中图法分类号： **文献标识码：** A **文章编号：** 1006-8961(XXXX)XX-0001-22

论文引用格式： Ao Sheng, Wen Chenglu, Li Wen, Liu Dunqiang, Xing Leyuan, Li Mingzhe, Guo Yulan, Wang Cheng. XXXX. Deep learning for LiDAR point cloud processing: a survey. Journal of Image and Graphics, XX(XX):0001-0022(敖晟, 温程璐, 李文, 刘敦强, 邢乐园, 李明哲, 郭裕兰, 王程. XXXX. 激光雷达智能处理关键技术研究进展. 中国图象图形学报, XX(XX):0001-0022)[DOI:10.11834/jig.250664]

激光雷达智能处理关键技术研究进展

敖晟[1+],温程璐[1+],李文[1],刘敦强[1],邢乐园[1],李明哲[1],郭裕兰[2],王程[1]

1. 厦门大学信息学院,厦门 361005; 2. 中山大学电子与通信工程学院,深圳 518107

摘 要： 激光雷达作为三维环境感知的核心传感器,在自动驾驶、机器人、增强现实等领域发挥着不可替代的作用。随着人工智能技术的快速发展,激光雷达智能处理技术已成为研究热点。本文围绕三维目标检测、激光雷达定位、人体动作捕捉与语言推理四大关键任务,对国内外研究进展进行了系统梳理与深入分析。首先,本文总结了该领域的核心任务定义与关键挑战。其次,本文结合任务特性,对相关技术进行了系统分类与方法解析,深入比较各类方法在不同场景下的适用性与性能优势。本文提及的算法、数据集和评估指标已汇总至https://github.com/aosheng1996/DL4LiDAR。接下来,本文对国内外研究进展进行了对比分析,指出国外研究在模型体系与数据构建方面基础坚实,国内研究在算法效率与工程化落地方面发展迅速。最后,本文从算法融合、任务扩展与系统优化三个层面展望了激光雷达智能处理的未来发展趋势,以期为学术界与工业界提供理论参考,推动激光雷达智能处理技术的进一步发展。

关键词： 激光雷达;三维目标检测;激光雷达定位;人体动作捕捉;激光雷达语言推理

Deep learning for LiDAR point cloud processing: a survey

Ao Sheng[1+], Wen Chenglu[1+], Li Wen[1], Liu Dunqiang[1], Xing Leyuan[1], Li Mingzhe[1], Guo Yulan[2], Wang Cheng[1]

1. School of Informatics, Xiamen University, Xiamen, 361005, China; 2. School of Electronics and Communication Engineering, Sun Yat-sen University, Shenzhen, 518107, China

Abstract: LiDAR, as a high-precision 3D sensing technology, has become a cornerstone in a wide array of intelligent systems such as autonomous vehicles, robotics, and augmented reality. Compared to traditional RGB or RGB-D sensors, LiDAR offers superior performance in long-range depth estimation, illumination invariance, and structural scene understanding, particularly under challenging environmental conditions. With the surge in AI-driven perception, the intelligent processing of LiDAR point clouds has rapidly emerged as a key research frontier. This paper provides a systematic survey of recent developments in LiDAR-based perception technologies, focusing on four representative tasks: 3D object detection, LiDAR localization, human motion capture, and language-guided spatial reasoning. In the domain of 3D object detection, deep learning models have evolved along three primary lines: point-based, voxel-based, and multi-view approaches. Each presents unique trade-offs between geometric fidelity and computational efficiency. More recently, advanced architectures incorporating attention mechanisms, BEV representations, and multi-sensor fusion strategies have achieved significant improvements in accuracy and robustness. Label-efficient learning paradigms, such as semi-supervised, self-supervised, and domain adaptive learning, have also gained traction to mitigate the high cost of annotated 3D data. LiDAR-

收稿日期：2025-12-31;**修回日期：**2026-01-28

＊通信作者：共同一作 *通信作者:王程 cwang@xmu.edu.cn

基金项目：国家自然科学基金项目(62501502;42571514)

Supported by: National Natural Science Foundation of China(62501502;42571514)

based localization has progressed in both absolute and relative positioning tasks. Absolute localization methods rely on map-based retrieval or direct pose regression using neural networks, often enhanced by feature descriptors or transformer architectures. Relative localization, or LiDAR odometry, estimates frame-to-frame 13456motion and is fundamental to LiDAR SLAM systems. Research has expanded into geometry-aware learning, differentiable pose estimation, and multi-temporal consistency. Domestic studies have demonstrated strong progress in real-time, lightweight localization solutions through efficient model design and self-supervised learning frameworks, especially for edge deployment. Human motion capture using LiDAR addresses the challenge of estimating dynamic human poses in sparse, noisy point clouds. The field has evolved from single-frame pose regression to more advanced spatiotemporal modeling techniques that integrate SMPL body priors, inverse kinematics solvers, and transformer-based temporal encoders. Multi-modal fusion with IMU and vision sensors has further enhanced robustness in occluded or long-range scenes. The construction of large-scale datasets and task-specific benchmarks has greatly supported research and practical applications in surveillance, animation, and sports analytics. Language-driven LiDAR reasoning represents a novel and rapidly developing task that combines natural language understanding with spatial localization. Models are designed to infer 3D positions or regions in a point cloud scene based on descriptive language inputs. Pioneering frameworks like Text2Pos and Text2Loc adopt contrastive learning or coarse-to-fine alignment strategies, while newer approaches integrate scene graphs, multi-modal transformers, or large language models to enhance semantic comprehension. This direction supports applications in human-robot interaction, navigation, and open-world instruction following. The algorithms, datasets, and evaluation metrics mentioned in this paper have been summarized athttps://github. com/aosheng1996/DL4LiDAR. Comparative analysis of international and domestic research reveals complementary emphases: international efforts are characterized by systematic theoretical modeling, dataset construction, and general-purpose frameworks, while domestic work emphasizes computational efficiency, real-world deployment, and task-specific performance. Notably, Chinese research has made significant strides in lightweight model design, regression-based localization, and motion capture under sparse LiDAR input. Looking forward, the future of intelligent LiDAR processing lies in three major trajectories: (1) algorithmic fusion, involving unified representation spaces across point clouds, images, and language, enabling cross-modal semantic reasoning; (2) task expansion, pushing LiDAR perception beyond detection and mapping toward richer interaction, behavior understanding, and cognitive reasoning; and (3) system optimization, balancing accuracy, generalization, and efficiency for deployment in real-time, resource-constrained environments. Research in neural architecture search, unsupervised pretraining, and end-to-end multi-task learning will be instrumental in meeting these goals. In conclusion, LiDAR intelligent processing is rapidly evolving into a comprehensive and interdisciplinary research field, integrating geometric computation, deep learning, and cross-modal cognition. With ongoing advancements in algorithms, data, and hardware systems, LiDAR will continue to be central to building safe, interpretable, and generalizable 3D intelligent perception systems.

Key words: LiDAR; 3D object detection; LiDAR localization; human motion capture; language-driven LiDAR reasoning

0 引 言

激光雷达(light detection and ranging,LiDAR)作为一种主动式遥感技术,通过发射激光束并捕获其回波信号,能够直接、精确地获取周围环境的三维几何信息。近年来,随着自动驾驶、机器人技术以及增强现实等领域的迅猛发展(Tang等,2023),LiDAR凭借其高精度、抗干扰能力强及不受光照条件影响等优势,已成为环境感知与理解的核心传感器之一(Tan等,2025)。海量三维点云数据的产生,极大地推动了激光雷达智能处理技术的进步,使其成为计算机视觉与机器人学交叉领域的前沿研究方向。

然而,LiDAR点云数据本身存在着无序、稀疏以及密度不均等固有特性,这对智能信息处理技术带来了严峻挑战。早期方法多依赖于手工设计的特征和传统的几何模型,虽在特定场景下有效,但泛化能力有限,难以应对复杂多变的应用环境。随着深度学习技术在视觉领域的突破,研究者们开始致力于将其迁移至三维点云数据处理任务中。这一趋势催生了激光雷达智能处理技术的革命性变革,研究方向也从初期的目标检测任务逐步拓展至全局定位、人体动作捕捉乃至新兴的语言感知等多个层面。

当前,激光雷达智能处理技术已成为推动自动

驾驶、智能机器人和增强现实等技术发展的关键驱动力。在自动驾驶领域，激光雷达的高精度感知能力是实现L4/L5级自动驾驶的必要条件；在机器人领域，激光雷达的三维环境建模能力是实现自主导航与操作的基础；在增强现实领域，激光雷达的精确空间定位能力为虚拟内容与真实环境的无缝融合提供了可能。然而，尽管技术发展迅速，激光雷达智能处理仍面临诸多挑战，包括数据处理效率、环境鲁棒性、标注成本以及跨场景泛化能力等问题。面向上述挑战与需求，本文将对激光雷达智能处理技术的研究进展进行系统性的梳理与综述。文章将重点围绕三维目标检测、激光雷达定位、激光雷达人体动作捕捉以及激光雷达语言推理任务，深入阐述每种任务的国内外研究现状，对比分析不同技术路线的特点与趋势，并展望未来的发展方向。图1展示了当前激光雷达智能处理关键技术的分类体系。通过对现有成果的归纳总结，旨在为相关领域的研究者提供清晰的技术发展脉络，推动激光雷达智能处理技术的进一步创新与应用。

1 背　景

1.1 三维目标检测

三维目标检测是环境感知的基石，为后续的轨迹预测、路径规划等模块提供输入（Li等，2024）。该任务旨在从单帧或多帧LiDAR点云中，自动识别出感兴趣的物体实例（如车辆、行人、骑行者），并预测每个实例在三维空间中的精确位置、空间尺寸（长宽高）和朝向角，如图2所示。然而，三维目标检测面临多重挑战：首先，点云的稀疏性与密度不均导致目标特征提取困难，尤其在远距离或遮挡场景下；其次，动态环境中的目标运动与形状变化增加了检测的复杂性；最后，点云标注成本高昂，限制了大规模训练数据的获取。此外，恶劣天气条件（如雨、雪、雾）会进一步降低点云质量，影响检测精度。这些挑战要求算法在保持高精度的同时，具备良好的实时性与环境适应性。

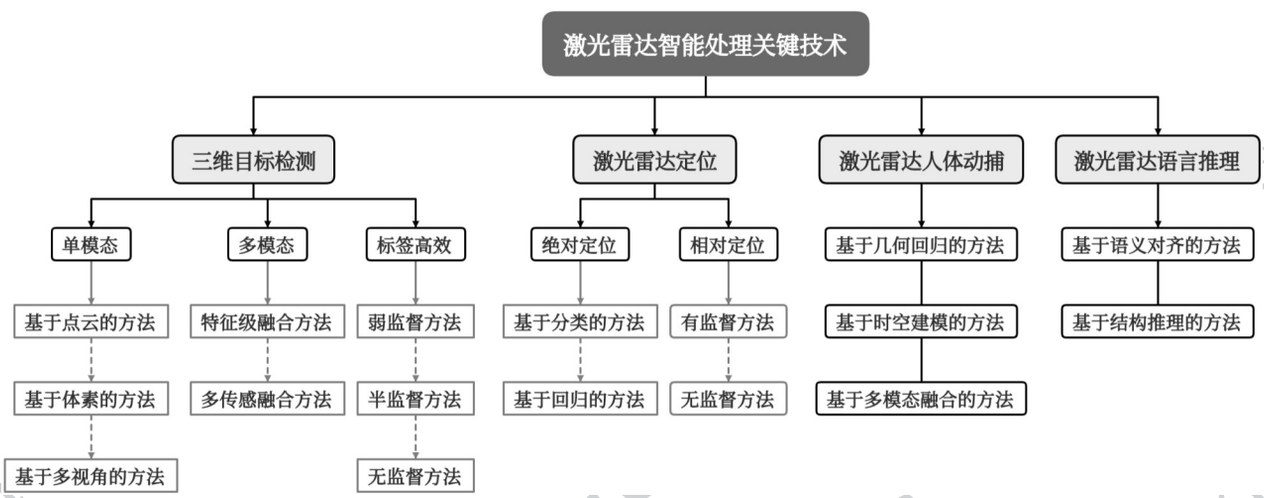

图1　激光雷达智能处理技术的分类体系

Fig. 1　A taxonomy of intelligent processing techniques for LiDAR

图2　三维目标检测示意图

Fig. 2　A illustration of 3D object detection

1.2 激光雷达定位

定位问题是自动驾驶、机器人技术和增强现实等领域的核心挑战，其本质是回答“我在哪里”这一基本问题（Guo等，2025）。与以相机为主要传感器的定位方法相比，激光雷达在复杂光照条件下能够提供更加稳定和鲁棒的环境观测（如图3所示）。因此，激光雷达定位技术已经成为当前的主流定位方案之一，其可以消除定位系统对基于全球卫星导航系统（global navigation satellite system，GNSS）和惯性导航系统（inertial navigation system，INS）的依赖性，

提供全天时、稳定的定位能力，在自动驾驶等应用场景中确保长期可靠的定位精度。此任务面临的挑战错综复杂：动态环境中移动的车辆、行人会污染本应用于定位的静态场景几何，引入估计误差；季节更替、天气变化、光照转换会导致同一地点的点云外观发生显著改变，对定位系统的长期鲁棒性与泛化能力提出严峻考验；此外，不依赖高精度先验地图的定位方法需从数据中学习场景本质，如何避免过拟合特定轨迹、实现跨场景的泛化，仍是一个开放问题。

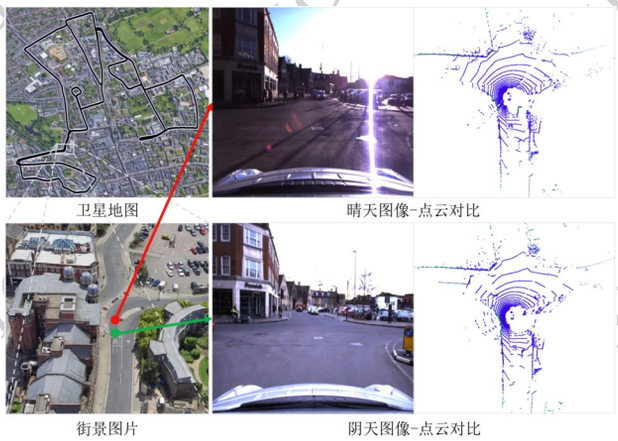

图3　不同光照条件下同一地点相机与LiDAR观测结果对比

Fig. 3　Comparison between image-based and LiDAR-based localization

1.3　激光雷达人体动作捕捉

激光雷达人体动作捕捉旨在利用LiDAR的高精度测距能力，实现对人体三维姿态的实时重建与跟踪的任务（Cheng等，2022）。该技术通过分析点云数据中人体各关节的位置变化，构建人体动作的三维模型，为虚拟现实、人机交互、安防监控等应用提供技术支持。与传统视觉方法相比，激光雷达人体动作捕捉不受光照条件影响，能够提供更精确的深度信息，尤其适用于室内场景和低光照环境。然而，激光雷达人体动作捕捉面临显著挑战：首先，点云稀疏性导致人体关键部位（如手部、面部）的细节信息不足；其次，人体相互遮挡或与环境物体遮挡严重影响姿态估计精度（如图4所示）；最后，不同人体形态与动作模式的差异性要求算法具有良好的泛化能力。

1.4　激光雷达语言推理

作为一项新兴的交叉模态任务，激光雷达语言推理旨在通过语言指令理解环境并执行相应操作。

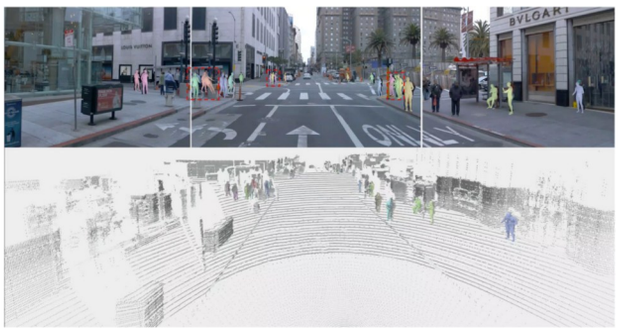

图4　远距离激光雷达人体动作捕捉效果不理想

Fig. 4　Long-range LiDAR-based human motion capture shows inferior performance

该任务要求系统能够根据文本描述（如“找到红色的汽车”）在点云数据中准确定位目标，并理解其空间关系（如“汽车在树的左边”）。这一技术是实现人机交互与智能环境理解的关键，为机器人执行复杂任务提供了新的可能性（Feng等，2025）。激光雷达语言推理面临的核心挑战包括：首先，语言与视觉模态之间的语义鸿沟使得跨模态对齐困难；其次，场景的复杂性与多样性要求模型具备强大的泛化能力。这些挑战要求算法能够同时处理多模态信息，并在语义层面实现深度理解与精确匹配。

1.5　数据集

为了支撑上述四项任务的算法研发与评估，研究人员构建了一系列具有代表性的基准数据集，如表1所示。在三维目标检测领域，数据集从早期的KITTI发展到如今的nuScenes和Waymo Open，其点云规模从数万帧激增至数十万帧，覆盖场景与天气条件也更为丰富。在激光雷达定位领域，Oxford Radar RobotCar与NCLT（north campus long-term dataset）等数据集通过提供长时间、大规模的城市与校园连续序列，为研究定位系统的长期鲁棒性提供了重要基准。新兴的激光雷达人体动作捕捉数据集（如LiDARHuman26M、HSC4D（human-centered 4d scene capture））与语言推理数据集（如KITTI360 Pose、CityRefer），则分别从精细三维姿态标注和密集跨模态关联两个维度填补了空白，为算法理解复杂场景语义与人体动态开辟了新途径。这些数据集的公开极大促进了算法公平比较、可复现性研究与技术迭代，构成了当前激光雷达感知与理解研究的重要基石。

2 国内外研究现状

2.1 三维目标检测

2.1.1 基于单模态的三维目标检测

激光雷达点云能够精确获取周围环境的三维信息，但存在固有的不规则性与稀疏性问题。针对这一问题，目前主要有三种解决思路：直接处理原始点云、将点云划分为体素或转换为多视角二维图像进行处理。

1）基于点云的方法

该类方法直接提取原始点云数据的特征，可最大程度保留点的精确位置信息，核心思路是通过对称函数或注意力机制聚合点状特征。Qi等人（2017）提出的PointNet++是该方向的奠基性工作，通过分层采样与分组策略实现点云的分层特征学习，解决了PointNet无法捕捉局部特征的缺陷，成为后续基于点的检测方法的基础框架。在此基础上，Xie等人（2021）提出的VoteNet引入投票机制来预测实例的中心点坐标，通过多轮投票优化目标定位精度，在室内场景小目标检测中表现优异。Shi等人（2019）构建了一种两阶段的端到端框架PointRCNN，通过点云分割生成候选框并细化框内特征，摆脱了传统方法对人工设计候选框的依赖，成为国内该领域的代表性工作。然而，基于点的方法存在固有局限：无序点云的处理依赖大量计算资源，导致模型推理速度慢，难以满足自动驾驶实时性需求。

2）基于体素的方法

为解决点云处理效率问题，该类方法将点云按固定分辨率划分为体素网格，通过三维卷积或稀疏

表1 现有的三维目标检测、激光雷达定位、激光雷达人体动捕和激光雷达语言推理数据集总结

Table 1 A summary of existing datasets for 3D object detection, LiDAR-based localization, LiDAR-based human motion capture, and LiDAR-based language reasoning

三维目标检测数据集					
数据集	**年份**	**场景**	**点云数量**	**传感器**	**网址**
KITTI	2012	城市	14,999	激光雷达+相机	https://www.cvlibs.net/datasets/kitti/eval_object.php?obj_benchmark=3d
nuScenes	2019	城市	400,000	激光雷达+相机+毫米波雷达	https://www.nuscenes.org/
Waymo Open	2020	城市/郊区	23,000	激光雷达+相机	https://waymo.com/open/
激光雷达定位数据集					
数据集	**年份**	**场景**	**长度**	**传感器**	**网址**
Oxford Radar RobotCar	2020	城市	10km	激光雷达+相机+毫米波雷达	https://oxford-robotics-institute.github.io/radar-robotcar-dataset/
NCLT	2016	校园	5.5km	激光雷达+相机	https://robots.engin.umich.edu/nclt/
激光雷达人体动作捕捉数据集					
数据集	**年份**	**场景**	**点云数量**	**传感器**	**网址**
LiDARHuman26M	2022	室外	184,048	激光雷达	http://www.lidarhumanmotion.net/lidarcap/
HSC4D	2022	室外/室内	50,000	激光雷达	http://www.lidarhumanmotion.net/data-hsc4d/
CIMI4D	2023	室内攀岩	~180,000	激光雷达+相机	http://www.lidarhumanmotion.net/cimi4d/
激光雷达语言推理数据集					
数据集	**年份**	**场景**	**点云数量**	**文本数量**	**网址**
KITTI360Pose	2022	城市	14,934	43,381	https://text2pos.github.io/
CityRefer	2023	城市	~5,000	~35,000	https://github.com/ATR-DBI/CityRefer
CityAnchor	2025	城市	1,448	1,448	https://github.com/WHU-USI3DV/CityAnchor

卷积提取特征，实现结构化处理。国际上的开创性工作包括Zhou等人（2017）提出的VoxelNet，这是首个端到端基于体素的检测器，通过体素特征编码模块学习体素特征，再利用3D卷积生成目标候选框，从而大幅提升处理效率。Yang等人（2018）将激光雷达点云投影至鸟瞰图（bird′s-eye-view，BEV）并通过CNN实现三维目标检测，其计算成本仅与体素总数相关，在车载硬件上可实现实时推理。Yan等人（2018）采用稀疏卷积处理体素特征，并引入特征金字塔增强多尺度感知，在车载GPU上实现20 FPS的实时推理。针对恶劣天气下的性能退化问题，Huang等人（2024）利用多教师知识蒸馏聚合邻车共享特征，提升雨天条件下的检测鲁棒性，填补了国内极端天气单模态检测的技术空白。然而，该类方法存在体素稀疏性问题：远处或小目标对应的体素多为空，导致特征信息丢失，检测精度下降。

3）基于多视角的方法

该类方法将3D点云投影为BEV或前视图，借助成熟的CNN进行特征提取，在效率与精度之间实现了良好平衡。一个代表性工作是Meyer等人（2019）提出的LaserNet，其在常规2D检测器基础上引入多尺度特征融合模块，通过BEV与前视图的特征互补增强检测鲁棒性，并采用概率建模量化检测不确定性，提升复杂场景下的适应能力。国内研究方面，Deng等人（2020）通过稀疏体素金字塔与分层RoI池化，在保证精度的同时降低模型复杂度，适配低算力车载平台。在协同感知方面，Li等人（2023）基于多教师知识蒸馏实现跨车特征聚合，进而提升信息交互效率与检测性能。此外，Wu等人（2023）提出的VirConvNet通过抗噪子流形卷积与随机体素丢弃机制，增强模型对深度补全噪声和低纹理区域的鲁棒性，显著改善遮挡目标的检测效果。尽管多视角方法在效率与精度之间取得了平衡，但投影过程不可避免地导致三维信息丢失，在拥挤场景中仍易出现漏检或误检。

2.1.2　基于多模态融合的三维目标检测

由于单模态点云存在稀疏性且缺乏纹理信息，许多研究采用LiDAR与图像、LiDAR与雷达或多传感器协同的融合策略，以提升感知系统的鲁棒性。融合方法正从简单的特征拼接向深度语义对齐演进，并呈现出向统一BEV表征发展的趋势。

1）特征级融合方法

早期研究侧重于简单的特征拼接，核心思路是分别提取单模态特征，再通过拼接或加权实现融合。国际上的经典工作包括Chen等人（2017）提出的MV3D（multi-view 3d object detection network），将图像特征与LiDAR的BEV特征、前视图特征分别提取后，在BEV空间进行特征拼接，利用不同视角信息互补提升检测精度。随后，Ku等人（2018）进一步提出感兴趣区域（region of interest，RoI）级融合策略，在目标候选框区域内联合提取点云与图像局部特征，实现细粒度交互，解决了全局融合中冗余信息干扰的问题。Vora等人（2020）提出的PointPainting是里程碑式工作，通过语义分割掩码为每个LiDAR点附加语义标签，实现点级特征增强，成为后续多模态融合的基准框架。

国内研究在多模态融合方面同样取得了显著进展。Xu等人（2021）通过自适应注意力机制将图像语义特征动态注入点云，实现了高效的纹理补充，但对点云本身的稀疏性问题改善有限。在RoI级融合方面，Cai等人（2023）首先生成空间对齐的体素、BEV和图像特征，经模态特定的上下文编码后在对象级别进行融合，避免了复杂的跨模态对齐，有效提升了检测精度。近年来，多模态融合研究逐步迈向深度耦合与结构创新。国际上的代表性工作如Liu等人（2022）采用独立骨干网络处理图像与点云，将两者均转换为BEV特征后进行加权融合，避免了跨模态变换的信息丢失，显著提升了复杂场景下的检测精度。Bai等人（2022）提出的TransFusion引入Transformer融合模块，该方法首先基于LiDAR生成初始检测候选框，再以查询机制自适应地关联图像特征，实现了目标层级的动态融合，在严重遮挡场景中展现出更强的鲁棒性。Wang等人（2021）提出的PointAugmenting则通过跨模态特征增强策略，将图像信息作为辅助信号动态注入点云特征，在不显著增加模型复杂度的情况下提升了点云语义表达能力。

2）多传感器协同融合方法

除激光雷达与图像融合外，国际研究进一步拓展至雷达、红外等传感器的多模态协同。Chen等人（2023）提出的FUTR3D构建了多模态统一查询空间，融合多视角图像、稀疏激光雷达点云与三维雷达数据，利用多尺度上下文信息优化目标边界框预测，在雨雾天气下的检测精度较双模态方法提升20%。

国内研究方面，Cao等人(2025)提出的RAFDet融合图像与雷达信息，利用雷达射频(RF)图像增强空间细节，并通过雷达点云提供的精确深度信息校正视觉深度预测。然而，这类方法的深度估计精度仍存在局限。在协同感知领域，Zhao等人(2023)在BEV空间中融合激光雷达与图像特征，显著增强了检测的鲁棒性。此外，Xie等人(2024)提出的DenseSeq-Fusion采用前景掩码引导的深度补全策略，将图像转换为稠密且精确的虚拟点云用于三维检测。然而，该方法高度依赖高质量的深度补全模型，在图像纹理匮乏区域(如纯色车身)易产生补全误差，进而导致检测框偏移。Huang等人(2025)发布了首个融合激光雷达、摄像头与4D雷达的模拟数据集V2X-R(vehicle-to-everything-radar)，为多模态协同感知研究提供了重要支持。

2.1.3 标签高效的三维目标检测

三维点云标注成本极高，成为算法落地的核心瓶颈。很多工作围绕减少标注依赖核心问题，聚焦弱监督、半监督、无监督方法研究，实现从减少标注量向摆脱标签依赖的技术转变。

1)弱监督检测方法

弱监督目标检测假设训练集无实例级边界框标注，仅利用点标注、目标级标签等少量监督信号。国际上的早期工作包括Abrar等人(2018)提出的3D重建辅助检测思路，通过学习带纹理的3D重建模型，从重建结果中提取目标结构特征，间接辅助检测任务，在室内场景中实现无框标注的检测。Gojcic等人(2021)进一步简化标注需求，将仅含类别信息的目标级标签作为监督信号，通过点云聚类与几何约束生成伪边界框，大幅降低了标注成本，但定位精度偏低。国内研究方面，Chen等人(2015)提出的3DOP(3d object proposals)是早期降低三维标注依赖的代表性工作，通过融合立体视觉与激光雷达信息生成三维候选框。该方法无需端到端三维监督，为后续标签高效与多模态三维检测研究奠定了基础。Liu等人(2019)提出Pseudo-LiDAR方法，通过图像深度估计将二维视觉信息重建为伪点云，并直接采用LiDAR检测框架完成三维目标检测。该方法在无需真实激光雷达标注的情况下显著提升检测性能，但对深度估计误差较为敏感。针对伪标签质量问题，Ding等人(2020)提出的SDFLabel(signed distance field labeling)方法利用符号距离函数(signed distance field，SDF)对目标表面进行几何建模，在无需人工三维标注的情况下自动生成高质量伪标签，用于训练三维检测模型，为几何先验驱动的弱监督检测提供了新的解决思路。Xia等人(2023)提出的CoIn(contrastive instance mining)仅需每场景标注一个实例，通过对比实例特征挖掘与空间分布分析生成伪标签，使用2%标注量即达全监督性能，展现了弱监督学习的潜力。

2)半监督检测方法

半监督目标检测旨在利用少量标注数据与大量未标注数据共同训练神经网络模型。现有方法主要分为两类：一类是一致性正则化方法，通过对未标注数据施加变换(如点云旋转、缩放等)，约束模型在不同扰动下输出一致的预测结果。Tarvainen等人(2017)提出的平均教师框架是该类方法的经典范式，通过学生模型进行梯度更新、教师模型采用指数移动平均(exponential moving average，EMA)更新权重的方式实现一致性学习，已被广泛应用于三维目标检测任务。Singh等人(2018)在此基础上引入尺度不变性约束，进一步提升了模型对多尺度目标的检测稳定性。另一类是伪标签方法，其核心思想是利用教师模型为未标注数据生成伪标签，进而指导学生模型的训练。国际上的代表性工作如Wang等人(2021)引入交并比(intersection over union，IoU)预测机制，通过筛选高置信度的伪标签以过滤低质量预测，在KITTI数据集仅使用1%标注数据的情况下，三类物体的检测性能已超越全监督基线。国内研究方面，Wang等人(2025)同样引入IoU学习动态过滤低置信度预测，提升训练稳定性。Liu等人(2021)采用多类别Focal损失替代传统的交叉熵损失，有效缓解了伪标签过程中存在的类别不平衡问题，展现出强大的学习效率与泛化能力。

3)无监督检测方法

无监督目标检测无需任何人工标注数据，通过自监督预训练学习通用特征，并将其迁移至下游检测任务。国际上的开创性工作包括Sanghi等人(2020)采用互信息最大化策略学习点云的判别性表示，预训练模型在迁移到检测任务后，显著提升了对小目标的检测能力。Afham等人(2022)引入跨模态对比学习，利用图像与点云之间的语义关联进行自监督预训练，使模型在多个数据集上展现出良好的泛化性能。Gadelha等人(2020)通过划分点云并构

建正负样本对，利用对比学习框架提取有效特征。PointContrast（Xie等，2020）作为早期代表性工作，将对比学习与重建任务相结合，进而在未标注数据上学习鲁棒的三维特征，成为后续研究的重要基准。Huang等人（2021）从实例判别视角出发，利用多视角和多层次特征增强对比学习的多样性；Henzler等人（2021）提出耦合多实例检测网络，通过源域与目标域间的特征对齐，实现跨场景下的无标注检测。Chen等人（2024）提出一种基于偏最小二乘的增强策略，利用高置信度伪标签中的尺寸信息对源域数据进行几何缩放，无需目标域标注即可提升跨域检测性能。国内在无监督方向的研究也取得了重要进展，Wu等人（2024）提出的CPD（common-sense prototype-based detection）利用常识原型与运动先验实现无监督检测，在KITTI和WOD（waymo open）数据集上接近全监督性能。但少数类别（如骑自行车者）因实例稀少，检测精度仍显著偏低，未来需加强稀有类别数据采集以进一步提升模型泛化能力。

2.1.4　总结

综合表2中各方法在主要公开数据集上的性能表现，并结合当前研究进展，可得出以下结论：

1）基于点云的方法能够实现较高的检测精度，但受限于点云的无序性与非结构化特性，其计算复杂度较高，难以满足实时性要求。相比之下，基于体素化与多视角投影的方法更适用于对实时性要求严格的场景。

2）多模态融合方法在复杂场景的数据集上，其综合性能指标普遍优于单一模态方法。当前技术发展呈现出从早期简单的特征拼接，向基于统一鸟瞰图表征或Transformer架构的深度、自适应融合方向演进的趋势。

3）部分先进方法仅需少量标注甚至无需人工标注，即可达到接近全监督基准模型的性能。然而，此类方法在数据中分布稀疏的类别上，其泛化能力与检测精度仍有待进一步提升。

2.2　激光雷达定位

根据先验信息的使用方式和输出结果的性质，激光雷达定位问题通常可分为两类：相对定位与绝对定位。相对定位旨在从已知初始位姿出发，估计并追踪系统的连续运动（Bresson等，2017）。与之相反，绝对定位则是在没有先验信息的情况下，直接确定系统在世界坐标系中的全局位姿（Lu等，2014）。当系统位姿完全未知，或相对定位中的连续位姿估计丢失时，绝对定位就显得尤为必要。这两种关键能力相互补充，共同构成了鲁棒定位系统的基石。

2.2.1　绝对激光雷达定位

绝对激光雷达定位的目标是从单帧激光雷达扫描点云中直接估计传感器在全球坐标系中的6自由度位姿，包括三维位置和三维方向。根据实现方式的不同，该类方法可以进一步分为基于分类的方法和基于回归的方法两大范畴。

1）基于分类的方法

基于分类的方法（Uy等，2018；Ao等，2021；Xia等，2023）通常以检索或位置识别的形式实现，将定位问题视为离散匹配问题。这类方法首先将环境离散化为已知位置或特征组成的数据库，构建一个结构化的场景表示。在推理阶段，通过提取实时扫描数据的特征描述子，并在预先构建的数据库中进行相似性匹配，找到最佳对应的场景位置，从而确定系统的全局位姿（Ao等，2022；Luo等，2023）。尽管这种方法在某些结构化环境中表现良好，但其定位精度受限于数据库的离散化粒度，且在大规模环境中面临着存储和计算效率的挑战（Ao等，2023）。

2）基于回归的方法

基于回归的方法（Kendall等，2015）将定位问题构建为连续的函数逼近问题。这类方法通过学习从传感器输入到位姿参数或其相关几何实体的连续映射，无需显式的搜索步骤即可直接回归目标值。这种端到端的学习范式不仅提高了定位效率，还避免了离散化带来的信息损失，因此受到研究者的广泛关注。

在国际研究中，绝对位姿回归（absolute pose regression，APR）作为代表性的回归方法，其发展深受计算机视觉领域先驱工作的影响（Wang等，2023；Goswami等，2025）。PoseNet（Kendall等，2015）首次证明了使用卷积神经网络直接从图像回归相机位姿的可行性，为激光雷达领域的相关研究提供了重要启示。国内研究者在绝对位姿回归方面同样做出了重要贡献。早期工作PosePN（Yu等，2022）创新性地提出了通用编码器与场景特异性回归器相结合的策略，这一设计巧妙地平衡了模型的泛化能力和场景适应性。STCLoc（Yu等，2023a）首次引入了时空约束，代表了从单帧建模到序列理解的重要转变。该方法利用注意力机制处理连续点云序列，不仅有效

表2　三维目标检测方法在公开数据集上的综合性能分析

Table 2　A comprehensive performance analysis of various 3D object detection methods across different datasets

类别	方法	传感器模态	运行时间(ms)	KITTI Car (容易/适中/困难)	nuScenes (mAP / NDS)	Waymo Vehicle (L1 / L2)
单模态	PointRCNN(Shi等,2019)	LiDAR	100	85.94 / 75.76 / 68.32	–	–
	Point-GNN(Shi等,2020)	LiDAR	640	**88.33 / 79.47 / 77.29**	–	–
	VoxelNet(Zhou等,2018)	LiDAR	220	77.47 / 65.11 / 57.73	–	–
	SECOND(Yan等,2018)	LiDAR	50	83.13 / 73.66 / 66.20	–	52.11 / –
	LaserNet(Meyer等,2019)	LiDAR	30	–	–	–
多模态	MV3D(Chen等,2017)	Fusion	240	71.09 / 62.35 / 55.12	**70.2 / 72.9**	–
	BEVFusion(Liu等,2022)	Fusion	–	–	46.4 / 58.1	–
	PointPainting(Vora等,2020)	Fusion	–	82.11 / 71.70 / 67.08	66.5 / 70.7	–
	FusionPainting(Xu等,2021)	Fusion	–	–	66.8 / 71.0	**67.41 / 62.70**
	PointAugmenting(Wang等,2021)	Fusion	542	–	–	
标签高效	CPD++ (Wu等,2024)	LiDAR	100	84.20 / 67.90 / 62.53	–	–
	SDFLabel (Ding等,2020)	Camera	–	–	–	–
	Pseudo-LiDAR (Liu等,2019)	Camera	–	54.53 / 34.05 / 28.25	–	–
	3DOP (Chen等,2015)	Camera	–	–	–	–

注:加粗字体表示各类方法中各列最优结果,"–"表示数据未提供。

捕捉了帧间的运动连续性,还通过时空一致性解决了单帧观测中的歧义性问题。受神经科学启发的NIDALoc(Yu等,2023b)展现了国内研究在跨学科融合方面的创新思维。该方法模仿大脑中的位置细胞、头朝向细胞和网格细胞的功能机制,构建了一个更具生物合理性的定位框架。DiffLoc(Li等,2024)创新性地将扩散模型引入位姿回归任务,通过迭代的去噪过程从随机初始化的位姿逐步生成精确的位姿估计。

场景坐标回归(scene coordinate regression,SCR)作为另一种重要的技术路线,通过引入中间表示层在精度和效率之间取得了更好的平衡。SCR方法首先利用神经网络预测输入点云中每个点在全局坐标系中的对应坐标,建立密集的3D-3D对应关系,然后通过鲁棒的位姿求解算法如随机样本一致(random sample consensus,RANSAC)计算最终的6-DoF(degrees of freedom)位姿。这种两阶段的方法既利用了深度学习的强大表示能力,又保留了传统几何方法的精度优势。

DSAC(differentiable sample consensus)系列工作(Eric等,2021,2023)在SCR发展中具有里程碑意义,首次将可微分的随机样本一致性算法引入视觉重定位任务,建立了一个完整的"学习+优化"端到端训练框架。该框架允许梯度通过RANSAC模块反向传播,使得网络能够学习更适合位姿估计的特征表示。这一创新为后续研究奠定了基础,推动了SCR方法的快速发展。近年来,Transformer架构(Ashish等,2017)在SCR中的应用进一步提升了模型性能。有研究工作(Wang等,2024)引入了分层级的场景坐标分类与回归策略,结合Transformer的自注意力机制,有效捕捉了点云中长距离的几何依赖关系,增强了对复杂场景的结构理解能力。这种方法在处理大

规模室外环境时表现出色，特别是在具有重复结构和相似几何特征的环境中。

国内在场景坐标回归方面同样取得了显著进展，形成了一系列具有特色的技术路线。SGLoc（Li等，2023）作为开创性工作，专注于构建强大的场景几何编码器，提出了多层次的特征融合机制和几何结构感知模块。为了满足实际应用中对效率的迫切需求，LightLoc（Li等，2025）通过精心设计的轻量级网络架构和高效的回归策略，大幅降低了计算开销和内存占用，为边缘计算平台的部署奠定了基础。RALoc（Yang等，2025）针对旋转估计这一长期挑战提出了专门的解决方案。该方法通过显式地建模方向变化，引入了旋转感知机制，包括旋转等变的特征学习和方向敏感的距离度量。

表3　Oxford数据集上的平均平移误差（m）与旋转误差（°）

Table 3　Average translation error（m） and rotation error（°） on the Oxford dataset

类别	方法	15-13-06-37	17-13-26-39	17-14-03-00	18-14-14-42	Average [m/°]
APR	PointLoc（Wang等，2022）	12.42/2.26	13.14/2.50	12.91/1.92	11.31/1.98	12.45/2.17
	NIDALoc（Yu等，2023b）	5.45/1.40	7.63/1.56	6.68/1.26	4.80/1.18	6.14/1.35
	HypLiLoc（Wang等，2023）	6.88/1.09	6.79/1.29	5.82/0.97	3.45/0.84	5.74/1.05
	DiffLoc（Li等，2024）	3.57/**0.88**	3.65/**0.68**	4.03/**0.70**	2.86/**0.60**	3.53/**0.72**
SCR	SGLoc（Li等，2023）	3.01/1.91	4.07/2.07	3.37/1.89	2.12/1.66	3.14/1.88
	LightLoc（Li等，2025）	**2.33**/1.21	**3.19**/1.34	**3.11**/1.24	**2.05**/1.20	**2.67**/1.25
	RALoc（Yang等，2025）	3.19/4.10	3.87/3.96	3.32/3.87	2.59/3.71	3.24/3.91

注：加粗字体表示各类方法中各列最优结果。

2.2.2　相对激光雷达定位

相对激光雷达定位，即激光雷达里程计（lidar odometry，LO），专注于估计相邻两帧激光雷达扫描之间的相对运动变换。作为同时定位与建图（simultaneous localization and mapping，SLAM）系统的核心组件，LO在自动驾驶、移动机器人等领域具有重要应用价值。根据训练过程中监督信号的来源，LO方法可分为有监督学习和无监督学习两大类，每种范式各有其优势和适用场景。

1）有监督学习方法

监督学习范式依赖于精确的位姿真值标签，通过端到端的方式学习从点云数据到相对位姿参数的映射函数。国际上的代表性工作DeepPCO（Wang等，2019）在这一方向做出了重要贡献，提出了并行神经网络结构，分别从单帧点云中学习空间结构特征和从连续帧对中学习时间一致性。这种双流架构显式地建模了点云的空间特性和运动连续性，相较于单流基线显著提高了精度，特别是在快速运动和大旋转变化的场景中。LodoNet（Zheng等，2020）则探索了不同的技术路径，通过在距离图像上引入2D关键点检测与匹配机制，将成熟的图像特征匹配思想适配到激光雷达数据处理中。该方法首先在2D投影空间检测稳定的关键点并建立对应关系，然后将这些匹配投影回3D空间，通过几何一致性约束优化相对位姿估计。这种混合方法结合了深度学习的特征学习能力和传统几何优化的精度优势，实现了更加稳定的里程计估计。

国内在激光雷达里程计方面的研究工作主要聚焦于有监督学习的LO方法，取得了系列重要成果。LO-Net（Li等，2019）作为早期具有影响力的代表作，建立了将点云投影至距离图像并使用卷积神经网络进行端到端位姿估计的基础范式。PWCLO-Net（pyramid，warping，and cost volume lidar odometry network）（Wang等，2021）采用分层嵌入结构和掩码优化技术，实现了在3D点云中的精细运动对齐。EfficientLO（Wang等，2022）在保持高精度的同时满足了实时性要求，通过优化网络结构和计算流程，显著降低了计算开销。TransLO（Liu等，2023）创新性地将窗口掩码点Transformer架构引入LO任务。该方法通过自注意力机制有效融合了局部几何上下文和长程依赖关系，提升了在复杂动态环境中的鲁棒性。PWDLO（probability-weighted diffusion lidar odometry）（Lu等，2025）和DiffLO（Huang等，2025）展现了国内在生成模型用于里程计细化方面的最新进

展。PWDLO探索了条件扩散模型在里程计中的应用，通过由粗到精的迭代优化过程逐步改善位姿估计。DiffLO则进一步将语义感知与扩散模型相结合，通过多尺度特征编码和条件去噪过程，显著提升了运动估计的精度和鲁棒性。

2)无监督学习方法

尽管监督学习方法取得了令人瞩目的性能，但其对精确位姿真值的依赖限制了在真实场景中的大规模应用。为了突破这一限制，研究者开始探索自监督和无监督学习范式，致力于从数据本身挖掘监督信号。

在自监督/无监督学习LO方面，国际研究呈现出多样化的技术路线。Cho等人(2020)的开创性工作证明了仅基于点云几何一致性约束的无监督学习的可行性。通过施加点对点和点对面距离约束，该方法强制相邻帧的点云在估计的位姿变换下实现最佳对齐，从而为网络训练提供自我监督信号。这种基于几何一致性的方法虽然简单有效，但在动态环境中容易受到运动物体的干扰。Nubert等人(2021)进一步推进了自监督学习的研究，提出了一个通用的训练框架，结合了时序循环一致性和扫描到扫描的对齐策略。该方法通过构建更长的时序约束，增强了运动估计的平滑性和一致性，在精度和计算效率之间取得了良好平衡。特别值得一提的是，该框架对不同的点云编码网络保持兼容性，为后续研究提供了灵活的基线系统。

2.2.3 总结

综合本节内容及表3性能对比，可得出以下结论：

1)在绝对定位中，基于检索的方法受限于场景表征的离散化粒度，而基于回归的方法在定位精度与场景适应性之间取得了较优的平衡。

2)在相对定位方面，无监督/自监督学习范式显著降低了对人工标注数据的依赖，是推动相关技术落地的重要方向。

3)现有方法在极端天气、高度动态或结构重复的大规模场景中，其精度与可靠性仍面临考验。

2.3 激光雷达人体动捕

2.3.1 基于几何回归的方法

LiDAR人体动捕技术早期主要集中在对单帧稀疏点云数据的人体检测与关键点回归，方法上多依赖于传统几何特征提取与骨架预测模块，尚未充分考虑跨时间帧之间的动态一致性。Furst等人(2021)提出的HPERL(human pose estimation using rgb and lidar)系统通过融合RGB图像与LiDAR点云，在自动驾驶场景中实现了精确的三维姿态估计。该系统结合多模态输入，提高了点云稀疏条件下的人体检测准确率，为后续融合方法提供了理论支撑与工程实践基础。然而，该类方法局限于单帧静态分析，对复杂场景中运动连续性和交互物理性的建模仍显不足。随着技术演进，国内研究逐渐向更具结构约束的回归模型拓展。LiDARCap(lidar-based motion capture)(Li等，2022)则突破了以往依赖标记或RGB输入的限制，实现了基于单一激光雷达设备的无标记人体姿态估计。该系统利用点云特征编码器提取局部几何信息，联合逆运动学求解器与蒙皮多人线性模型(skinned multi-person linear model，SMPL)，在远距离(超过15米)和高稀疏环境下仍能够稳定重建三维姿态。其后续工作LiDARCapV2(Zhang等，2024)进一步引入了人-物交互建模，通过构建人物接触图与物理约束，提升了交互场景下的姿态估计准确性。

2.3.2 基于时空建模的方法

随着对时序连续性和动态建模需求的提升，研究开始引入序列建模机制，以捕捉跨帧的人体动作变化与结构一致性。此类方法广泛借助Transformer、时序卷积网络等模型框架，在提升姿态估计精度的同时，增强了系统对复杂场景中长时依赖的建模能力。

LPFormer(Ye等，2024)代表了时空序列建模在该领域的先进应用。该方法通过两阶段处理流程，在第一阶段完成人体边界框检测及多尺度点云特征提取，在第二阶段引入Transformer结构对关键点进行时序预测。模型通过多任务联合训练同时优化骨架与身体形状预测，在保持空间结构准确的同时，有效提升了姿态的连续性与稳定性。另一代表性方法LiveHPS(Ren等，2024)则侧重于构建稳健的时空几何建模机制。该方法引入激增机制处理点云帧间分布变化，结合连续帧中的几何约束与动态特征，有效缓解了遮挡与噪声干扰所带来的姿态估计误差。在后续版本LiveHPS++(Ren等，2025)中，进一步引入了三阶段网络结构，显著降低了对干净点云分割结果的依赖，使其在动态复杂环境中依然具备较高精度与连贯性。该方法还融合知识蒸馏策略，兼顾建

模精度与实时性，特别适合低延迟的边缘计算与交互应用。该类方法普遍通过构建帧间一致性约束、显式建模点云连续性与骨架稳定性，突破了早期方法在动态姿态跟踪中的性能瓶颈，为实现高质量、低延迟的人体重建提供了可靠技术基础。

此外，LIP（lidar-inertial perceiving）（Ren 等，2022）通过激光雷达与惯性测量单元（Inertial Measurement Unit，IMU）的融合，实现了在无GPS环境下的连续运动捕捉，为室外巡检、体育训练等应用提供了技术支持。该类方法普遍通过构建帧间一致性约束、显式建模点云连续性与骨架稳定性，突破了早期方法在动态姿态跟踪中的性能瓶颈，为实现高质量、低延迟的人体重建提供了可靠技术基础。

2.3.3　基于多模态融合的方法

在实际应用中，LiDAR点云往往受到遮挡、稀疏性以及多目标干扰等因素的影响，单模态点云难以满足精确、稳定的动作捕捉需求。因此，近年来的研究越来越多地引入多模态信息与语义增强机制，以提升系统整体感知能力与鲁棒性。

SLOPER4D（scene-level outdoor pose estimation and reconstruction in 4d）（Dai 等，2023）提出了一种结合预扫描三维场景模型与人体运动建模的全局4D姿态估计方法。该框架支持在城市道路场景中引入人体与环境之间的几何接触约束，通过优化人体与场景交互关系，提升在复杂地形下的姿态估计准确率。该研究构建的数据集结合城市街景结构与人体动态，填补了该领域在城市级别大尺度场景建模方面的空白。进一步地，CIMI4D（contact-aware imu-lidar motion capture in 4d）（Yan 等，2023）针对攀爬、搬运等人-物高交互动作，提出结合物理接触约束的姿态估计优化机制。该方法考虑人体与物体间的接触区域，通过物理一致性优化算法，使得预测的人体姿态在动作过程中更加符合现实世界的力学约束，在极端运动状态下展现出优越的稳定性。

HSC4D（human-centered 4d scene capture）（Dai 等，2022）与 HiSC4D（hierarchical human-centered 4d scene capture）（Dai 等，2024）分别实现了 IMU 与 LiDAR的协同建模。IMU提供的局部加速度与角速度信息，有效弥补了LiDAR点云中由于遮挡造成的数据缺失。通过多模态数据协同训练，系统能够在遮挡严重或远距场景下依然稳定捕捉人体姿态，同时显著降低了惯性传感器累积误差，实现对大型室内外场景的4D人体中心化感知。RELI11D（rgb-enhanced lidar inertial 11d）（Yan 等，2024）进一步拓展了跨模态一致性建模。该方法通过激光雷达与相机的外参标定，将三维姿态投影至二维图像中，并引入图像中的人体轮廓与二维关键点作为附加监督信号，以优化三维姿态预测质量。FusionPose（Cong 等，2023）面向大规模场景的标注难题，提出了一个弱监督融合框架。该方法通过基于交叉注意力的图像-点云特征融合模块实现自适应模态对齐，并利用时序一致性约束与2D投影监督，在无需3D标注的情况下实现了鲁棒的多人姿态估计。ImmFusion（Chen 等，2023）则首次将毫米波雷达引入融合框架，通过基于Transformer的动态特征融合与模态掩码训练，有效解决了雨、雾、暗光条件下的重建退化问题，为全天候鲁棒动作捕捉提供了新的技术路径。为促进激光雷达人体动捕研究，国内团队构建了多个具有影响力的数据集。LiDARHuman26M（Li 等，2022）是首个带有精确三维姿态标注的远距离激光雷达数据集，覆盖30米范围，包含13名受试者的20项日常活动。LiDARHuman51M（Zhang 等，2024）进一步扩展至人物交互场景，包含10名受试者的20种交互动作，提供了丰富的接触与交互标注。

2.3.4　总结

综合本节内容及表4性能对比，可得出以下结论：

1）现有技术逐渐从单帧几何回归演进至序列时空建模，显著提升了姿态的时序连续性与估计精度。

2）基于多模态融合的方法有效克服了单一点云在遮挡及恶劣条件下的局限性，增强了系统鲁棒性。

3）现有方法在严重遮挡、密集人-物交互以及远距离探测场景下，其精度与稳定性仍有待提升。

2.4　激光雷达语言推理

2.4.1　基于语义对齐的方法

Kolmet 等人（2022）提出了 Text2Pos 框架，首次系统性地定义了“语言驱动的激光雷达定位”任务，如图5所示。Text2Pos的目标是在三维点云地图中，根据语言指令定位出精确的空间位置。该框架采用粗到细的多模态匹配策略，首先通过语义信息筛选候选区域，随后在几何层面执行高精度匹配，从而实现从语言到三维空间的跨模态映射。具体而言，Text2Pos构建了一种两阶段神经结构：在粗匹配阶段，模型利用语言嵌入与点云全局特征之间的相似

表4　CIMI4D数据集上3D人体重建结果

Table 4　Comparisons of 3D Human Reconstruction on the CIMI4D dataset

类别	方法	传感器模态	ACCEL	MPJPE	PMPJPE	PVE	PCK0.3
几何回归	LiDARCapV2(Zhang等,2024)	LiDAR	70.72	389.97	267.76	364.26	0.50
时空建模	LiveHPS(Ren等,2024)	LiDAR	**72.57**	**190.86**	**148.00**	**225.65**	**0.59**
多模态	FusionPose(Cong等,2023)	LiDAR+RGB	68.82	322.21	232.44	435.72	0.45
	ImmFusion(Chen等,2023)	LiDAR+RGB	58.54	242.20	189.92	330.09	0.53

注:加粗字体表示各类方法中各列最优结果。

性计算,初步锁定可能包含目标的区域;在细匹配阶段,则采用几何对齐、局部特征比对等方式完成精确定位。该方法首次明确了语言与三维坐标之间的推理路径,验证了语言描述作为空间约束信号的可行性与有效性。

为了支撑该任务的训练与评估,Kolmet等人(2022)同时构建了KITTI360Pose数据集。该数据集基于KITTI360(Liao等,2022)城市级激光雷达数据扩展而来,涵盖德国卡尔斯鲁厄市九个城区、总行驶距离达80公里。研究人员利用语义分割标签(建筑、车辆、行人、交通灯等)结合自动化模板生成技术,构造了约4.3万条位置-语言对样本。描述内容包括方向(如"右侧"、"前方")、目标物类型(如"灰色建筑"、"蓝色轿车")与空间关系(如"靠近"、"后方"等),文本格式结构清晰、语义可控。相较人工标注,该自动生成机制有效提升了数据规模与一致性,同时保留了语义多样性与语言表达复杂度。Text2Pos不仅为任务定义与数据标准化提供了范式,也奠定了语言与三维空间对齐任务的研究基石。其提出的语义检索—几何精对结构,也成为后续研究的通用框架。

在此基础上,Text2Loc(Xia等,2024)进一步推动了语言与点云语义嵌入空间的统一建模。该方法引入预训练语言模型T5(Ni等,2021)作为语言编码器,构建双塔结构分别对语言与点云进行编码,然后将两者映射至共享语义空间中。与Text2Pos显式比对不同,Text2Loc摒弃了几何对齐过程,通过对比损失强化配对样本之间的语义一致性,并惩罚非配对项的距离接近问题,从而实现更高效的跨模态学习。Text2Loc展现了强大的语义建模能力与泛化性能,特别在语言歧义、描述不完整或目标遮挡的情况下,依然能够借助语义嵌入信息完成准确定位。实验证明,该方法在KITTI360Pose数据集上取得了最优性能,标志着语言驱动的点云定位方法已从几何主导逐步过渡到以语义为核心的表征方式。

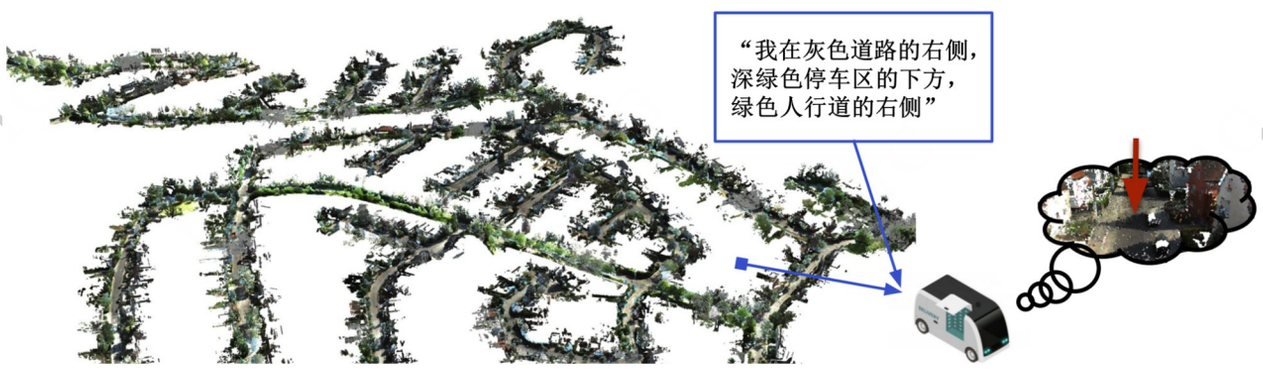

图5　激光雷达语言推理方法(Kolmet等,2022)

Fig. 5　LiDAR-based language reasoning methods(Kolmet et al., 2022)

近年来,国内学者在激光雷达语言推理领域也取得了快速进展,逐步形成了以多层次特征对齐、跨模态注意力机制、语言增强定位及大模型融合为核心的研究体系。与国际研究相比,国内工作更强调模型结构优化与跨模态特征细粒度对齐,注重算法的可扩展性与在自动驾驶场景中的实际应用。

MNCL(multi-level negative contrastive learning)(Liu等,2025)提出了一种多层次负样本对比学习框架,通过语言信息过滤与边界感知特征增强,实现了高精度的语言引导定位。该方法意识到传统的粗到细匹配策略在候选区域筛选阶段存在噪声累积问题,因而引入语言作为过滤器,在全局检索阶段即强化与描述相关的空间区域。MNCL的多层次对比损失函数能够同时约束全局语义一致性与局部几何差异,从而显著提升模型的判别能力。

CMMLoc(cross-modal matching localization)(Xu等,2025)对现有的框架进行了深入分析,指出以往方法往往忽略了语言与点云特征间的部分相关性,导致语义对齐偏移与空间一致性不足。为此,CMMLoc提出了基于跨模态匹配先验的Transformer结构,引入空间整合策略以增强子地图表征。然后,该模型通过预对齐策略和方位方向融合模块进一步强化语言与空间特征的对应关系。该方法在多尺度场景下表现出优越的定位精度,充分体现了跨模态语义一致性建模的重要性。

Des4Pos(Shang等,2025)从特征层面对语言与点云的融合机制进行了深入研究,提出了多尺度特征注意模块与级联残差跨模态融合模块。其中,多尺度特征注意模块结合多尺度自注意与跨注意机制,强化了局部几何特征与全局语义信息的交互;级联残差跨模态融合模块通过级联残差连接实现跨模态特征的动态融合,保持了语义差异性同时提升了融合的稳定性。此外,Des4Pos设计了渐进式文本编码器,借助CLIP模型的语言-视觉先验,将文本从图文联合空间投影至点云-文本联合空间,实现了更稳健的跨模态对齐。

2.4.2 基于结构推理的方法

尽管多模态编码与语义空间对齐技术显著提升了定位性能,但在面对复杂语言结构(如多重关系、嵌套方位、交叉目标)时,传统双塔模型往往难以对语言中的空间逻辑与关系进行深入建模。尤其在存在多个相似对象或需要基于上下文关系推断时,模型需要具备更强的结构理解与语义解析能力。

为解决这一问题,研究开始引入结构化语义建模机制,通过构建场景语义图谱与语言关系图谱,实现语言与点云之间更高级别的语义结构对齐与空间推理。SceneGraphLoc(Miao等,2024)是该方向的代表性研究,首次将场景语义图引入到激光雷达语言推理任务中。该方法将点云数据中识别出的实体(如建筑、车辆、树木等)构建为图中的节点,通过边连接描述对象间的空间关系(如“前方”、“旁边”、“靠近”等),形成三维语义图结构。模型在训练过程中从语言描述中抽取目标对象与关系成分,构造语言图谱,再通过跨图对齐机制在语言图与场景图之间寻找匹配路径,实现结构级的跨模态推理。

国内研究在结构推理方法方面也做出了重要创新。MambaPlace(Shang等,2024)进一步引入结构化状态空间模型(Gu等,2024),将选择性状态机制应用于点云与语言融合任务。其核心创新包括:多策略扫描Mamba模拟视网膜式聚焦机制以强化点云空间关系建模;文本注意Mamba增强语言中方位与目标关键词的语义关联;以及级联跨模态注意Mamba实现多尺度跨模态特征融合与定位偏移预测。MambaPlace展示了基于动态状态建模的跨模态推理能力,显著提升了复杂环境下的空间定位鲁棒性。该研究标志着国内学者开始将新一代序列建模架构引入语言定位任务,为三维多模态融合开辟了新的技术路线。

LangLoc提出了全新的“语言驱动定位”任务定义,旨在通过自然语言直接推断用户的空间位置与朝向(Liao等,2022)。不同于传统依赖配对语料的学习方式,LangLoc利用大语言模型自动生成空间描述,通过提取场景中的关键空间属性(如方位、距离、相对关系)形成结构化文本,再利用语言-点云对齐网络完成定位。LangLoc框架支持纯语言输入与视觉-语言联合输入两种模式,兼具可解释性与灵活性。实验结果表明,该方法在无配对数据的条件下仍能实现较高精度的语言定位,展示了大模型在多模态空间推理中的潜力。

2.4.3 总结

综合本节内容及表5性能对比,可得出以下结论:

1)现有方法已从基于几何匹配的定位,演进为基于共享语义空间的深度对齐,显著提升了跨模态理解的精度与泛化能力。

2)引入结构化建模与新型序列网络,成为处理复杂空间语义关系、实现高级推理的重要发展趋势。

3)面对隐含语义、常识依赖等复杂语言描述,现有方法的推理能力与鲁棒性仍有待加强。

表5　KITTI360Pose数据集上不同方法的性能对比

Table 5　Comparisons of existing methods on the KITTI360Pose dataset

方法	定位召回率(ϵ<5/10/15m)					
	验证集			测试集		
	k=1	k=5	k=10	k=1	k=5	k=10
NetVLAD(Arandjelović等,2016)	0.18/0.33/0.43	0.29/0.50/0.61	0.34/0.59/0.69	0.12/0.15/0.17	0.22/0.32/0.34	0.24/0.29/0.31
PointNetVLAD(Uy等,2018)	0.21/0.28/0.30	0.44/0.58/0.61	0.54/0.71/0.74	0.13/0.17/0.18	0.28/0.37/0.39	0.32/0.39/0.44
Test2Pos(Kolmet等,2022)	0.14/0.25/0.31	0.36/0.55/0.61	0.48/0.68/0.74	0.13/0.20/0.30	0.33/0.42/0.49	0.43/0.61/0.65
RET(Wang等,2023)	0.19/0.30/0.37	0.44/0.62/0.67	0.52/0.72/0.78	0.16/0.25/0.29	0.35/0.51/0.56	0.46/0.65/0.71
Text2Loc(Xia等,2024)	0.37/0.57/0.63	0.68/0.85/0.87	0.77/0.91/0.93	0.33/0.48/0.52	0.60/0.75/0.78	0.70/0.84/0.86
Des4Pos(Shang等,2025)	**0.45/0.63/0.69**	**0.76/0.89/0.92**	**0.84/0.94/0.96**	**0.40/0.54/0.57**	**0.68/0.80/0.82**	**0.77/0.87/0.89**

注:加粗字体表示各类方法中各列最优结果。

3　国内外研究进展比较

通过对国内外研究现状的系统梳理,可以清晰地看到,激光雷达智能处理技术的发展既呈现出全球共通的演进脉络,也因科研环境、应用驱动和资源禀赋的差异,展现出各自鲜明的特色与优势。国内外研究进展的对比如表6所示。对于三维目标检测任务,国际研究较早地确立了基于点、体素和视图的三大基础技术路线,并构建了KITTI、Waymo等具有全球影响力的公开数据集,为领域发展奠定了坚实的算法框架与评估基准。国内研究则在跟进这些主流范式的同时,展现出强烈的应用导向,特别是在解决点云处理效率、恶劣天气鲁棒性以及面向车载边缘计算平台的模型轻量化方面贡献了大量创新工作。

在激光雷达定位领域,国际上的研究起步更早,尤其在基于滤波和优化的传统SLAM方法以及高精度地图依赖的定位方案上积累了深厚的理论基础。近年来,基于学习的无地图绝对定位与里程计成为全球热点,国外团队在自监督学习、跨域泛化以及面向开放世界的基础模型探索上提供了许多奠基性的思路(如DSAC系列工作)。相比之下,国内研究虽起步稍晚,但发展迅猛,尤其在基于回归的无地图定位方向上呈现出强大的创新活力。国内团队在绝对位姿回归中引入时空建模、神经科学启发机制,在激光雷达里程计中广泛应用Transformer、扩散模型等先进架构,在定位精度与鲁棒性上取得了国际领先的成果,显示出在特定技术路径上的深度挖掘与赶超之势。

在激光雷达人体动作捕捉这一新兴方向,国际研究呈现出"理论先行、多模态深度融合"的特点,早期便系统探索了基于物理的建模、人-物-场景交互约束以及隐式神经表征等前沿方向,并构建了多个涵盖复杂交互行为的大规模数据集。国内研究虽起步相对较晚,但紧密围绕安防监控、体育康复、虚拟现实等实际应用需求,在远距离稀疏点云下的姿态估计、轻量化网络设计以及自主数据体系建设方面取得了显著进展,形成了从数据到模型再到应用的完整研究体系。

在最为前沿的激光雷达语言推理领域,国内外研究几乎处于同一起跑线。国际研究率先定义了任务范式并构建了首个基准数据集,在跨模态对齐与推理机制上进行了初步探索。国内研究则迅速跟进,并更侧重于模型结构的精细优化与跨模态特征的细粒度对齐,提出了多种基于注意力机制、对比学习乃至状态空间模型的融合网络,在定位精度和语义理解深度上展现出竞争力,部分工作已开始探索与大语言模型的结合,显示出前瞻性布局。

综上所述,国际研究在激光雷达智能处理领域长期以来发挥着引领作用,尤其在提出基础理论、构建权威数据集和开拓新研究方向方面贡献卓著。而国内研究则凭借其对国家重大应用需求的深刻理解、在特定技术点上的集中攻关以及高效的工程化实现能力,正逐渐从"跟跑"向"并跑"乃至部分领域的"领跑"转变,形成了与国际研究互补互鉴、共同推

动学科发展的良好格局。未来,随着技术的不断深化和应用场景的持续拓展,这种全球协作与良性竞争并存的态势将进一步加速激光雷达智能处理技术的创新与落地。

表6　国内外研究进展比较

Table 6　Comparison of domestic and international research progress

任务	国际研究进展	国内研究进展	国内外对比
三维目标检测	确定了基于点、体素、视图的三大基础技术路线,构建了权威数据集,引领了多模态深度融合方向。	聚焦效率优化与模型轻量化,提出多种知识蒸馏、协同感知策略,强调适配车载平台和恶劣环境。	国际研究全面;国内注重工程导向。
激光雷达定位	在基于滤波/优化的SLAM和高精地图依赖定位上积累深厚,重点关注无监督学习与跨域泛化。	在基于回归的无地图定位中发展迅猛,专注于网络架构设计以提升定位精度、鲁棒性和实时性。	国际理论扎实;国内在特定技术上实现了重点突破。
激光雷达人体动捕	强调物理建模与人-物-场景交互,引入隐式表征并建立大规模数据集支撑复杂行为理解。	从LiDARCap起步,发展LiveHPS、LPFormer等系列方法,应用导向强。	国际注重前沿探索;国内重视实际适配与部署。
激光雷达语言推理	率先定义任务范式(Text2Pos),并构建首个基准数据集(KITTI360Pose)和检索-对齐框架。	提出了一系列精细化融合网络,并开始探索利用大语言模型生成描述或增强推理。	国际率先构建范式;国内注重性能平衡。

4　发展趋势和展望

随着自动驾驶、机器人、增强现实等领域的快速发展,激光雷达智能处理技术正迎来前所未有的发展机遇。基于前文对国内外技术现状的系统梳理,本文从算法融合、任务扩展、系统优化三个层面,展望激光雷达智能处理技术的未来发展方向。

在算法融合层面,未来研究将致力于构建统一的多模态表征框架,实现跨模态语义的端到端对齐。通过图-语言-点云联合嵌入空间的设计,促进语义与几何的深度耦合,而非依赖手工设计的对齐策略。同时,借助大语言模型与视觉-语言模型的先验知识,增强系统对复杂场景的语义推理能力。

在任务扩展层面,随着技术不断成熟,激光雷达智能处理的任务边界正逐步拓宽。从早期的目标检测、定位,延伸至人体动捕、语言定位、场景理解等高层次任务。未来有望在智慧城市、人机协作、远程巡检、虚拟现实等更多领域实现深度融合与创新应用,构建以激光雷达为核心的全息感知与交互系统。这一扩展不仅体现在任务类型上,也体现在应用场景的广度与深度上,将为智能系统提供更丰富的交互与理解能力。

在系统优化层面,需在保持精度的前提下,推进模型压缩、神经架构搜索、动态推理等轻量化技术的研究,构建兼顾效率与性能的实时激光雷达处理系统。同时,针对现有算法大多针对特定场景或传感器进行优化,面临跨域、跨设备泛化能力不足的挑战,未来应重点探索领域自适应、元学习、自监督预训练等机制,构建对天气变化、传感器差异、场景动态性具有强适应性的通用模型。

5　结　论

本文系统回顾了激光雷达智能处理技术的研究进展。首先,梳理了激光雷达智能处理的基本任务范畴与核心挑战。其次,围绕三维目标检测、激光雷达定位、人体动作捕捉与语言推理等关键方向,深入分析了国内外研究现状。进而,通过对主流数据集上性能表现的比较与分析,揭示了不同方法的特性。最后,对未来研究方向进行了展望。

致谢:本文由中国图象图形学学会成像探测与感知专业委员会组织撰写,该专委会链接为https://www.csig.org.cn/16/201704/49321.html。

参考文献(References)

Ao S, Hu Q, Yang, B, Markham, A and Guo Y. 2021. Spinnet: Learning a general surface descriptor for 3d point cloud registration//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Nashville: IEEE: 11753-11762 [DOI: 10.1109/

CVPR46437.2021.01158].

Ao S, Guo Y, Hu Q, Yang B, Markham A and Chen Z. 2022. You only train once: Learning general and distinctive 3D local descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(3), 3949-3967 [DOI: 10.1109/TPAMI.2022.3180341].

Ao S, Hu Q, Wang H, Xu K and Guo Y. 2023. Buffer: Balancing accuracy, efficiency, and generalizability in point cloud registration// Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Vancouver: IEEE: 1255-1264 [10.1109/CVPR52729.2023.00127].

Abrar H. Abdulnabi, Bing Shuai, Zhen Zuo, Lap-Pui Chau, and Gang Wang. 2017. Multimodal recurrent neural networks with information transfer layers for indoor scene labeling//IEEE Transactions on Multimedia, 20 (7): 1656-1671 [DOI: 10.1109/TMM. 2017. 2774007]

Afham M, Dissanayake I, Dissanayake D, Dharmasiri A, Thilakarathna K, and Rodrigo R. 2022. Crosspoint: Self-supervised cross-modal contrastive learning for 3d point cloud understanding// Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. New Orleans, USA: IEEE: 9892-9902 [DOI: 10.1109/CVPR52688.2022.00967]

Bai X, Hu Z, Zhu X, Huang Q, Chen Y, Fu H, and Tai C L. 2022. TransFusion: Robust LiDAR-Camera Fusion for 3D Object Detection with Transformers//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. New Orleans, USA: IEEE: 1080-1089 [DOI: 10.1109/CVPR52688.2022.00116]

Barnes D, Gadd M, Murcutt P, Newman P, and Posner I. 2020. The Oxford Radar RobotCar Dataset: A Radar Extension to the Oxford RobotCar Dataset//Proceedings of the IEEE International Conference on Robotics and Automation (ICRA). Paris, France : IEEE: 6433-6438 [DOI: 10.1109/ICRA40945.2020. 9196884]

Brachmann E and Rother C. 2021. Visual camera re-localization from RGB and RGB-D images using DSAC. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44 (9): 5847-5865 [DOI: 10.1109/TPAMI.2021.3070754]

Brachmann E, Cavallari T, and Prisacariu V A. 2023. Accelerated coordinate encoding: Learning to relocalize in minutes using RGB and poses//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, Canada: IEEE:5044-5053 [DOI: 10.1109/CVPR52729.2023.00488]

Bresson G, Alsayed Z, Yu L, and Glaser S. 2017. Simultaneous localization and mapping: A survey of current trends in autonomous driving. IEEE Transactions on Intelligent Vehicles, 2(3): 194-220 [DOI: 10.1109/TIV.2017.2749181]

Caesar H, Bankiti V, Lang A H, Vora S, Liong V E, Xu Q, Krishnan A, Pan Y, Baldan G, and Beijbom O. 2020. nuScenes: A multimodal dataset for autonomous driving//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 11621-11631 [DOI: 10.1109/CVPR42600.2020.01164]

Cai Q, Pan Y, Yao T, Ngo C W, and Mei T. 2023. Objectfusion: Multimodal 3d object detection with object-centric fusion//Proceedings of the IEEE/CVF international conference on computer vision. Paris, France: IEEE: 18021-18030 [DOI: 10.1109/ICCV51070.2023. 01656]

Cao X, Wang P, Zhang Z, Tu H, Chen Y, and Liang Z. 2025. RAF-Det: A Novel Camera-Radar Fusion Framework for Robust 3D Object Detection in Autonomous Driving//ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Hyderabad, India: IEEE: 1-5 [DOI: 10.1109/ICASSP49660.2025.10888004]

Cheng J M, Xie W J, Shen Z Q, Li L and Liu X P. 2022. Multimodal Human Motion Synchronization Dataset. Journal of Computer-Aided Design & Computer Graphics, 34(11):1713 - 1722 (程景铭,谢文军,沈子祺,李琳,刘晓平. 2022. 多模态人体运动同步数据集. 计算机辅助设计与图形学学报,34(11):1713 - 1722).

Chen S, Wang R, Li X, Wu Y, Liu H, Chen J, and Ma H. 2024. PLS: Unsupervised Domain Adaptation for 3d Object Detection Via Pseudo-Label Sizes//ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Seoul, Korea: IEEE: 6370-6374 [DOI: 10.1109/ICASSP48485. 2024.10446579]

Chen X, Ma H, Wan J Li, B, and Xia T. 2017. Multi-view 3d object detection network for autonomous driving//Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 1063-6919 [DOI: 10.1109/CVPR.2017.691]

Chen X, Zhang T, Wang Y, and Zhao H. 2023. Futr3d: A unified sensor fusion framework for 3d detection//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Vancouver, Canada: IEEE: 172-181 [DOI: 10.1109/CVPRW59228. 2023.00022]

Cho Y, Kim G, and Kim A. 2020. Unsupervised geometry-aware deep LiDAR odometry//2020 IEEE International Conference on Robotics and Automation (ICRA). Paris, France: IEEE: 2145-2152 [DOI: 10.1109/ICRA40945.2020.9197366]

Choy C, Park J, and Koltun V. 2019. Fully convolutional geometric features//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korean: IEEE: 8958-8966 [DOI: 10.1109/ICCV.2019.00905]

Dai Y, Lin Y, Lin X, Wen C, Xu L, Yi H, Shen S, Ma Y, and Wang C. 2023. SLOPER4D: A scene-aware dataset for global 4D human pose estimation in urban environments//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE: 682-692 [DOI: 10.1109/CVPR52729.2023. 00073]

Dai Y, Wang Z, Lin X, Wen C, Xu L, Shen S, Ma Y, and Wang C. 2022. HSC4D: Human-centered 4D scene capture in large-scale

indoor-outdoor space using wearable IMUs and LiDAR//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE: 6782-6792 [DOI: 10.1109/CVPR52688.2022.00667]

Deng J, Shi S, Li P, Zhou W, Zhang Y, and Li H. 2021. Voxel r-cnn: Towards high performance voxel-based 3d object detection//Proceedings of the AAAI conference on artificial intelligence. 35(2): 1201-1209 [DOI: 10.1609/aaai.v35i2.16207]

Diaz-Ruiz C A, Xia Y, You Y, Nino J, Chen J, Monica J, Chen X, Luo K, Wang Y, and Emond M. 2022. Ithaca365: Dataset and driving perception under repeated and challenging weather conditions//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, USA: IEEE: 21383-21392 [DOI: 10.1109/CVPR52688.2022.02069]

Feng M T, Shen J H, Wu Z J, Peng W X, Zhong H, Guo Y L, Shu X B, Zhang H, Dong W S and Wang Y N. 2025. Advancements in 3D vision understanding using multimodal large language models. Journal of Image and Graphics, 30(6):1744-1791 (冯明涛,沈军豪,武子杰,彭伟星,钟杭,郭裕兰,舒祥波,张辉,董伟生,王耀南. 2025. 多模态大模型驱动的三维视觉理解技术前沿进展. 中国图象图形学报,30(6):1744-1791) [DOI:10. 11834/jig. 240588]

Fürst M, Gupta S T, Schuster R, Wasenmüller O, and Stricker D. 2021. HPERL: 3D human pose estimation from RGB and LiDAR//25th International Conference on Pattern Recognition (ICPR). IEEE: 7321-7327 [DOI: 10.1109/ICPR48806.2021.9412785]

Gadelha M, RoyChowdhury A, Sharma G, Kalogerakis E, Cao L, Learned-Miller E, Wang R, and Maji S. 2020. Label-efficient learning on point clouds using approximate convex decompositions//European Conference on Computer Vision. Cham: Springer International Publishing: 473-491 [DOI: 10.1007/978-3-030-58607-2_28]

Geiger A, Lenz P, and Urtasun R. 2012. Are we ready for autonomous driving? The KITTI Vision Benchmark Suite//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Providence, USA: IEEE: 3354-3361 [DOI: 10.1109/CVPR.2012.6248074]

Gojcic Z, Litany O, Wieser A, Guibas L J, and Birdal T. 2021. Weakly supervised learning of rigid 3D scene flow// Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5692-5703. Nashville, USA: IEEE: 5692-5703 [DOI: 10.1109/CVPR46437.2021.00564]

Goswami R G, Patel N, Krishnamurthy P, and Khorrami F. 2025. FlashMix: Fast map-free LiDAR localization via feature mixing and contrastive-constrained accelerated training//2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Tucson, USA: IEEE: 2011-2020 [DOI: 10.1109/WACV61041.2025.00202]

Gu A and Dao T. 2024. Mamba: Linear-time sequence modeling with selective state spaces//First conference on language modeling. [DOI: 10.48550/arXiv.2312.00752]

Guo Z Y, L R, Z X, W Z 2025. LiDAR SLAM system based on ground segmentation and loop closure optimization in dynamic environment, 45(S1):302-308 (郭致远,刘瑞,赵轩,王姝. 2025. 动态场景下基于地面分割与回环优化的激光雷达定位与建图系统. 计算机应用,45(S1):302-308)

Henzler P, Reizenstein J, Labatut P, Shapovalov R, Ritschel T, Vedaldi A, and Novotny D. 2021. Unsupervised learning of 3d object categories from videos in the wild//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 4698-4707 [DOI: 10.1109/CVPR46437.2021.00467]

Huang S, Xie Y, Zhu S C, and Zhu Y. 2021. Spatio-temporal self-supervised representation learning for 3d point clouds//Proceedings of the IEEE/CVF international conference on computer vision, Montreal, Canada: IEEE: 6515-6525 [DOI: 10.1109/ICCV48922.2021.00647.]

Huang X, Wang J, Xia Q, Chen S, Yang B, Li X, Wang C, and Wen C. 2025. V2X-R: Cooperative LiDAR-4D Radar Fusion with Denoising Diffusion for 3D Object Detection//2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, USA: IEEE: 27390-27400 [DOI: 10.1109/CVPR52734.2025.02551]

Huang X, Wu H, Li X, Fan X, Wen C, and Wang C. 2024. Sunshine to rainstorm: Cross-weather knowledge distillation for robust 3d object detection//Proceedings of the AAAI Conference on Artificial Intelligence, 38 (3): 2409-2416 [DOI: 10.1609/aaai. v38i3.28016]

Huang Y, Liu C, Zhu M, Ao S, Wen C, and Wang C. 2025. DiffLO: Semantic-Aware LiDAR Odometry with Diffusion-Based Refinement//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE: 17050-17059 [DOI: 10.1109/CVPR52734.2025.01589]

Kendall A, Grimes M, and Cipolla R. 2015. PoseNet: A convolutional network for real-time 6-DOF camera relocalization//Proceedings of the IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE: 2938-2946 [DOI: 10.1109/ICCV.2015.336]

Kolmet M, Zhou Q, Ošep A, and Leal-Taixé L. 2022. Text2pos: Text-to-point-cloud cross-modal localization//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition: 6687-6696 [DOI: 10.48550/arXiv.2203.15125]

Li C C, Chen G, Hou Z X, Huang K and Zhang W. 2024. Survey of 3D object detection algorithms for autonomous driving. Journal of Image and Graphics, 29(11):3238-3264 (李昌财,陈刚,侯作勋,黄凯,张伟. 2024. 自动驾驶中的三维目标检测算法研究综述. 中国图象图形学报, 29(11):3238-3264) [DOI:10. 11834/jig. 230779]

Li J, Zhang J, Wang Z, Shen S, Wen C, Ma Y, Xu L, Yu J, and Wang C. 2022. LiDARCap: Long-range marker-less 3D human motion capture with LiDAR point clouds//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE: 20470-20480 [DOI: 10.1109/CVPR52688.2022.01985]

Li Q, Chen S, Wang C, Li X, Wen C, Cheng M, and Li J. 2019. LO-Net: Deep real-time LiDAR odometry//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE: 8473-8482 [DOI: 10.1109/CVPR.2019.00867]

Li W, Liu C, Yu S, Liu D, Zhou Y, Shen S, Wen C, and Wang C. 2025. LightLoc: Learning outdoor LiDAR localization at light speed//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE: 6680-6689 [DOI: 10.1109/CVPR52734.2025.00626]

Li W, Yang Y, Yu S, Hu G, Wen C, Cheng M, and Wang C. 2024. DiffLoc: Diffusion Model for Outdoor LiDAR Localization//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 15045-15054 [DOI: 10.1109/CVPR52733.2024.01425]

Li W, Yu S, Wang C, Hu G, Shen S, and Wen C. 2023. SGLoc: Scene geometry encoding for outdoor LiDAR localization//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, Canada: IEEE: 9286-9295 [DOI: 10.1109/CVPR52729.2023.00896]

Li Z, Liang H, Wang, H, Zhao M, Wang J, and Zheng X. 2023. MKD-cooper: Cooperative 3D object detection for autonomous driving via multi-teacher knowledge distillation. IEEE Transactions on Intelligent Vehicles, 9 (1): 1490-1500 [DOI: 10.1109/TIV.2023.3310580]

Liao Y, Xie J, and Geiger A. 2022. Kitti-360: A novel dataset and benchmarks for urban scene understanding in 2d and 3d. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45 (3): 3292-3310 [DOI: 10.1109/TPAMI.2022.3179507]

Liu D, Huang S, Li W, Shen S, and Wang C. 2025. Text to point cloud localization with multi-level negative contrastive learning//Proceedings of the AAAI Conference on Artificial Intelligence, 39 (5): 5397-5405 [DOI: 10.1609/aaai.v39i5.32574]

Liu J, Wang G, Jiang C, Liu Z, and Wang H. 2023. TransLO: A window-based masked point transformer framework for large-scale LiDAR odometry//Proceedings of the AAAI Conference on Artificial Intelligence. 37(2): 1683-1691 [DOI: 10.1609/aaai.v37i2.25256]

Liu Y C, Ma C Y, He Z, Kuo C W, Chen K, Zhang P, Wu B, Kira Z and Vajda P 2021. Unbiased teacher for semi-supervised object detection//Proceedings of the International Conference on Learning Representations (ICLR). [DOI: 10.48550/arXiv.2102.09480]

Liu Z, Tang H, Amini A, Yang X, Mao H, Rus D, and Han S. 2022. BEVFusion: Multi-Task Multi-Sensor Fusion with Unified Bird's-Eye View Representation//Proceedings of the IEEE/CVF international conference on robotics and automation. London, United Kingdom: IEEE: 2774-2781 [DOI: 10.1109/ICRA48891.2023.10160968]

Lu D and Schnieder E. 2014. Performance evaluation of GNSS for train localization. IEEE Transactions on Intelligent Transportation Systems, 16(2): 1054-1059 [DOI: 10.1109/TITS.2014.2349353]

Lu S, Zhuo G, Zheng L, Zhu J, and Bai J. 2025. A Generative Hierarchical Optimization Framework for LiDAR Odometry Using Conditional Diffusion Models. IEEE Sensors Journal (Early Access): 1-1 [DOI: 10.1109/JSEN.2025.3575156]

Luo L, Zheng S, Li Y, Fan Y, Yu B, and Cao S. 2023. BEVPlace: Learning LiDAR-based place recognition using bird's eye view images//Proceedings of the IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 8666-8675 [DOI: 10.1109/ICCV51070.2023.00799]

Maddern W, Pascoe G, Linegar C, and Newman P. 2017. 1 Year, 1000 km: The Oxford RobotCar Dataset. The International Journal of Robotics Research, 36 (1): 3-15 [DOI: 10.1177/0278364916679498]

Meng Q, Wang W, Zhou T, Shen J, Jia Y, and Van Gool L. 2021. Towards a weakly supervised framework for 3D point cloud object detection and annotation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44 (8): 4454-4468. [DOI: 10.1109/TPAMI.2021.3063611]

Meyer G P, Laddha A, Kee E, Vallespi-Gonzalez C, and Wellington C. K. 2019. LaserNet: An Efficient Probabilistic 3D Object Detector for Autonomous Driving//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition: 12677-12686 [DOI: 10.1109/CVPR.2019.01296]

Miao Y, Engelmann F, Vysotska O, Tombari F, Pollefeys M, and Baráth DB. 2024. Scenegraphloc: Cross-modal coarse visual localization on 3d scene graphs//European Conference on Computer Vision. Cham: Springer Nature Switzerland: 127-150 [DOI: 10.1007/978-3-031-73242-3_8]

Ni J, Abrego GH, Constant N, Ma J, Hall KB, Cer D, and Yang Y. 2021. Sentence-t5: Scalable sentence encoders from pre-trained text-to-text models. [DOI: 10.48550/arXiv.2108.08877]

Nubert J, Khattak S, Hutter M. Self-supervised learning of lidar odometry for robotic applications//IEEE international conference on robotics and automation (ICRA). IEEE, 2021: 9601-9607.

Qi C R, Su H, Mo K, and Guibas L J. 2017. PointNet: Deep learning on point sets for 3D classification and segmentation//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE: 652-660 [DOI: 10.1109/CVPR.2017.16]

Qi C R, Yi L, Su H, and Guibas L J. 2017. PointNet++: Deep hierarchical feature learning on point sets in a metric space//Advances

in Neural Information Processing Systems, 30. Long Beach: Curran Associates, Inc.: 5099-5108 [DOI: 10.5555/3295222.3295263]

Radford A, Kim J, Hallacy C, Ramesh A, Goh G, Agarwal S, Sastry G, Askell A, Mishkin P, Clark J, Krueger G, and Sutskever I. 2021. Learning transferable visual models from natural language supervision//Proceedings of the 38th International Conference on Machine Learning, 139: 8748-8763 [DOI: 10.48550/arXiv.2103.00020]

Ren Y, Han X, Yao Y, Long X, Sun Y, and Ma Y. 2024. LiveHPS++: Robust and coherent motion capture in dynamic free environment// European Conference on Computer Vision (ECCV). Springer: 127-144 [DOI: 10.1007/978-3-031-73397-0_8]

Ren Y, Han X, Zhao C, Wang J, Xu L, Yu J, and Ma Y. 2024. LiveHPS: LiDAR-based scene-level human pose and shape estimation in free environment//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE: 1281-1291 [DOI: 10.1109/CVPR52733.2024.00128]

Sanghi A. 2020. Info3d: Representation learning on 3d objects using mutual information maximization and contrastive learning//European conference on computer vision. Glasgow, UK: IEEE: 626-642. [DOI: 10.1007/978-3-030-58526-6_37]

Shang T, Li Z, Xu P, and Qiao J. 2024. Mambaplace: Text-to-point-cloud cross-modal place recognition with attention mamba mechanisms. [DOI: 10.48550/arXiv.2408.15740]

Shang T, Li Z, Xu P, Deng Z, and Zhang R. 2025. Text-driven 3d lidar place recognition for autonomous driving. [DOI: 10.48550/arXiv.2503.18035]

Shi S, Guo C, and Jiang L 2020. PV-RCNN: Point-Voxel Feature Set Abstraction for 3D Object Detection//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 10526-10535 [DOI: 10.1109/CVPR42600.2020.01054]

Shi S, Wang X, and Li H. 2019. PointRCNN: 3D Object Proposal Generation and Detection from Point Cloud//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 770-779 [DOI: 10.1109/CVPR.2019.00086]

Shi W, Chen C, Li K, Xiong Y, Cao X, and Zhou Z. 2025. LangLoc: Language-Driven Localization via Formatted Spatial Description Generation. IEEE Transactions on Image Processing, 34: 1737-1752 [DOI: 10.1109/TIP.2025.3546853]

Tan Z, Niu Z Y, Zhang J P, Chen X Y L and Hu D W. 2025. New opportunities in SLAM-Gaussian splatting technology. Journal of Image and Graphics, 30(6):1792-1807 (谭臻,牛中颜,张津浦,陈谢沅澧,胡德文. 2025. SLAM新机遇—高斯溅射技术. 中国图象图形学报,30(6):1792-1807)[DOI:10.11834/jig.240443]

Tang L F, Zhang H, Xu H and Ma J Y. 2023. Deep learning-based image fusion: a survey. Journal of Image and Graphics, 28(01): 0003-0036 (唐霖峰,张浩,徐涵,马佳义. 2023. 基于深度学习的图像融合方法综述. 中国图象图形学报,28(01):0003-0036) [DOI:10.11834/jig.220422]

Tarvainen A and Valpola H. 2017. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results//Proceedings of the 31st International Conference on Neural Information Processing Systems: 1195-1204 [DOI: 10.48550/arXiv.1703.01780]

Uy M A and Lee G H. 2018. PointNetVLAD: Deep point cloud based retrieval for large-scale place recognition//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, USA: IEEE: 4470-4479 [DOI: 10.1109/CVPR.2018.00470]

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser Ł, and Polosukhin I. 2017. Attention is all you need//Advances in Neural Information Processing Systems, 30. Long Beach, USA: CA, Inc.: 5998-6008 [DOI: 10.5555/3295222.3295349]

Vora S, Lang A H, Helou B, and Beijbom O. 2020. Joint 3D Proposal Generation and Object Detection from View Aggregation//Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Madrid, Spain: IEEE: 978-1-5386-8095-7 [DOI: 10.1109/IROS.2018.8594049]

Vora S, Lang A H, Helou B, and Beijbom O. 2020. Pointpainting: Sequential fusion for 3d object detection//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Seattle, USA: IEEE: 4604-4612 [DOI: 10.1109/CVPR42600.2020.00466]

Wang G, Fan H, and Kankanhalli M. 2023. Text to point cloud localization with relation-enhanced transformer//Proceedings of the AAAI Conference on Artificial Intelligence, 37(2): 2501-2509 [DOI: 10.1609/aaai.v37i2.25347]

Wang G, Wu X, Jiang S, Liu Z, and Wang H. 2022. Efficient 3D Deep LiDAR Odometry. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(5): 5749-5765 [DOI: 10.1109/TPAMI.2022.3207015]

Wang G, Wu X, Liu Z, and Wang H. 2021. PWCLO-Net: Deep LiDAR odometry in 3D point clouds using hierarchical embedding mask optimization//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE: 15910-15919 [DOI: 10.1109/CVPR46437.2021.01565]

Wang H, Cong Y, Litany O, Gao Y, and Guibas L. 2021. 3dioumatch: Leveraging iou prediction for semi-supervised 3d object detection//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 14615-14624 [DOI: 10.1109/CVPR46437.2021.01438]

Wang S, Kang Q, She R, Wang W, Zhao K, Song Y, and Tay W P. 2023. HypLiLoc: Towards effective LiDAR pose

regression with hyperbolic fusion//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, Canada: IEEE: 5176-5185 [DOI: 10.1109/CVPR52729.2023.00501]

Wang S, Laskar Z, Melekhov I, Li X, Zhao Y, Tolias G, and Kannala J. 2024. HSCNet++: Hierarchical scene coordinate classification and regression for visual localization with transformer. International Journal of Computer Vision, 132(7): 2530-2550 [DOI: 10.1007/s11263-023-01982-9]

Wang W, Saputra M R U, Zhao P, Gusmao P, Yang B, Chen C, Markham A, and Trigoni N. 2019. DeepPCO: End-to-end point cloud odometry through deep parallel neural network//2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Macau, China: IEEE: 3248-3254 [DOI: 10.1109/IROS40897.2019.8967756]

Wang W, Wang B, Zhao P, Chen C, Clark R, Yang B, and Markham A. 2022. PointLoc: Deep pose regressor for LiDAR point cloud localization. IEEE Sensors Journal, 22(1): 959-968 [DOI: 10.1109/JSEN.2021.3128683]

Wang Y, Zhang B, Wan Y, Yue Z, and Zhang Y. 2025. Confidence-Aware Superpixel Self-Supervised Learning for Cross-Modal Point Cloud Recognition. IEEE Geoscience and Remote Sensing Letters, 22(000): 1-5 [DOI: 10.1109/LGRS.2025.3587236]

Wu H, Wen C, Shi S, Li X, and Wang C. 2023. Virtual sparse convolution for multimodal 3d object detection//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition: 21653-21662 [DOI: 10.1109/CVPR52729.2023.02074]

Wu H, Zhao S, Huang X, Wen C, Li X, and Wang C. 2024. Commonsense prototype for outdoor unsupervised 3d object detection//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 14968-14977 [DOI: 10.1109/CVPR52733.2024.01418]

Xia Q, Deng J, Wen C, Wu H, Shi S, Li X, and Wang, C. 2023. CoIn: Contrastive instance feature mining for outdoor 3d object detection with very limited annotations//Proceedings of the IEEE/CVF International Conference on Computer Vision: 6231-6240 [DOI: 10.1109/ICCV51070.2023.00575]

Xia Y, Shi L, Ding Z, Henriques JF, and Cremers D. 2024. Text2loc: 3d point cloud localization from natural language//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition: 14958-14967 [DOI: 10.1109/CVPR52733.2024.01417]

Xia Y, Xu Y, Li S, Wang R, Du J, Cremers D, and Stilla U. 2021. SOE-Net: A self-attention and orientation encoding network for point cloud based place recognition//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE: 11348-11357 [DOI: 10.1109/CVPR46437.2021.01119]

Xie C, Lin C, and Zheng X. 2023. Dense sequential fusion: Point cloud enhancement using foreground mask guidance for multimodal 3-d object detection. IEEE Transactions on Instrumentation and Measurement, 73: 1-15 [DOI: 10.1109/TIM.2023.3332935]

Xie Q, Lai Y K, Wu J, Wang Z, Lu D, Wei M, and Wang J. 2021. Venet: Voting enhancement network for 3d object detection//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 3692-3701 [DOI: 10.1109/ICCV48922.2021.00369]

Xie S, Gu J, Guo D, Qi C R, Guibas L, and Litany O. 2020. Pointcontrast: Unsupervised pre-training for 3d point cloud understanding//European conference on computer vision. Cham: Springer International Publishing: 574-591 [DOI: 10.1007/978-3-030-58580-8_34]

Xu S, Zhou D, Fang J, Yin J, Bin Z, and Zhang L. 2021. FusionPainting: Multimodal fusion with adaptive attention for 3D object detection//International Intelligent Transportation Systems Conference (ITSC). Indianapolis, USA: IEEE: 3047-3054 [DOI: 10.1109/ITSC48978.2021.9564951]

Xu Y, Qu H, Liu J, Zhang W, and Yang X. 2025. CMMLoc: Advancing Text-to-PointCloud Localization with Cauchy-Mixture-Model Based Framework//Proceedings of the Computer Vision and Pattern Recognition Conference: 6637-6647 [DOI: 10.48550/arXiv.2503.02593]

Yan M, Wang X, Dai Y, Cai S, Fan S, Lin X, Dai Y, Shen Si, Wen C, Xu L, Ma Y, and Wang C. 2024. RELI11D: A comprehensive multimodal human motion dataset and method//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE: 2250-2262 [DOI:10.1109/CVPR52733.2024.00219]

Yan Y, Mao Y, and Li B. 2018. Second: Sparsely Embedded Convolutional Detection. Sensors, 18 (10): 3337 [DOI: 10.3390/s18103337]

Yang B, Li Z, Li W, Cai Z, Wen C, Zang Y, Müller M, and Wang C. 2024. LiSA: LiDAR localization with semantic awareness//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 15271-15280 [DOI: 10.1109/CVPR52733.2024.01446]

Yang B, Luo W, and Urtasun R. 2018. PIXOR: Real-time 3D Object Detection from Point Clouds//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA: IEEE: 7652-7660 [DOI: 10.1109/CVPR.2018.00798]

Yang Y, Li W, Ao S, Xu Q, Yu S, Guo Y, Zhou Y, Shen S, and Wang C. 2025. RALoc: Enhancing outdoor LiDAR localization via rotation awareness//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Honolulu, USA: IEEE: 3304-3313 [DOI: pending]

Yang Z, Sun Y, Liu S, and Jia J. 2020. 3DSSD: Point-Based 3D Single Stage Object Detector//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 11037-11045 [DOI: 10.1109/CVPR42600.2020.01105]

Ye D, Xie Y, Chen W, Zhou Z, Ge L, and Foroosh H. 2024. LPFormer: LiDAR pose estimation transformer with multi-task network//IEEE International Conference on Robotics and Automation (ICRA). IEEE: 16432-16438 [DOI: 10.1109/ICRA57147.2024.10611405]

Yu S, Sun X, Li W, Wen C, Yang Y, Si B, Hu G, and Wang C. 2024. NIDALoc: Neurobiologically inspired deep LiDAR localization. IEEE Transactions on Intelligent Transportation Systems, 25(5): 4278-4289 [DOI: 10.1109/TITS.2023.3324700]

Yu S, Wang C, Lin Y, Wen C, Cheng M, and Hu G. 2023. STCLoc: Deep LiDAR Localization with Spatio-Temporal Constraints. IEEE Transactions on Intelligent Transportation Systems, 24(1): 489-500 [DOI: 10.1109/TITS.2022.3213311]

Yu S, Wang C, Wen C, Cheng M, Liu M, Zhang Z, and Li X. 2022. LiDAR-based localization using universal encoding and memory-aware regression. Pattern Recognition, 128(1): 108685 [DOI: 10.1016/j.patcog.2022.108685]

Zhang J Y, Mao Q H, Shen S Q, Wen C L, Xu L, and Wang C. 2024. LiDARCapV2: 3D human pose estimation with human - object interaction from LiDAR point clouds. Pattern Recognition, 156: 110848 [DOI: 10.1016/j.patcog.2024.110848]

Zhang X, Fan Z, Shen Y, Li Y, An Y, and Tan X. 2024. MAEMOT: Pretrained MAE-based antiocclusion 3-D multiobject tracking for autonomous driving. IEEE Transactions on Neural Networks and Learning Systems, 36(10): 10721-10735 [DOI: 10.1109/TNNLS.2024.3480148]

Zhao B, Zhang W, and Zou Z. 2023. Bm2cp: Efficient collaborative perception with lidar-camera modalities. Conference on Robot Learning: 1022 - 1035 [DOI: 10.48550/arXiv.2310.14702]

Zheng C, Lyu Y, Li M, and Zhang Z. 2020. LodoNet: A deep neural network with 2D keypoint matching for 3D LiDAR odometry estimation//Proceedings of the 28th ACM International Conference on Multimedia (MM). Seattle, USA: ACM: 2391-2399 [DOI: 10.1145/3394171.3413771]

Zheng J, Shi X, Gorban A, Mao J, Song Y, and Qi C. 2022. Multi-modal 3D human pose estimation with 2D weak supervision in autonomous driving//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE: 4477-4486 [DOI: 10.1109/CVPRW56347.2022.00494]

Zheng W, Tang W, Chen S, Jiang L, and Fu C W. 2021. CIA-SSD: Confident IoU-Aware Single-Stage Object Detector From Point Cloud//Proceedings of the AAAI conference on artificial intelligence, 35(4): 3555-3562 [DOI: 10.1609/aaai.v35i4.16470]

Zhou Y and Tuzel O. 2017. VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection//Proceedings of the IEEE conference on computer vision and pattern recognition: 4490-4499 [DOI: 10.1109/CVPR.2018.00472]

作者简介

敖晟,男,助理教授,主要研究方向为激光雷达数据处理。E-mail: aosh@xmu.edu.cn

温程璐,女,教授,主要研究方向为激光雷达数据处理。E-mail: clwen@xmu.edu.cn

李文,男,博士研究生,主要研究方向为激光雷达数据处理。E-mail: liwen777@stu.xmu.edu.cn

刘敦强,男,博士研究生,主要研究方向为激光雷达数据处理。E-mail: dqliu@stu.xmu.edu.cn

邢乐园,女,博士研究生,主要研究方向为激光雷达数据处理。E-mail: xly_0622@163.com

李明哲,女,博士研究生,主要研究方向为激光雷达数据处理。E-mail: limingzhe@stu.xmu.edu.cn

郭裕兰,男,教授,主要研究方向为计算机视觉。E-mail: guoyulan@sysu.edu.cn

王程,通信作者,男,教授,主要研究方向为激光雷达数据处理。E-mail: cwang@xmu.edu.cn