获得不同类别的实例级标签. Chen 等^[119]利用不同 LiDAR 光束距离分布不均匀的特点, 提出一种距离感知实例分割网络 RangeSeg, 其重型解码器只计算远、小目标所在的距离图像的顶部, 提高小目标的检测精度; 轻型解码器计算整个距离图像, 降低计算成本; 最后利用分辨率加权距离函数对结果进行

DBSCAN 聚类, 得到实例级分割结果. 该算法不仅显著地提高计算速度, 而且在小对象和远对象上的实例级分割获得了优越的性能. Du 等^[120]提出一种基于数据融合的城市环境实例分割模型 SMS, 它集成了分类、优化和分割任务; 该算法构建了一种轻量级、鲁棒的城市环境实例分割模型, 通过整合不同传感器收集的信息, 提高模型的学习能力和资源效率, 不仅能够提高分类准确率, 还能在不需要实例标签的情况下完成不同个体的分组. Zhang 等^[121]提出 FreePoint, 通过无监督算法减少标注依赖, 降低成本, 结合几何、颜色和自监督深度特征提高伪标签质量, 并通过 id-as-feature 策略生成高质量伪标签. FreePoint 中设计了 2 步训练策略及相应的损失函数, 克服粗略掩模不准确性, 有效地提升模型的性能. 图 5 所示为点云实例分割相关算法完整的发展时间线.

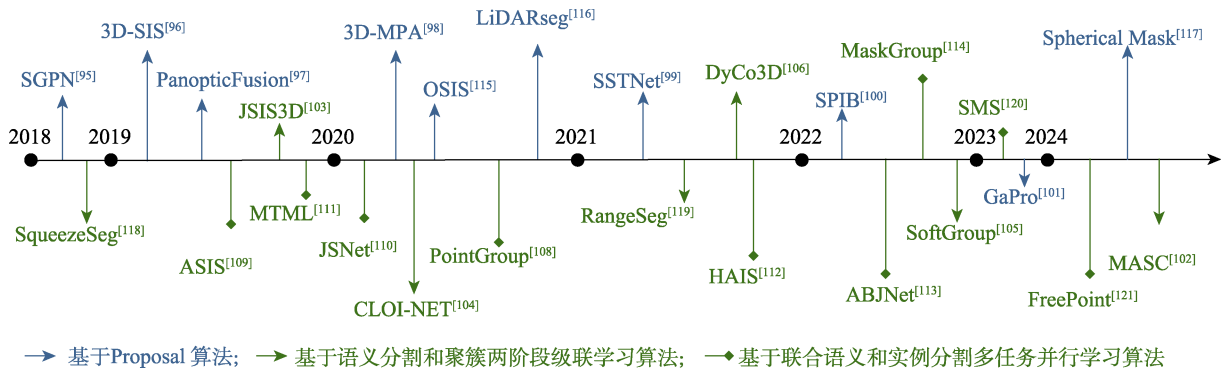


图 5 点云实例分割时间线图

4.2 点云全景分割

2019 年, Kirillov 等^[122]提出关于全景分割的开创性工作之后, 室外大规模复杂室外场景下点云实例分割研究集中在 3D 点云全景分割方向. 尤其是伴随着自动驾驶技术的迅猛发展, 面向大规模复杂室外场景的点云实例分割算法取得了显著进展. 点云全景分割融合语义和实例分割, 具有更为细致和全面的场景理解能力, 该任务既需要预测背景(物质)类别(如道路、天空等不可计数对象)的语义标签, 还需要对前景(物体)类别(如汽车、建筑物和路灯等可计数对象)进行实例分割, 预测语义标签和实例 ID; 而点云实例分割一般只对特定独立的前景类别进行实例提取^[122]. 类似地, 基于深度学习的大规模室外场景的点云全景分割分为基于 proposal 和基于 proposal-free 的算法, 其主要区别如图 6 所示, 前者通常需要更复杂和更慢的非极大值抑制(non-maximum suppression, NMS)或

聚簇等后处理过程, 后者一般直接端到端地输出语义类别和实例信息.

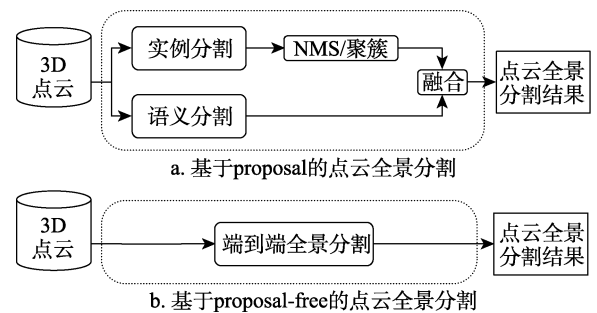


图 6 基于 proposal 和 proposal-free 的算法框架比较^[123]

基于 proposal 的点云全景分割算法. 为了填补点云全景分割公开标注数据的空白, Behley 等^[124]基于 KITTI 数据集补充了点云全景分割训练和评估所需的密集逐点语义和实例信息标签, 并且提出 2 种基线算法, 即表现优异的语义分割算法

KPConv^[125] 和 RangeNet++^[126] 与物体检测器 PointPillars(P.P)^[127] 的组合, 用于预测实例级别信息. 然而, 这些算法需要将检测结果转换为点级实例标签, 后处理模块在不同分支间的结果融合中需要解决冲突, 降低了分割性能又需要额外的计算资源, 以及可能导致过分割问题; 相比之下, 基于 proposal-free 算法可以通过预测各实例之间的相对位置避免这些问题. 为了解决包括距离依赖的稀疏性、严重的遮挡、大的规模变化和重新投影错误在内的多重挑战, Sirohi 等^[17] 提出一个自顶向下的高效点云全景分割架构 EfficientLPS, 构建一个全新基于范围图像的共享主干网络, 以强化几何变换建模能力, 并聚合了丰富的语义、多尺度的范围感知特征; 该框架不但整合尺度不变的语义和实例分割头, 以及一个由全景边缘损失函数监督的全景融合模块, 还构建了一个规范化的伪标记框架, 通过在未标记的数据上进行训练进一步提高性能. Chen 等^[128] 提出融合 3D 点云目标检测、语义分割和全景分割的多任务流式学习框架 PolarStream, 通过极坐标柱更紧凑地表达点云扇区, 解决了现有流式算法浪费内存和计算效率低的问题. Li 等^[94] 提出一个实时高精度的点云全景分割框架 Panoptic-PHNet, 融合细粒度体素特征和具有不同感受野的 2D BEV 特征, 引入一个 K -近邻(K -nearest neighbor, KNN)变换器进行特征增强, 并设计聚类伪热图直接产生实例中心进行分组, 获得了较高的分割精度.

基于 proposal-free 的点云全景分割算法又可以分为直接法和间接法. 直接法是一种不需要转换形式, 直接将点云输入到网络中实现分割的算法. Gasperini 等^[129] 提出一种无需候选框的点云全景分割算法 Panoster, 通过在语义分割网络中添加一个实例分支实现全景分割能力, 该实例分支通过学习基础的聚类算法来解析对象, 直接从网络中输出实例 ID, 实现了端到端的全景分割, 无需后处理步骤, 其结构灵活、开销小、性能优异. GP-S3Net^[130] 是一种基于图神经网络的全景分割框架, 通过图网络将传统的无监督聚类问题转化为一个有监督的图边分类问题, 可以从语义分割结果中直接获得实例; 该算法在公开数据集上取得了当时最先进效果, 尤其是对道路上较小的动态物体分割. Shi 等^[131] 引入联邦学习, 提出一种结合全景卷积融合网络(panoptic convolutional fusion network, PCF-Net)和动态分组联邦学习(dynamic grouping federated learning, DGFL)算法的点云全景分割框架, 其中,

PCF-Net 通过一个可学习的卷积模块直接输出实例级的全景预测; DGFL 算法在全局平均之前基于分组进行参数传递, 在保证通信效率的同时使模型参数的共享次数最大化. Marcuzzi 等^[132] 提出基于掩模的 3D 点云全景分割算法 MaskPLS, 使用基于 3D 稀疏卷积的主干网络来提取多尺度特征, 并通过掩模注意力结合可学习的查询和不同分辨率的特征, 以及直接预测一组非重叠二进制掩模及其语义类别来完成全景分割; 该算法支持端到端学习, 不依赖启发式后处理, 并且在处理不同传感器数据集时无需更改任何超参数, 鲁棒性好、性能优越. Wang 等^[133] 提出一种基于高斯混合模型的算法 GMM-PanopticSeg, 通过显式建模实例内的方差改进 3D 点云的全景分割性能, 能够提高点嵌入的区分能力, 并增强对遮挡和噪声的鲁棒性.

间接法是一种将不规则点云转化为规则结构, 并将其输入到网络中进行分割的算法, 可以很好地弥补 CNN 不能应用在点云的缺陷, 但在转化阶段中算法可能会损失某些关键特征; 它包括基于投影、基于体素和基于多表示融合的算法.

(1) 基于投影的算法. Milioto 等^[134] 提出一种基于 RangeNet 的全景分割算法, 结合共享编码器和语义/实例解码器进行语义和实例分割并行的多任务学习, 融合几何信息、语义嵌入, 以及新颖的基于范围图像三线性上采样模块, 显著地提高了处理速度和全景分割质量. Panoptic-PolarNet^[135] 是一种高效的全景分割框架, 使用极坐标 BEV 表示点云数据, 并通过单一的编解码器网络生成语义预测、中心热图和偏移回归; 通过投票机制集成中间结果, 生成全景分割结果; 还采用适应性实例增强和对抗性点云剪枝算法提高网络的学习能力: 该框架不仅能够有效地解决城市街景中实例之间的遮挡问题, 还能在不牺牲准确性的情况下实现实时或接近实时的处理速度. SMAC-Seg^[136] 是一种基于范围图像的端到端的实时全景分割系统, 为了恢复投影过程中丢失的信息, 提出一种新颖的带有交叉局部空间注意力(cross local space attention, CLSA)的卷积; 考虑其相对于邻近对象的相对位置, 引入新的质心感知排斥损失来有效地监督网络区分每个对象聚类与其邻近对象; 还增加了一个可学习的稀疏多方向注意力聚类(sparse multi-directional attention clustering, SMAC)模块, 动态地聚合多个方向上的点, 将点重新定位到它们的中心, 加速聚类过程. 为了解决点云全景分割模型在不同领域数据集巨大领域差距带来的性能下

降, Besic 等^[137]提出一种用于点云全景分割的无监督领域自适应(unsupervised domain adaptation, UDA)算法, 采用 2 种互补的领域自适应策略解决 UDA 任务, 即基于数据的和基于模型的自适应策略. 基于数据的自适应通过处理原始 LiDAR 扫描减小领域差距, 使其类似于目标领域的扫描, 而基于模型的指导网络提取对 2 个领域都有代表性的特征, 最终实现了准确的全景分割输出和有效的泛化性. 为了解决大多数点云分割算法都忽略了 3D 点云中物体的先验分布信息的问题, Xian 等^[138]提出一种基于位置引导的 LiDAR 全景分割算法, 采用 BEV 表示法将 3D 点云投影到 2D 图像中, 首先通过全局特征提取(global feature mining, GFM)得到全局特征, 再利用位置引导特征提取(location guided feature mining, LGFM)将 BEV 投影分为 8 个区域, 并对每个区域输入依赖的卷积核来提取局部特征, 最后将全局特征和局部特征融合后进行实例分割. CFNet^[139]是一种基于 BEV 视图表示的中心聚焦网络, 采用中心聚焦特征编码(center focusing feature encoding, CFFE)算法, 通过平移 LiDAR 点并填充中心点, 明确地理解原始 LiDAR 点与虚拟实例中心之间的关系; 还提出一种快速的中心去重模块(center deduplication module, CDM), 通过选择每个实例的单一中心来利用冗余检测到的中心, 使得到的实例中心更加准确. 实验结果证明, 该算法性能出色、计算效率高.

(2) 基于体素的算法. Zhao 等^[107]提出将语义分割和聚类级联以实现点云全景分割, 包括一个基于语义分割网络 Cylinder3D^[140]先提取语义类别掩模, 再使用传统点云聚类算法来分割每个实例对象; 并通过 4 种传统聚簇算法的消融实验, 证明了结合语义分割和聚类算法的级联框架的有效性. 之后, 为了提高聚簇阶段算法鲁棒性, Zhao 等^[141]又提出一种新的分割和合并的点云聚类算法, 首先在每个均匀划分的局部区域中进行聚类, 然后通过对边缘点对进行投票来合并局部聚类的小组件, 进一步提高了点云全景分割算法的性能. Chen 等^[142]提出一个基于柱状(pillar-based)视图表示法的点云全景分割算法, 并通过学习柱级亲和力(pillar-level affinity)确定柱是否属于同一个实例, 联合优化语义分割和类别无关的实例分类提出一种局部聚类算法获取实例 ID, 提高了分割的准确性. Hong 等^[143]提出一种基于动态偏移网络(dynamic shifting network, DS-Net), 以统一的方式解析对象和场景, 首先采用圆柱卷积高效地从点云中提取

网格级特征, 然后由语义和实例分支共享, 最后基于动态偏移模块对实例分支产生的复杂分布的回归中心进行聚类; 其中, 动态偏移模块鲁棒性强, 对参数设置不敏感, 且计算开销小. 为了应对大型物体的长距离依赖性建模和紧密分布的小物体分割的挑战, Xu 等^[144]提出一种稀疏跨尺度注意力网络(sparse cross-scale attention network, SCAN), 通过全局体素编码的注意力网络对齐多尺度稀疏特征, 捕捉实例上下文的长距离关系, 提高大型物体过分割问题的回归精度; 并采用一种新颖的稀疏类不可知的实例中心表示, 不仅保持对齐特征的稀疏性, 解决了小物体的欠分割问题, 还通过稀疏卷积减少了网络的计算量. Li 等^[145]提出一种无聚类全景分割(cluster-free panoptic segmentation, CPSeg)网络, 通过共享编码器、双解码器和无聚类实例分割头, 将前景点按照学习到的嵌入特征动态地分组为柱状体, 然后通过比较成对嵌入特征找到连接的柱状体, 将传统的基于提案或聚类的实例分割转化为成对嵌入比较矩阵上的二元分割问题, 在无需大量后处理的情况下取得了优异的分割性能.

(3) 基于多表示融合的算法. Su 等^[123]提出一个简单而有效的端到端点云统一全景分割(point cloud unified panorama segmentation, PUPS)框架, 通过基于范围-点-体素融合网络 RPVNet 提取特征, 在训练中引入二分匹配, 确保分类器能够以排他性方式预测实例分组, 避免了手工设计和后处理, 如锚点和 NMS. 此外, PUPS 还利用变换器解码器迭代地细化点分类器, 并引入一种感知上下文的 CutMix 增强策略克服类别不平衡问题. Panoptic-FusionNet^[146]是一种基于相机和 LiDAR 融合的点云全景分割网络, 通过融合来自相机的密集像素特征和来自 LiDAR 的稀疏点云特征, 提高全景分割任务的性能; 该算法能够克服单一 LiDAR 传感器存在的点稀疏性和缺乏纹理信息的问题.

基于 proposal-free 的算法中, 估计实例中心点算法仍然存在挑战, 如对于部分遮挡物体的点云数据, LiDAR 扫描遗漏的单个点可能导致实例中心的重大变化. 该类算法没有明确的对象候选框, 因此聚类精度相对较低, 而且复杂的后处理也导致时间和计算资源成本高. 尽管如此, 未来点云全景分割算法仍趋向于朝着无需提案和无需聚类的方向发展.

图 7 所示为点云全景分割相关算法的完整的发展时间线. 更进一步, 在 SematicKITTI 数据集上, 将基于 proposal 和基于 proposal-free 的全景分

割算法通过统一的评估指标进行整理和归纳, 包括全景质量 (panoptic quality, PQ), 平均交并比

(mean intersection over union, MIoU) 这 2 个指标, 结果如表 5 所示. 其中, PQ 值由低到高排列.

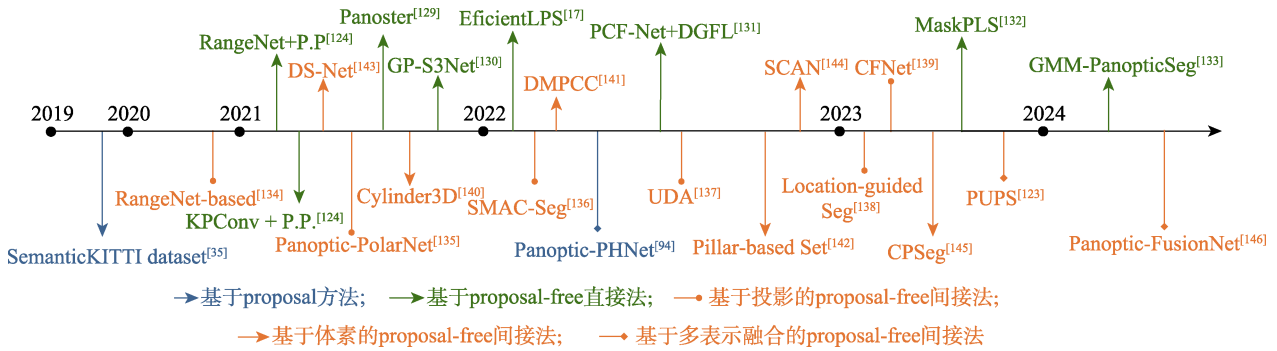


表 5 在 Semantic KITTI 数据集上点云全景分割算法对比

算法	出版年	PQ/%	MIoU/%	类别		
				proposal-free	间接法	表示方式
RangeNet+P.P. ^[124]	2021	37.1	52.4	×		
Panoptic RangeNet ^[134]	2020	38.0	50.9	√	√	投影
KPConv+P.P. ^[124]	2021	44.5	58.8	×		
Panoster ^[129]	2021	52.7	59.9	√	×	
Panoptic-PolarNet ^[135]	2021	54.1	59.5	√	√	投影
DS-Net ^[143]	2021	55.9	61.6	√	√	体素
SMAC-Seg ^[136]	2022	56.2	63.3	√	√	投影
DMPCC ^[141]	2022	56.5	68.2	√	√	体素
CPSeg ^[145]	2023	56.9	62.6	√	√	体素
EfficientLPS ^[17]	2022	57.4	61.4	×		
Mask-PLS ^[132]	2023	58.2	62.5	√	×	
GP-S3Net ^[130]	2021	60.0	70.8	√	×	
Panoptic-PHNet ^[94]	2022	61.5	66.0	×	√	多表示融合
SCAN ^[144]	2022	61.5	67.7	√	√	体素
CFNet ^[139]	2023	63.4	68.3	√	√	投影
PUPS ^[123]	2023	65.7	68.5	√	√	多表示融合
GMM-PanopticSeg ^[133]	2024	60.3	64.2	√	×	
Panoptic-FusionNet ^[146]	2024	64.3	65.4	√	√	多表示融合

4.3 小结

点云分割算法能够实现整个场景语义理解和分割, 对于后续自动化构建高精度和高质量的 BIM 发挥着关键作用, 决定了 BIM 的精确性和细节. 然而, 由于 3D 扫描传感器限制、随机误差和遮挡等情况下常存在点云不完整的问题, 场景关键对象缺损, 进而显著地影响最终 BIM 的重建效果. 当前, 许多工作探讨了如何通过点云补全算法或者点云生成算法^[147-151]有效地缓解该问题, 提高重建元素的结构完整性和几何精度. 如 Mahmoud 等^[152]对建筑物内非结构化物体进行 Scan2BIM 重建, 经过点云语义分割和实例聚簇后, 采用 Snowflake-Net 点云

生成算法增强不完整的点云簇, 提取到了更完整且精确的对象.

5 自动化 BIM 建模

Scan2BIM 的关键步骤是: 首先将扫描点云数据经过全局点云对齐和点云实例分割; 然后通过几何建模和关系建模分别提取几何形状参数和关系, 将其导入 BIM 软件中完成自动化 BIM 建模过程; 最后依据 BIM 互操作标准写入文件. 其中, 几何建模中, 为每个设施构建几何形状、位置和方向等几何信息, 而关系建模将构建设施之间的关系^[153]. 下

面首先概述当前 BIM 互操作性标准的研究进展,特别是大规模复杂室外场景相关标准的发展;然后综合评述 BIM 建模方面的研究进展和成果,包括 BIM 几何建模和关系建模。

5.1 BIM 互操作标准

BIM 生产流程可能涉及不同人员交互、多个软件应用程序或服务数据交换等过程,双方可能需要不同格式或类型信息,而交互过程中极易发生数据丢失和沟通问题,导致项目效率低下,而 BIM 标准可以解决异构数据交换问题。面向大规模复杂室外场景的 Scan2BIM 所有模块中, BIM 标准至关重要,它决定了最终的 BIM 产品输出效果。

BIM 标准就是一种标准化可交换数据格式,允许以最小数据损失存储和交换信息^[10],促进不同利益相关者之间的知识共享和互操作性,帮助项目节约成本并提高生产力。基于标准编码的 BIM 文件定义包含语义注释、空间位置以及几何化信息的对象元素,如道路和路灯等,并明确定义这些元素之间的排列关系(包括聚合关系、拓扑关系和方向关系)^[11]等。如在道路建模中,各个交通基础设施资产的位置与道路对齐相关,通常对应于道路中心,这个扩展允许有效地数字化交通运输领域项目。

BIM 标准涉及各种组织使用交换协议和原则来开发一个共同框架。开放地理空间联盟(open geospatial consortium, OGC)和 buildingSMART 已经开发了多个开放的 BIM 标准,如工业建筑开放标准(industry foundation classes, IFC), CityGML, gbXML 和 LandXML 等^[10]。其中, IFC 是一个最广泛接受的开放和通用的 BIM 数据格式,提供了对 BIM 元素及其关联关系、属性和描述性信息的严格和权威的语义定义,用于在异构系统之间交换和共享 BIM 对象信息。

Costin 等^[154]通过文献回顾,分析了室外基础设施在 BIM 应用中的互操作性问题,指出 BIM 信息交换标准的紧迫性。IFC 从 4.1 版本开始涵盖铁路、公路、港口和水路工程内基础设施建设的常见元素,直到 IFC4.3 RC1(4.3.rc.1)版本基本修订完成 IFC[®]已经提供了一整套交通基础设施领域的标准化描述方式,包括交通基础设施的构件(如人行道、路面),特性(如类型、功能、形态、材料),关系(如支撑和连接),以及计划、活动、空间和成本

等概念。当前,研究者一直致力于探索如何更加出色地设计室外基础设施的 IFC 对象实体。Cho 等^[155]为道路排水开发 IFC,定义了表示道路系统所需的元素、对象和关系,持续扩展了 BIM 在公路设计和网络管理中的协同作用;Borrmann 等^[156]对构建多尺度隧道模型,以及如何扩充 IFC 标准以适应隧道工程的需求进行研究,引入类似城市地理标记语言 CityGML 的多层次地理信息系统(geographic information system, GIS)表示方式,实现了模型在不同细节层级间的自动同步更新,即在较高层次所做的改动能够自动地反映到更详细的层次。这项初步探索为 IFC 应用到隧道工程领域迈出了重要的一步,并为其在 GIS 应用中的扩展提供了一定的基础。

5.2 BIM 建模

BIM 对设施进行整体编码和表示,由于不可能重建现实对象的每个细节,因此 BIM 建模过程通常只提取对目标设施至关重要的几何和关系这 2 类信息,而忽略其他信息。下面归纳 BIM 几何建模和关系建模的相关工作。

5.2.1 BIM 几何建模

几何建模是通过点云数据构建设施实例的 3D 形状简化表示的过程。在几何建模之前需要先定义几何表示形式,其决定了 BIM 建模输出表示形式。

常见的几何形状表示的分类维度包括显式与隐式、参数化与非参数化^[153]。显式表示通常指直接编码对象的形状,可以准确地描述用于构建 BIM 所需的几何形状;隐式表示通过中间表示间接编码对象形状,并从数据或预先定义的 BIM 形状模型库中提取特征来编码形状。例如,三角网格是一种显式表达形式,而表面法线的直方图可隐式地表达对象形状。参数化表示使用具有少量参数的模型来描述形状,包括直线、平面、曲面、矩形和立方体等线性几何形状;非参数化表示本身没有参数,通常以三角形为主的多边形网格或体积的表面形式呈现。

特别地,上述 2 类分类维度是非独立可组合的,显式表示可进一步细分为参数化和非参数化显示表示,但隐式表示大多数是非参数化。显示表示适合用于建模 3D 对象,而隐式表示最常用于 3D 对象识别和分类^[155]。因此在 BIM 几何建模过程中,通常使用显式几何形状表示作为建模表示形

① <https://technical.buildingsmart.org/standards/ifc/ifc-schema-specifications/ifc-release-notes>

式,大体上分为基于表面的和基于体素的.边界表示(boundary representation, B-rep)^[157]是常见的基于表面的表示,它通过定义一组几何表面元素(如点、线、面)及其对象边界表示 3D 实体,还包括具体的几何定义,如面的形状、边的曲线和顶点的位置,并使用对应连接信息描述元素之间的拓扑关系.例如,非均匀有理 B 样条(non-uniform rational B-splines, NURBS)^[158]是常见的 B-Rep 算法,其具备在表示、数学推导和描述从复杂到简单的一系列原始几何的能力;常见的三角网格也是一种 B-rep,网格的面是表面元素,面的相邻关系对拓扑关系进行编码.表面建模在准确地表示复杂几何时倾向于提供更高的细节度.

构造实体几何(constructive solid geometry, CSG)^[159]是常见的基于体素表示算法,它通过实体几何描述形状,通常基于布尔运算操作组合,如立方体、圆柱体和球体等基本几何原形构建复杂形状;实体体积建模倾向于具有更简单的数据结构;但两者之间也可能存在重叠,如一个球体既可以被视为表面表示,也可以视为体素表示.

为了建模 3D 物体几何形状,常见的算法是利用基本形状的闭合形式描述,提供一个目标函数评估基本形状与点的接近程度.已有多种自动及半自动几何建模算法,通常包括基于模板拟合算法和直接拟合算法 2 大类.

(1) 基于模板拟合算法.在 BIM 模板库的基础上,通过元启发式优化算法参数化并自动调整先验 BIM 的形状和姿态参数,以适应 3D 点云实例对象.如果调整后的先验参数化 BIM 成功地与场景中的实例对象匹配,即可获得该对象的全参数化 BIM^[160-162].该类优化算法通常是基于规则的,经典算法包括粒子群优化^[163],遗传算法(genetic algorithm, GA)^[164],模拟退火(simulated annealing, SA)^[165]等. Rausch 等^[162]采用 GA 和 SA 这 2 种优化算法,分别对现浇混凝土底座的 CSG 和 NURBS 这 2 种显式参数化表示的 BIM 模板拟合到点云过程进行实验,结果表明, NURBS-GA 为最优结果.其中,基于 CSG 几何形状表示的参数包括长度、宽度、深度、X 轴位置、Y 轴位置、Z 轴位置,围绕 Z 轴旋转;基于 NURBS 的几何体表示无需明确参数化,只考虑矩形柱状几何体的 8 个角控制点作为参数,用于建立对象表面的控制点来变形;最后,将 CSG 和 NURBS 参数分别输入 Revit[®]软件中即可完成参数化几何建模,生成满足需求的 BIM.

(2) 直接拟合算法.其是一种由数据驱动的算

法,通常基于传统拟合算法直接拟合或检测实例点云的基础几何元素,如点、线、平面、曲面、球体、圆柱体以及立方体等;提取对应的关键几何参数后,输入 BIM 设计软件中即可完成 BIM 参数化重建^[20,166-168].线拟合、平面拟合、一次曲面以及二次曲面等不同类型的几何元素使用的拟合算法不同,常见的算法包括 RANSAC、霍夫变换和最小二乘优化算法、直线段检测算法、区域生长算法、alpha-shape 算法等^[169-172]. Mahmoud 等^[167]首先对建筑物扫描点云进行分割和聚类,划分得到房间和各类物体的单体化点云对象;然后提取对象边界框,获得位置、高度、长度和宽度参数,结合随机抽样和 RANSAC 进行平面检测,提取房间边界线段,基于线特征检测各墙体轮廓线并推导出房间角点,并采用点云平面投影和 alpha-shape 算法提取轮廓点等;最后综合以上参数和几何信息,利用 Revit[®]和 Dynamo 工具协助自动生成 BIM.

5.2.2 BIM 关系建模

常见 BIM 相关的几何对象关系包括聚合关系(如门包含在墙中)、拓扑关系(如墙连接到屋顶),以及方向关系(如墙垂直于地板),这些关系通常以树形或图形结构表示.几何对象关系建模算法分为全局优化算法和局部启发式算法 2 类^[173].全局优化算法通常是基于模型的,将场景作为一个整体进行推理,试图找到在最大后验估计或能量最小化方面最优的全局几何解释.与这种可能在模型上非常复杂且计算代价高昂的整体算法相对立,局部启发式算法使用自下而上的推理,场景中的元素被独立处理,并可能使用预定义的元素之间关系相关先验信息作为潜在错误过滤标准.

6 总结与展望

本文以文献综述的方式,对大规模室外基础设施领域的深度学习 Scan2BIM 进行了全面的梳理、深度的解析和系统的总结.从深度学习在室外大规模基础设施领域应用出发,介绍了 3D 点云获取,大规模场景点云对齐、点云分割和自动化 BIM 建模 4 个关键模块.在每个模块中,不仅概述具体的技术原理和算法流程,还对比各自的优缺点,揭示了它们在实际应用中的效果和局限,以及目前面临的挑战和未来的改进方向.通过这种全面而系统的文献综述,为深度学习在大规模室外场景建模的 Scan2BIM 算法研究和应用提供了完整、全面和清晰的理论框架和技术途径.

随着 3D 数据采集技术的发展, 在实际应用中, 如何从原始 3D 数据中重建几何和语义丰富的 BIM 成为热点研究方向. 得益于 AI 的不断演进, Scan2BIM 在大规模室外场景建模已经取得了显著的进步. (1) 得益于各类型传感器以及 3D 合成算法的发展, 采集方式和来源的多样化使得 3D 点云数据获取更加高效和便捷, 采集样本和虚拟样本在数量和质量上均提升显著; (2) 在深度学习算法迅猛发展的背景下, 大规模复杂室外场景下的点云配准和点云分割算法取得了较多突破性进展, 不仅显著地提高处理效率和配准精度, 而且大幅增强算法的鲁棒性, 实现了算法综合性能的全面提升; (3) 在自动化建模方向, BIM 互操作标准逐步扩展到道路、桥梁和铁路等大规模复杂室外场景, BIM 几何对象建模和关系建模算法在不断地探索降低计算代价的同时, 持续提升精度和细节表现. 目前, 该研究领域仍然面临着一系列问题和挑战, 未来, 对应的大规模 Scan2BIM 发展趋势和潜力集中在多源数据融合建模, 精度、鲁棒性与性能的协同优化, 构建端到端 Scan2BIM 通用框架和大模型应用与探索 4 个方向.

6.1 多源数据融合建模

Scan2BIM 的最终目标是尽可能真实地还原物体的几何信息、空间关系和物理属性等信息, 如物体的位置、几何、关系、语义、表面结构和材质等精细表达, 以增强 BIM 的真实性. 3D 点云数据能够精确地表达立体空间信息, 非常适用于精确测量和建模场景, 支持 Scan2BIM 对物体的空间位置、几何形状、尺寸和空间关系等建模需求. 但是, 单一点云数据难以捕捉所有细节, 而多源数据融合可以提供更全面的视角, 不同的数据可能在不同类型特征上具有优势, 融合这些数据可以相互增强, 互补彼此的不足. 因此, 融合多源数据, 如光学、图像视频、GPS 和 IMU 数据等, 是未来 Scan2BIM 的重点发展趋势之一, 以实现更精准和趋向于真实世界的 3D 重建模型.

多模态融合属于多源融合的范畴, 通过对齐和融合图像和点云这 2 种不同模态, 能够有效地增强三维场景语义与几何特征的完整性和丰富度, 有助于减少算法对单一数据源可能存在的信息缺失或噪声的依赖, 提高算法的鲁棒性. 因此, 多模态点云配准和点云分割算法是当前的研究热点之一. 然而, 多模态数据融合过程也带来了更优模型学习结构和更高计算能力的挑战, 需进一步探索多模态 Scan2BIM 的建模效能优化路径.

除此之外, GPS 能够提供绝对位置信息, 可用于确定物体位置信息和空间关系; IMU 通过集成陀螺仪与加速度计等多传感器数据, 实现对运动姿态的动态捕捉. 当前, GPS 与 IMU 的融合策略已被广泛应用于 SLAM 系统, 可辅助确定物体位置和运动轨迹. 未来, 可继续探索如何在大规模室外场景的 Scan2BIM 中, 简洁高效地融合 GPS 和 IMU 等多源异构数据, 提高全局点云配准算法的可靠性和准确率, 为 Scan2BIM 后续过程提供更准确的输入, 从而提高整体算法逆向建模效果.

6.2 精度、鲁棒性与性能的协同优化

受到噪声和异常值较多、分布密度不规则、物体大小不一、遮挡、不完整、不同的初始位姿、计算资源消耗较大等影响, 目前, 极少算法能在精度、计算性能与鲁棒性等多方面均表现出色. 未来, 精度、鲁棒性与性能的协同优化依然是该领域内亟待解决的课题. 围绕该主题的方向如下:

(1) 精度提升. 首先, 增加丰富多样的优质点云数据集, 是提升 Scan2BIM 各模块效果的一个有效的措施, 未来仍需探究更多潜在的可能性与有效性; 其次, 实际场景中通常存在更多的遮挡和不完整, 在大量数据缺失的环境中, 逆向构建高质量 BIM 是个巨大的挑战, 可探索结合点云补全或者点云生成算法弥补点云遮挡和不完整的缺陷, 进一步提升物体识别和分割质量和鲁棒性. 此外, 现实世界物体多姿多态, 几何拓扑形态复杂不规则, 空间上存在相邻甚至重叠情况, 可能处于运动变化状态等. 因此, 设计高效灵活的知识表示和学习框架, 以提取更有效的 3D 特征、精确对齐连续场景中的同一物体, 以及准确地识别点云分割算法中的边界, 都是具有挑战性的任务.

(2) 计算效率优化. 兼顾精度和效率的算法是具备应用推广价值的关键条件, 这对算法设计提出了更高层次的期望, 也是未来一个必不可少的研究方向. 对算法模型结构进行优化, 在保证模型精度的同时降低计算复杂度, 如模型参数优化、推理加速、将多阶段算法融合到一个端对端学习框架等, 都值得未来进一步研究.

(3) 鲁棒性增强. 一方面, 为了减少高成本和高精度数据集的限制和依赖, 适应更复杂和非结构化的环境变化, 减少算法对异常值和噪声的敏感性, 未来应继续探索无监督学习和小样本学习等自适应学习机制, 确保算法在真实场景中稳定可靠; 另一方面, 由于室内和室外场景之间存在巨大的差异, 现阶段室内场景所训练的模型无法直

接迁移到室外场景,未来应更多地关注基于大规模预训练模型和迁移学习改进算法的设计,使之既能适用于训练时的特定对象或类别,也能直接跨域泛化到实际新应用场景甚至开放场景。

6.3 构建端到端 Scan2BIM 通用框架

从 Scan2BIM 整体算法框架看,当前,大规模复杂室外场景下的 Scan2BIM 仍然面临着一定的不足和局限性:(1)无法建模全场景关键要素,多聚焦于单一要素或部分要素提取;(2)3D 扫描点云中存在许多缺损,造成点云分割后 BIM 建模精度存在误差;(3)现有研究未构成端到端的全自动化 Scan2BIM 过程,从点云数据采集至 BIM 生成之间的关键步骤衔接还需人工干预,影响自动化建模效率与精度;(4)在处理复杂几何、生成参数化模型以及确保几何关系的准确性方面,尤其是在考虑室外建筑设施多样性与空间位置复杂性时,当前算法容易忽略关键点云数据,未来的研究重点是以高效率处理非结构化点云数据,精确地提取出几何特征,并构建出既可参数化又能动态更新的 BIM;(5)高昂的数据采集成本限制了训练资料的丰富度,增加了深度学习模型应用难度。

因此,未来应利用深度学习算法和整合多源数据集,来高效地识别复杂室外场景关键要素,解决点云数据不全等问题,提升自动化提取效率和质量,实现室外建筑设施的端到端的全自动化 BIM 逆向建模。在此过程中,创新智能化 BIM 工具的开发、标准化的数据交换协议的确立,以及增强的用户互动设计,都将是模型优化的关键点。

6.4 大模型应用与探索

基于深度学习的 Scan2BIM 在大规模室外场景的 BIM 逆向重建中展现了较为卓越的能力,是当前研究的热点技术路线。大模型凭借其庞大的参数量和深度网络架构,能够发现数据细节和复杂模式,并通过广泛数据训练展现出强大的表达和迁移能力,推动了多个领域的创新和突破,虽然大模型在 Scan2BIM 方向的现有成果不多,探索也还不够成熟,但呈现出了巨大的发展潜力。因此,如何通过深入融合大模型提升 Scan2BIM 性能是未来的重要研究方向之一,可从 3 个方面进行探索:

(1) 基于大模型的点云数据生成。丰富、完整和高精度的点云数据集是实现大规模场景下的 Scan2BIM 的充分条件之一。目前,3D 点云获取依旧存在很多问题没有解决。首先,传统点云数据采集方式普遍存在采集和标注成本高、质量难保证等问题;其次,实际采集的点云虽然真实性强,但可

能存在噪声、遮挡和信息缺失等问题,而模拟器虽然能够大幅降低点云标注成本,但可能缺乏真实世界细节;相比之下,依靠大模型强大的表示和泛化能力,可实现高效地生成比较干净和完整的 3D 点云场景,作为采集式点云数据的补充。然而,现有的基于大模型的 3D 点云数据生成局限于单体化物体或室内小场景,并且细节性和复杂性仍然存在不小差距^[31]。因此,如何将应用范围扩展到大规模室外场景,以及保证数据质量和真实性是巨大的挑战,仍需进一步深入研究。

(2) 基于大模型的点云场景分割。大模型的泛化能力非常出色,能够在没有针对特定类别进行训练的情况下,通过学习到的通用特征识别和分割新的或未见过的物体类别。部分研究^[174]已开始利用大模型的高级特征提取能力,通过融合 3D 点云几何先验知识与 2D 图像语义解析能力,构建跨维度的精准空间映射关系,即使在复杂的开放室内环境中,也能实现零样本学习的鲁棒性分割,并精细地分割出 3D 物体。但是,目前的研究面临众多挑战,包括对大量计算资源的依赖、在处理大规模场景下复杂 3D 结构和罕见物体时泛化能力的局限,以及如何降低对用户交互的需求以实现全自动化分割。这些问题制约了模型在特定应用场景中的适用性和稳定性,需要进一步探索 and 解决。

(3) 基于大模型的点云场景理解。当前,有少量研究已开始关注如何利用大语言模型强大的语义理解能力理解 3D 场景语义和空间结构,理论上可辅助 Scan2BIM 各个阶段对 3D 场景感知和理解,为未来 Scan2BIM 发展开辟了新方向。PointLLM^[175]作为一个 3D 多模态大型语言模型,能够在零样本学习情况下理解 3D 点云,可以感知开放世界的对象类型、几何结构和外观等;GPT4Point^[176]通过 3D 点云与文本对齐,实现了点云语义理解和生成的统一框架。然而,该方向的研究才刚刚起步,仍存在诸多问题与挑战,如 3D 场景语言数据集在质量和数量上均存在不足,3D 场景理解还局限于小规模场景或单物体,语义类型和空间结构识别精确度和稳定性还不高,以及如何结合基于大模型的点云场景理解能力提高 Scan2BIM 的建模效果等。这些问题对于未来 Scan2BIM 的发展至关重要,需要投入更多的研究精力和资源。

综上所述,未来需着重提高算法和模型的通用性,以适应不同类型的基础设施和各种运行环境;同时,提升自动化流程中的各个环节,减少人为干预,增强系统的自主性和适应性。首先,点云

预处理、点云全局配准和点云实例分割等算法可以从融合多源数据集、模型计算优化,以及鲁棒性和可解释性上进行持续突破;特别地,大规模室外复杂场景点云语义实例分割的研究应当综合考虑室外建筑设施在不同视角下的3D真实映射,并设计对硬件资源更为友好的算法框架;最后,研究重心应趋向于识别出大规模场景中更细致化且更为真实的对象。因此,未来应致力于推进一个面向大规模复杂室外场景的基于端到端的自动化 Scan2BIM 通用框架,推动室外建筑设施运行维护进入一个更智能、更可持续的新纪元,为建筑行业的数字孪生和智慧数字化等研究提供持续且可靠的动力。

参考文献(References):

- [1] Shin J, Song J. Dismantling quantity estimation for nuclear power plant: scan-to-BIM versus conventional method[J]. KSCE Journal of Civil Engineering, 2024, 28(5): 1607-1621
- [2] Banfi F, Fai S, Brumana R. BIM automation: advanced modeling generative process for complex structures[J]. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2017, IV-2-W2: 9-16
- [3] Cheng Y J, Qiu W G, Duan D Y. Automatic creation of as-is building information model from single-track railway tunnel point clouds[J]. Automation in Construction, 2019, 106: Article No.102911
- [4] Park J, Kim J, Lee D, *et al.* Deep learning-based automation of scan-to-BIM with modeling objects from occluded point clouds[J]. Journal of Management in Engineering, 2022, 38(4): Article No.04022025
- [5] Perez-Perez Y, Golparvar-Fard M, El-Rayes K. Scan2BIM-NET: deep learning method for segmentation of point clouds for scan-to-BIM[J]. Journal of Construction Engineering and Management, 2021, 147(9): Article No.04021107
- [6] Xie Y, Li S Y, Liu T R, *et al.* As-built BIM reconstruction of piping systems using PipeNet[J]. Automation in Construction, 2023, 147: Article No.104735
- [7] He Wenjing, Yang Jian, Xiong Wuyue. BIM model method for automatic generation of 3D point clouds based on deep learning[J]. Journal of Civil Engineering and Management, 2021, 38(3): 133-139(in Chinese)
(何文景, 杨健, 熊吴越. 基于深度学习的三维点云自动生成 BIM 模型方法[J]. 土木工程与管理学报, 2021, 38(3): 133-139)
- [8] Xia T, Yang J, Chen L. Automated semantic segmentation of bridge point cloud based on local descriptor and machine learning[J]. Automation in Construction, 2022, 133: Article No.103992
- [9] Wang Yuchen, Yu Bin, Chen Xiaoyang, *et al.* Extraction and digital modeling of road geometric information using LiDAR data point clouds[J]. China Journal of Highway and Transport, 2023, 36(3): 45-60(in Chinese)
(王羽尘, 于斌, 陈晓阳, 等. 基于激光雷达点云的道路几何信息提取与数字化建模研究[J]. 中国公路学报, 2023, 36(3): 45-60)
- [10] Rashdi R, Martínez-Sánchez J, Arias P, *et al.* Scanning technologies to building information modelling: a review[J]. Infrastructures, 2022, 7(4): Article No.49
- [11] Czerniawski T, Leite F. Automated digital modeling of existing buildings: a review of visual object recognition methods[J]. Automation in Construction, 2020, 113: Article No.103131
- [12] Gong Jinggyu, Lou Yujing, Liu Fengqi, *et al.* Scene point cloud understanding and reconstruction technologies in 3D space[J]. Journal of Image and Graphics, 2023, 28(6): 1741-1766(in Chinese)
(龚靖渝, 楼雨京, 柳奉奇, 等. 三维场景点云理解与重建技术[J]. 中国图象图形学报, 2023, 28(6): 1741-1766)
- [13] Samavati T, Soryani M. Deep learning-based 3D reconstruction: a survey[J]. Artificial Intelligence Review, 2023, 56(9): 9175-9219
- [14] Mirzaei K, Arashpour M, Asadi E, *et al.* 3D point cloud data processing with machine learning for construction and infrastructure applications: a comprehensive review[J]. Advanced Engineering Informatics, 2022, 51: Article No.101501
- [15] Monji-Azad S, Hesser J, Löw N. A review of non-rigid transformations and learning-based 3D point cloud registration methods[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2023, 196: 58-72
- [16] Salzano A, Intignano M, Mottola C, *et al.* Systematic literature review of open infrastructure BIM[J]. Buildings, 2023, 13(7): Article No.1593
- [17] Sirohi K, Mohan R, Büscher D, *et al.* EfficientLPS: efficient LiDAR panoptic segmentation[J]. IEEE Transactions on Robotics, 2022, 38(3): 1894-1914
- [18] Liu J M, Wang G M, Liu Z, *et al.* RegFormer: an efficient projection-aware transformer network for large-scale point cloud registration[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2023: 8417-8426
- [19] Xiang B, Peters T, Kontogianni T, *et al.* Towards accurate instance segmentation in large-scale LiDAR point clouds[J]. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, 2023, X-1/W1-2023: 605-612
- [20] Wang Y C, Wang W C, Liu J Z, *et al.* Framework for geometric information extraction and digital modeling from LiDAR data of road scenarios[J]. Remote Sensing, 2023, 15(3): Article No.576
- [21] Tomsett C, Leyland J. Development and testing of a UAV laser scanner and multispectral camera system for eco-geomorphic applications[J]. Sensors, 2021, 21(22): Article No.7719
- [22] Soilán M, Sánchez-Rodríguez A, del Río-Barral P, *et al.* Review of laser scanning technologies and their applications for road and railway infrastructure monitoring[J]. Infrastructures, 2019, 4(4): Article No.58
- [23] 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
- [24] Deng Weiwen, Li Jiangkun, Ren Bingtao, *et al.* A survey on automatic simulation scenario generation methods for autonomous driving[J]. China Journal of Highway and Transport,

- 2022, 35(1): 316-333(in Chinese)
(邓伟文, 李江坤, 任秉韬, 等. 面向自动驾驶的仿真场景自动生成方法综述[J]. 中国公路学报, 2022, 35(1): 316-333)
- [25] Shah S, Dey D, Lovett C, *et al.* AirSim: high-fidelity visual and physical simulation for autonomous vehicles[C] //Proceedings of the 11th International Conference on Field and Service Robotics. Heidelberg: Springer, 2018: 621-635
- [26] Rong G D, Shin B H, Tabatabaee H, *et al.* LGSVL simulator: a high fidelity simulator for autonomous driving[C] //Proceedings of the 23rd IEEE International Conference on Intelligent Transportation Systems. Los Alamitos: IEEE Computer Society Press, 2020: 1-6
- [27] Dosovitskiy A, Ros G, Codevilla F, *et al.* CARLA: an open urban driving simulator[C] //Proceedings of the 1st Annual Conference on Robot Learning. Cambridge MA: MLResearch Press, 2017: 1-16
- [28] Li Q Y, Peng Z H, Feng L, *et al.* MetaDrive: composing diverse driving scenarios for generalizable reinforcement learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(3): 3461-3475
- [29] Swerdlow A, Xu R S, Zhou B L. Street-view image generation from a bird's-eye view layout[J]. IEEE Robotics and Automation Letters, 2024, 9(4): 3578-3585
- [30] Gao R Y, Chen K, Xie E Z, *et al.* MagicDrive: street view generation with diverse 3D geometry control[C] //Proceedings of the 12th International Conference on Learning Representations. Red Hook: Curran Associates Inc, 2024: 1-20
- [31] Yang Y, Sun F Y, Weihs L, *et al.* Holodeck: language guided generation of 3D embodied AI environments[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2024: 16277-16287
- [32] Geiger A, Lenz P, Urtasun R. Are we ready for autonomous driving? The KITTI vision benchmark suite[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2012: 3354-3361
- [33] Hackel T, Savinov N, Ladicky L, *et al.* SEMANTIC3D.NET: a new large-scale point cloud classification benchmark[J]. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2017, IV-1/W1: 91-98
- [34] Chang M F, Lambert J, Sangkloy P, *et al.* Argoverse: 3D tracking and forecasting with rich maps[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 8740-8749
- [35] Behley J, Garbade M, Milioto A, *et al.* SemanticKITTI: a dataset for semantic scene understanding of LiDAR sequences[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2019: 9296-9306
- [36] Caesar H, Bankiti V, Lang A H, *et al.* nuScenes: a multimodal dataset for autonomous driving[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2020: 11618-11628
- [37] Sun P, Kretschmar H, Dotiwalla X, *et al.* Scalability in perception for autonomous driving: waymo open dataset[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2020: 2443-2451
- [38] Pham Q H, Sevestre P, Pahwa R S, *et al.* A*3D dataset: towards autonomous driving in challenging environments[C] //Proceedings of the IEEE International Conference on Robotics and Automation. Los Alamitos: IEEE Computer Society Press, 2020: 2267-2273
- [39] Tan W K, Qin N N, Ma L F, *et al.* Toronto-3D: a large-scale mobile LiDAR dataset for semantic segmentation of urban roadways[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Los Alamitos: IEEE Computer Society Press, 2020: 797-806
- [40] Xiao P C, Shao Z L, Hao S, *et al.* PandaSet: advanced sensor suite dataset for autonomous driving[C] //Proceedings of the IEEE International Intelligent Transportation Systems Conference. Los Alamitos: IEEE Computer Society Press, 2021: 3095-3101
- [41] Houston J, Zuidhof G, Bergamini L, *et al.* One thousand and one hours: self-driving motion prediction dataset[C] //Proceedings of the Conference on Robot Learning. Cambridge MA: MLResearch Press, 2021: 409-418
- [42] Caesar H, Kabzan J, Tan K S, *et al.* NuPlan: a closed-loop ML-based planning benchmark for autonomous vehicles[OL]. [2024-05-22]. <https://arxiv.org/abs/2106.11810>
- [43] Xiao A R, Huang J X, Xuan W H, *et al.* 3D semantic segmentation in the wild: learning generalized models for adverse-condition point clouds[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2023: 9382-9392
- [44] Mao J G, Niu M Z, Jiang C H, *et al.* One million scenes for autonomous driving: ONCE dataset[OL]. [2024-05-22]. <https://arxiv.org/abs/2106.11037>
- [45] Cao A Q, de Charette R. MonoScene: monocular 3D semantic scene completion[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 3981-3991
- [46] Tian X Y, Jiang T, Yun L F, *et al.* Occ3D: a large-scale 3D occupancy prediction benchmark for autonomous driving[C] //Proceedings of the 37th International Conference on Neural Information Processing Systems. Red Hook: Curran Associates Inc, 2023: Article No.2809
- [47] Wei Y, Zhao L Q, Zheng W Z, *et al.* SurroundOcc: multi-camera 3D occupancy prediction for autonomous driving[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2023: 21672-21683
- [48] Wang X F, Zhu Z, Xu W B, *et al.* OpenOccupancy: a large scale benchmark for surrounding semantic occupancy perception[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2023: 17804-17813
- [49] Li Y M, Li S H, Liu X H, *et al.* SSCBench: a large-scale 3D semantic scene completion benchmark for autonomous driving[C] //Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. Los Alamitos: IEEE Com-

- puter Society Press, 2024: 13333-13340
- [50] Xiao A R, Huang J X, Guan D Y, *et al.* Transfer learning from synthetic to real LiDAR point cloud for semantic segmentation[C] //Proceedings of the 36th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2022: 2795-2803
- [51] Hurl B, Czarnecki K, Waslander S. Precise synthetic image and LiDAR (PreSIL) dataset for autonomous vehicle perception[C] //Proceedings of the IEEE Intelligent Vehicles Symposium. Los Alamitos: IEEE Computer Society Press, 2019: 2522-2529
- [52] Xu R S, Xiang H, Xia X, *et al.* OPV2V: an open benchmark dataset and fusion pipeline for perception with vehicle-to-vehicle communication[C] //Proceedings of the International Conference on Robotics and Automation. Los Alamitos: IEEE Computer Society Press, 2022: 2583-2589
- [53] Ros G, Sellart L, Materzynska J, *et al.* The synthia dataset: a large collection of synthetic images for semantic segmentation of urban scenes[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2016: 3234-3243
- [54] Sun T, Segu M, Postels J, *et al.* SHIFT: a synthetic driving dataset for continuous multi-task domain adaptation[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 21339-21350
- [55] Lin L Q, Liu Y L, Hu Y, *et al.* Capturing, reconstructing, and simulating: the urbanscene3D dataset[C] //Proceedings of the 17th European Conference on Computer Vision. Heidelberg: Springer, 2022: 93-109
- [56] Zhou Y C, Huang J W, Dai X L, *et al.* HoliCity: a city-scale data platform for learning holistic 3D structures[OL]. [2024-05-22]. <https://arxiv.org/abs/2008.03286v2>
- [57] Yang Guoqing, Lai Wentao, Huang Hui. On active labeling 3D point clouds via contrastive learning[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2023, 35(11): 1664-1673(in Chinese)
(杨国庆, 赖文韬, 黄惠. 结合对比学习的三维点云主动标注方法[J]. *计算机辅助设计与图形学学报*, 2023, 35(11): 1664-1673)
- [58] Xiao A R, Zhang X Q, Shao L, *et al.* A survey of label-efficient deep learning for 3D point clouds[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024, 46(12): 9139-9160
- [59] Huang L Y. Review on LiDAR-based SLAM techniques[C] //Proceedings of the International Conference on Signal Processing and Machine Learning. Los Alamitos: IEEE Computer Society Press, 2021: 163-168
- [60] Wu Yue, Yuan Yongzhe, Xiang Benhua, *et al.* Overview of the computational intelligence method in 3D point cloud registration[J]. *Journal of Image and Graphics*, 2023, 28(9): 2763-2787 (in Chinese)
(武越, 苑咏哲, 向本华, 等. 三维点云配准中的计算智能方法综述[J]. *中国图象图形学报*, 2023, 28(9): 2763-2787)
- [61] Besl P J, McKay N D. Method for registration of 3-D shapes[C] //Proceedings of the Sensor Fusion IV: Control Paradigms and Data Structures. Bellingham: Society of Photo-Optical Instrumentation Engineers, 1992: 586-606
- [62] Fan Linlin, Wang Junyi, Xu Zhigang, *et al.* Registration method of partial point cloud and whole point cloud of large workpiece[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2023, 35(9): 1323-1332(in Chinese)
(范林林, 王军义, 徐志刚, 等. 大型工件部分点云与整体点云的配准方法[J]. *计算机辅助设计与图形学学报*, 2023, 35(9): 1323-1332)
- [63] Biber P, Straßer W. The normal distributions transform: a new approach to laser scan matching[C] //Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. Los Alamitos: IEEE Computer Society Press, 2003: 2743-2748
- [64] Fischler M A, Bolles R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography[J]. *Communications of the ACM*, 1981, 24(6): 381-395
- [65] Mellado N, Aiger D, Mitra N J. Super 4PCS fast global point-cloud registration via smart indexing[J]. *Computer Graphics Forum*, 2014, 33(5): 205-215
- [66] Horn M, Engel N, Belagiannis V, *et al.* DeepCLR: correspondence-less architecture for deep end-to-end point cloud registration[C] //Proceedings of the 23rd IEEE International Conference on Intelligent Transportation Systems. Los Alamitos: IEEE Computer Society Press, 2020: 1-7
- [67] Wang Y, Solomon J M. Deep closest point: learning representations for point cloud registration[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2019: 3522-3531
- [68] Aoki Y, Goforth H, Srivatsan R A, *et al.* PointNetLK: robust & efficient point cloud registration using PointNet[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 7156-7165
- [69] Li J H, Zhang C H, Xu Z Y, *et al.* Iterative distance-aware similarity matrix convolution with mutual-supervised point elimination for efficient point cloud registration[C] //Proceedings of the 16th European Conference on Computer Vision. Heidelberg: Springer, 2020: 378-394
- [70] Yuan W T, Eckart B, Kim K, *et al.* DeepGMR: learning latent gaussian mixture models for registration[C] //Proceedings of the 16th European Conference on Computer Vision. Heidelberg: Springer, 2020: 733-750
- [71] Huang X S, Mei G F, Zhang J. Feature-metric registration: a fast semi-supervised approach for robust point cloud registration without correspondences[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2020: 11363-11371
- [72] Zeng A, Song S R, Nießner M, *et al.* 3DMatch: learning local geometric descriptors from RGB-D reconstructions[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 199-208
- [73] Deng H W, Birdal T, Ilic S. PPFNet: global context aware local features for robust 3D point matching[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 195-205

- [74] Gojic Z, Zhou C F, Wegner J D, *et al.* The perfect match: 3D point cloud matching with smoothed densities[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 5540-5549
- [75] Choy C, Dong W, Koltun V. Deep global registration[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2020: 2511-2520
- [76] Lu W X, Wan G W, Zhou Y, *et al.* DeepVCP: an end-to-end deep neural network for point cloud registration[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2019: 12-21
- [77] Lee J, Kim S, Cho M, *et al.* Deep hough voting for robust global registration[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2021: 15974-15983
- [78] 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, 2021, 43(12): 4338-4364
- [79] Yew Z J, Lee G H. 3DFeat-Net: weakly supervised local 3D features for point cloud registration[C] //Proceedings of the 15th European Conference on Computer Vision. Heidelberg: Springer, 2018: 630-646
- [80] Choy C, Park J, Koltun V. Fully convolutional geometric features[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2019: 8957-8965
- [81] Ao S, Hu Q Y, Yang B, *et al.* SpinNet: learning a general surface descriptor for 3D point cloud registration[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2021: 11748-11757
- [82] Li J X, Lee G H. USIP: unsupervised stable interest point detection from 3D point clouds[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2019: 361-370
- [83] Bai X Y, Luo Z X, Zhou L, *et al.* D3Feat: joint learning of dense detection and description of 3D local features[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2020: 6358-6366
- [84] Liu S C, Wang T, Zhang Y, *et al.* Rethinking of learning-based 3D keypoints detection for large-scale point clouds registration[J]. International Journal of Applied Earth Observation and Geoinformation, 2022, 112: Article No.102944
- [85] Ao S, Hu Q Y, Wang H Y, *et al.* BUFFER: balancing accuracy, efficiency, and generalizability in point cloud registration[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2023: 1255-1264
- [86] Wu Y, Zhang Y, Ma W P, *et al.* RORNet: partial-to-partial registration network with reliable overlapping representations[J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(11): 15453-15466
- [87] Liu S C, Wang T, Zhang Y, *et al.* Deep semantic graph matching for large-scale outdoor point cloud registration[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: Article No.5701412
- [88] Yu H, Li F, Saleh M, *et al.* CoFiNet: reliable coarse-to-fine correspondences for robust point cloud registration[C] //Proceedings of the 35th International Conference on Neural Information Processing Systems. Red Hook: Curran Associates Inc, 2021: Article No.1828
- [89] Qin Z, Yu H, Wang C J, *et al.* GeoTransformer: fast and robust point cloud registration with geometric transformer[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(8): 9806-9821
- [90] Lu F, Chen G, Liu Y L, *et al.* HRegNet: a hierarchical network for large-scale outdoor LiDAR point cloud registration[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2021: 15994-16003
- [91] Lu F, Chen G, Liu Y L, *et al.* HRegNet: a hierarchical network for efficient and accurate outdoor LiDAR point cloud registration[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(10): 11884-11897
- [92] Shi C H, Chen X Y L, Lu H M, *et al.* RDMNet: reliable dense matching based point cloud registration for autonomous driving[J]. IEEE Transactions on Intelligent Transportation Systems, 2023, 24(10): 11372-11383
- [93] Xue W Y, Lu F, Chen G. HDMNet: a hierarchical matching network with double attention for large-scale outdoor LiDAR point cloud registration[C] //Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Los Alamitos: IEEE Computer Society Press, 2024: 3381-3391
- [94] Li J K, He X, Wen Y, *et al.* Panoptic-PHNet: towards real-time and high-precision LiDAR panoptic segmentation via clustering pseudo heatmap[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 11799-11808
- [95] Wang W Y, Yu R, Huang Q G, *et al.* SGPN: similarity group proposal network for 3D point cloud instance segmentation[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 2569-2578
- [96] Hou J, Dai A, Nießner M. 3D-SIS: 3D semantic instance segmentation of RGB-D scans[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 4416-4425
- [97] Narita G, Seno T, Ishikawa T, *et al.* PanopticFusion: online volumetric semantic mapping at the level of stuff and things[C] //Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. Los Alamitos: IEEE Computer Society Press, 2019: 4205-4212
- [98] Engelmann F, Bokeloh M, Fathi A, *et al.* 3D-MPA: Multi-proposal aggregation for 3D semantic instance segmentation[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2020: 9028-9037
- [99] Liang Z H, Li Z H, Xu S C, *et al.* Instance segmentation in 3D scenes using semantic superpoint tree networks[C] //Proceedings of the IEEE/CVF International Conference on Computer

- Vision. Los Alamitos: IEEE Computer Society Press, 2021: 2763-2772
- [100] Liao Y B, Zhu H Y, Zhang Y G, *et al.* Point cloud instance segmentation with semi-supervised bounding-box mining[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(12): 10159-10170
- [101] Ngo T D, Hua B S, Nguyen K. GaPro: box-supervised 3D point cloud instance segmentation using Gaussian processes as pseudo labelers[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2023: 17748-17757
- [102] Liu C, Furukawa Y. MASC: multi-scale affinity with sparse convolution for 3D instance segmentation[OL]. [2024-05-22]. <https://arxiv.org/pdf/1902.04478>
- [103] Pham Q H, Nguyen T, Hua B S, *et al.* JSIS3D: joint semantic-instance segmentation of 3D point clouds with multi-task pointwise networks and multi-value conditional random fields[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 8819-8828
- [104] Agapaki E, Brilakis I. CLOI-NET: class segmentation of industrial facilities' point cloud datasets[J]. Advanced Engineering Informatics, 2020, 45: Article No.101121
- [105] Vu T, Kim K, Luu T M, *et al.* Softgroup for 3D instance segmentation on point clouds[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 2698-2707
- [106] He T, Shen C H, van den Hengel A. Dynamic convolution for 3D point cloud instance segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(5): 5697-5711
- [107] Zhao Y M, Zhang X, Huang X M. A technical survey and evaluation of traditional point cloud clustering methods for LiDAR panoptic segmentation[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2021: 2464-2473
- [108] Jiang L, Zhao H S, Shi S S, *et al.* PointGroup: dual-set point grouping for 3D instance segmentation[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2020: 4866-4875
- [109] Wang X L, Liu S, Shen X Y *et al.* Associatively segmenting instances and semantics in point clouds[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 4091-4100
- [110] Zhao L, Tao W B. JSNet: joint instance and semantic segmentation of 3D point clouds[C] //Proceedings of the 34th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2020: 12951-12958
- [111] Lahoud J, Ghanem B, Oswald M R, *et al.* 3D instance segmentation via multi-task metric learning[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2019: 9255-9265
- [112] Chen S Y, Fang J M, Zhang Q, *et al.* Hierarchical aggregation for 3D instance segmentation[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2021: 15447-15456
- [113] Hao W, Wang H X, Liang W, *et al.* Attention-Based joint semantic-instance segmentation of 3D point clouds[J]. Advances in Electrical and Computer Engineering, 2022, 22(2): 19-28
- [114] Zhong M, Chen X H, Chen X K, *et al.* MaskGroup: hierarchical point grouping and masking for 3D instance segmentation[C] //Proceedings of the IEEE International Conference on Multimedia and Expo. Los Alamitos: IEEE Computer Society Press, 2022: 1-6
- [115] Wong K, Wang S L, Ren M Y, *et al.* Identifying unknown instances for autonomous driving[C] //Proceedings of the Conference on Robot Learning. Cambridge MA: MLResearch Press, 2020: 384-393
- [116] Zhang F H, Guan C Y, Fang J, *et al.* Instance segmentation of LiDAR point clouds[C] //Proceedings of the IEEE International Conference on Robotics and Automation. Los Alamitos: IEEE Computer Society Press, 2020: 9448-9455
- [117] Shin S, Zhou K C, Vankadari M, *et al.* Spherical mask: coarse-to-fine 3D point cloud instance segmentation with spherical representation[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2024: 4060-4069
- [118] Wu B C, Wan A, Yue X Y, *et al.* SqueezeSeg: convolutional neural nets with recurrent CRF for real-time road-object segmentation from 3D LiDAR point cloud[C] //Proceedings of the IEEE International Conference on Robotics and Automation. Los Alamitos: IEEE Computer Society Press, 2018: 1887-1893
- [119] Chen T H, Chang T S. RangeSeg: range-aware real time segmentation of 3D LiDAR point clouds[J]. IEEE Transactions on Intelligent Vehicles, 2021, 7(1): 93-101
- [120] Du K Y, Meng J, Meng X, *et al.* A model for urban environment instance segmentation with data fusion[J]. Sensors, 2023, 23(13): Article No.6141
- [121] Zhang Z K, Ding J, Jiang L, *et al.* FreePoint: unsupervised point cloud instance segmentation[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2024: 28254-28263
- [122] Kirillov A, He K M, Girshick R, *et al.* Panoptic segmentation[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 9396-9405
- [123] Su S H, Xu J Y, Wang H Y, *et al.* PUPS: point cloud unified panoptic segmentation[C] //Proceedings of the 37th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2023: 2339-2347
- [124] Behley J, Milioto A, Stachniss C. A benchmark for LiDAR-based panoptic segmentation based on KITTI[C] //Proceedings of the IEEE International Conference on Robotics and Automation. Los Alamitos: IEEE Computer Society Press, 2021: 13596-13603
- [125] Thomas H, Qi C R, Deschaud J E, *et al.* KPConv: flexible and deformable convolution for point clouds[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2019: 6410-6419
- [126] Milioto A, Vizzo I, Behley J, *et al.* RangeNet++: fast and accu-

- rate LiDAR semantic segmentation[C] //Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. Los Alamitos: IEEE Computer Society Press, 2019: 4213-4220
- [127] Lang A H, Vora S, Caesar H, *et al.* PointPillars: fast encoders for object detection from point clouds[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 12689-12697
- [128] Chen Q, Vora S, Beijbom O. PolarStream: streaming LiDAR object detection and segmentation with polar pillars[C] //Proceedings of the 35th International Conference on Neural Information Processing Systems. Red Hook: Curran Associates Inc, 2021: Article No.2058
- [129] Gasperini S, Mahani M A N, Marcos-Ramiro A, *et al.* Panoster: end-to-end panoptic segmentation of LiDAR point clouds[J]. IEEE Robotics and Automation Letters, 2021, 6(2): 3216-3223
- [130] Razani R, Cheng R, Li E X, *et al.* GP-S3Net: graph-based panoptic sparse semantic segmentation network[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2021: 16056-16065
- [131] Shi C Z, Own C M, Zhou R M. Federated grouping panoptic segmentation for LiDAR point cloud based on convolutional network[C] //Proceedings of the 34th IEEE International Conference on Tools with Artificial Intelligence. Los Alamitos: IEEE Computer Society Press, 2022: 6-13
- [132] Marcuzzi R, Nunes L, Wiesmann L, *et al.* Mask-based panoptic LiDAR segmentation for autonomous driving[J]. IEEE Robotics and Automation Letters, 2023, 8(2): 1141-1148
- [133] Wang Y Z, Wang L G, Hu Q Y, *et al.* Panoptic segmentation of 3D point clouds with Gaussian mixture model in outdoor scenes[J]. Visual Intelligence, 2024, 2(1): Article No.10
- [134] Milioto A, Behley J, McCool C, *et al.* LiDAR panoptic segmentation for autonomous driving[C] //Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. Los Alamitos: IEEE Computer Society Press, 2020: 8505-8512
- [135] Zhou Z X, Zhang Y, Foroosh H. Panoptic-PolarNet: proposal-free LiDAR point cloud panoptic segmentation[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2021: 13189-13198
- [136] Li E X, Razani R, Xu Y X, *et al.* SMAC-Seg: LiDAR panoptic segmentation via sparse multi-directional attention clustering[C] //Proceedings of the International Conference on Robotics and Automation. Los Alamitos: IEEE Computer Society Press, 2022: 9207-9213
- [137] Basic B, Gosala N, Cattaneo D, *et al.* Unsupervised domain adaptation for LiDAR panoptic segmentation[J]. IEEE Robotics and Automation Letters, 2022, 7(2): 3404-3411
- [138] Xian G Z, Ji C Y, Zhou L, *et al.* Location-guided LiDAR-based panoptic segmentation for autonomous driving[J]. IEEE Transactions on Intelligent Vehicles, 2023, 8(2): 1473-1483
- [139] Li X Y, Zhang G, Wang B Y, *et al.* Center focusing network for real-time LiDAR panoptic segmentation[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2023: 13425-13434
- [140] Zhu X G, Zhou H, Wang T, *et al.* Cylindrical and asymmetrical 3D convolution networks for LiDAR segmentation[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2021: 9934-9943
- [141] Zhao Y M, Zhang X, Huang X M. A divide-and-merge point cloud clustering algorithm for LiDAR panoptic segmentation[C] //Proceedings of the International Conference on Robotics and Automation. Los Alamitos: IEEE Computer Society Press, 2022: 7029-7035
- [142] Chen Q, Vora S. Proposal-free LiDAR panoptic segmentation with pillar-level affinity[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Los Alamitos: IEEE Computer Society Press, 2022: 4528-4535
- [143] Hong F Z, Zhou H, Zhu X G, *et al.* LiDAR-based panoptic segmentation via dynamic shifting network[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2021: 13085-13094
- [144] Xu S J, Wan R, Ye M S, *et al.* Sparse cross-scale attention network for efficient LiDAR panoptic segmentation[C] //Proceedings of the 36th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2022: 2920-2928
- [145] Li E X, Razani R, Xu Y X, *et al.* CPseg: cluster-free panoptic segmentation of 3D LiDAR point clouds[C] //Proceedings of the IEEE International Conference on Robotics and Automation. Los Alamitos: IEEE Computer Society Press, 2023: 8239-8245
- [146] Song H M, Cho J, Ha J S, *et al.* Panoptic-FusionNet: camera-LiDAR fusion-based point cloud panoptic segmentation for autonomous driving[J]. Expert Systems with Applications, 2024, 251: Article No.123950
- [147] Nichol A, Jun H, Dhariwal P, *et al.* Point-E: a system for generating 3D point clouds from complex prompts[OL]. [2024-05-22]. <https://arxiv.org/abs/2212.08751>
- [148] Gong J Y, Liu F Q, Xu J C, *et al.* Optimization over disentangled encoding: unsupervised cross-domain point cloud completion via occlusion factor manipulation[C] //Proceedings of the 17th European Conference on Computer Vision. Heidelberg: Springer, 2022: 517-533
- [149] Wu Y, Hu X D, Zhang Y, *et al.* SACF-Net: skip-attention based correspondence filtering network for point cloud registration[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 33(8): 3585-3595
- [150] Liu F Q, Gong J Y, Zhou Q Y, *et al.* CloudMix: dual mixup consistency for unpaired point cloud completion[J]. IEEE Transactions on Visualization and Computer Graphics, 2025, 31(4): 2182-2195
- [151] Wang J, Cui Y, Guo D Y, *et al.* PointAttN: you only need attention for point cloud completion[C] //Proceedings of the 38th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2024: 5472-5480
- [152] Mahmoud M, Chen W, Yang Y, *et al.* Leveraging deep learning for automated reconstruction of indoor unstructured elements in

- scan-to-BIM[J]. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2024, 48(1): 479-486
- [153] Tang P B, Huber D, Akinci B, *et al.* Automatic reconstruction of as-built building information models from laser-scanned point clouds: a review of related techniques[J]. *Automation in Construction*, 2010, 19(7): 829-843
- [154] Costin A, Adibfar A, Hu H J, *et al.* Building information modeling (BIM) for transportation infrastructure—literature review, applications, challenges, and recommendations[J]. *Automation in Construction*, 2018, 94: 257-281
- [155] Cho G H, Won J S, Kim J U. The extension of IFC model schema for geometry part of road drainage facility[J]. *Journal of the Korea Academia-Industrial cooperation Society*, 2013, 14(11): 5987-5992
- [156] Borrmann A, Kolbe T H, Donaubaer A, *et al.* Multi-scale geometric-semantic modeling of shield tunnels for GIS and BIM applications[J]. *Computer-Aided Civil and Infrastructure Engineering*, 2015, 30(4): 263-281
- [157] Baumgart B G. *Winged edge polyhedron representation*[M]. Stanford, California: Stanford University, 1972
- [158] Ma W, Kruth J P. NURBS curve and surface fitting for reverse engineering[J]. *The International Journal of Advanced Manufacturing Technology*, 1998, 14(12): 918-927
- [159] Lin W C, Chen T W. CSG-based object recognition using range images[C] // *Proceedings of the 9th International Conference on Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 1988: 99-103
- [160] Romero-Jarén R, Arranz J J. Automatic segmentation and classification of BIM elements from point clouds[J]. *Automation in Construction*, 2021, 124: Article No.103576
- [161] Siddiqui Y, Thies J, Ma F C, *et al.* RetrievalFuse: neural 3D scene reconstruction with a database[C] // *Proceedings of the IEEE/CVF International Conference on Computer Vision*. Los Alamitos: IEEE Computer Society Press, 2021: 12548-12557
- [162] Rausch C, Haas C. Automated shape and pose updating of building information model elements from 3D point clouds[J]. *Automation in Construction*, 2021, 124: Article No.103561
- [163] Kennedy J, Eberhart R. Particle swarm optimization[C] // *Proceedings of the International Conference on Neural Networks*. Los Alamitos: IEEE Computer Society Press, 1995: 1942-1948
- [164] Holland J H. Genetic algorithms[J]. *Scientific American*, 1992, 267(1): 66-73
- [165] Bertsimas D, Tsitsiklis J. Simulated annealing[J]. *Statistical Science*, 1993, 8(1): 10-15
- [166] Zhang Y X, Ren X H, Zhang J X, *et al.* A method for deformation detection and reconstruction of shield tunnel based on point cloud[J]. *Journal of Construction Engineering and Management*, 2024, 150(3): Article No.04024006
- [167] Mahmoud M, Chen W, Yang Y, *et al.* Automated BIM generation for large-scale indoor complex environments based on deep learning[J]. *Automation in Construction*, 2024, 162: Article No.105376
- [168] Yang L, Lin Y C, Cai H B, *et al.* From scans to parametric BIM: an enhanced framework using synthetic data augmentation and parametric modeling for highway bridges[J]. *Journal of Computing in Civil Engineering*, 2024, 38(3): Article No.04024008
- [169] Che E Z, Jung J, Olsen M J. Object recognition, segmentation, and classification of mobile laser scanning point clouds: a state of the art review[J]. *Sensors*, 2019, 19(4): Article No.810
- [170] Golub G H, van Loan C F. An analysis of the total least squares problem[J]. *SIAM Journal on Numerical Analysis*, 1980, 17(6): 883-893
- [171] von Gioi R G, Jakubowicz J, Morel J M, *et al.* LSD: a fast line segment detector with a false detection control[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 32(4): 722-732
- [172] Edelsbrunner H, Mücke E P. Three-dimensional alpha shapes[J]. *ACM Transactions on Graphics (TOG)*, 1994, 13(1): 43-72
- [173] Pătrăucean V, Armeni I, Nahangi M, *et al.* State of research in automatic as-built modelling[J]. *Advanced Engineering Informatics*, 2015, 29(2): 162-171
- [174] Yin Y D, Liu Y Z, Xiao Y, *et al.* SAI3D: segment any instance in 3D scenes[C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2024: 3292-3302
- [175] Xu R S, Wang X L, Wang T, *et al.* PointLLM: empowering large language models to understand point clouds[C] // *Proceedings of the 18th European Conference on Computer Vision*. Heidelberg: Springer, 2024: 131-147
- [176] 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*. Los Alamitos: IEEE Computer Society Press, 2024: 26407-26417