

DOI: 10.3969/j.issn.1671-0673. 2026. 04. 002

基于车载激光雷达 BEV 旋转不变特征的城市道路场景地点识别与定位方法

周梦鸽, 曹雪峰, 何静怡

(信息工程大学, 河南 郑州 450001)

摘要: 面向城市道路场景自主定位任务, 提出一种基于车载激光雷达鸟瞰图 (BEV) 旋转不变特征的地点识别与定位方法。首先, 将车载激光雷达三维点云通过归一化点密度方法转换为二维 BEV 图像; 其次, 分析激光雷达 BEV 平移与旋转不变性特征, 设计旋转平移等变的 BEV 图像局部特征提取网络 (RE-Net) 和全局特征聚合方法; 最后, 实现基于车载激光雷达 BEV 旋转不变特征的城市道路场景地点识别与定位。实验结果表明, 该方法在 NCLT 上的 Recall@1 相比主流方法至少提升 5 个百分点, 在自采集城市广角激光雷达数据集上 Recall@1 高于 90%, 并证明了训练数据多样性比数据规模更能提升模型泛化能力。

关键词: 地点识别; 激光雷达; 鸟瞰图; 旋转不变; 城市道路场景

中图分类号: P237; TP391

文献标识码: A

文章编号: 1671-0673(2026)04-0391-08

Place Recognition and Localization Method for Urban Road Scenes Based on Vehicle-Mounted LiDAR BEV Rotation-Invariant Features

ZHOU Mengge, CAO Xuefeng, HE Jingyi

(Information Engineering University, Zhengzhou 450001, China)

Abstract: A place recognition and localization method based on vehicle-mounted LiDAR bird's-eye view (BEV) rotation-invariant features is proposed for autonomous positioning in urban road scenes. Firstly, three-dimensional point clouds acquired by vehicle-mounted LiDAR are transformed into two-dimensional BEV images through a normalized point-density representation. Subsequently, the translation and rotation invariance properties of LiDAR BEV features are analyzed. On this basis, a rotation-translation-equivariant local feature extraction network for BEV images, termed rotation equivariant local feature extraction network (RE-Net), is designed, together with a global feature aggregation strategy. Finally, place recognition and localization in urban road scenes are performed based on the rotation-invariant BEV features extracted from vehicle-mounted LiDAR data. The experimental results show that the proposed method improves Recall@1 on the NCLT dataset by at least 5 percentage points compared with mainstream methods, achieves a Recall@1 of over 90% on the self-collected urban wide-angle LiDAR dataset, and demonstrates that training data diversity contributes more than data scale to model generalization.

Key words: place recognition; LiDAR; bird's-eye view; rotation-invariance; urban road scenes

激光雷达能够主动获取车辆周围环境的三维点云数据, 相比车载摄像机对光照变化、弱纹理或

无纹理场景更具鲁棒性, 在激光雷达地点识别 (LiDAR Place Recognition, LPR)、全局定位 (Global Lo-

收稿日期: 2026-04-20; 修回日期: 2026-05-08

作者简介: 周梦鸽 (2000—), 女, 硕士生, 主要研究方向为激光雷达地点识别。E-mail: blingzhou0707@163.com

通信作者: 曹雪峰 (1983—), 男, 副教授, 博士, 主要研究方向为无人平台环境感知。E-mail: CAO_Xue_Feng@163.com

calization)等任务中有独特优势^[1]。但因激光点云存在稀疏、无序和非结构化的特点,如何对其进行有效表征是激光雷达地点识别面临的重要挑战^[2]。

根据激光点云表征方式的不同,现有的LPR方法大致可分为基于特征、基于投影和基于分割等3类^[3]。基于投影的LPR方法兼顾效率与性能,是当前受到重要关注的研究方向。2021年,Xu等^[4]构建多层BEV,利用频域相似性度量提取旋转不变的全局特征。但是,当输入点云之间的重叠度较低时,尤其是在非重叠区域包含相似结构的场景中,其配准容易失败。Luo等^[5]设计了一种基于Log-Gabor滤波器的鸟瞰图特征变换(Bird's-Eye View Feature Transform, BVFT)描述,以应对BEV图像旋转和强度变化。BVFT特征聚合依赖手工设计的词袋模型,导致在陌生环境中的泛化能力较差。2023年,Luo等^[6]设计了基于群卷积的旋转等变特征提取与NetVLAD^[7]全局特征聚合方法。Xu等^[8]提出针对大视角差异的旋转平移不变性的Gram表示方法,但面对复杂多变的实际场景,需要进一步提升描述符对环境外观变化的适应能力。2025年,Shen等^[9]利用不同高度的多层BEV密度图像生成旋转不变的全局描述符。同年,Wang等^[10]利用扩散模型从BEV特征中迭代恢复绝对位姿。Zhang等^[11]提出自监督BEV定位框架,显著降低了数据采集与预处理成本。但面对城市场景的复杂性和激光扫描的多样性,现有研究在适应性、泛化性等方面仍存在不足。

现有地点识别研究使用的点云数据主要由固定安置在车辆上的激光雷达采集,水平视场角达到360°,垂直视场角集中在25°~40°,最远探测距离在100 m左右。随着激光雷达新结构与小型化发展,出现了通过旋转本体实现广角远距离高点频扫描的激光雷达设备,显著增加了对道路周围景物侧立面和更远距离的观测能力,为激光雷达BEV地点识别提供了新的研究场景。已有Wild-Places^[12]数据集等研究验证了广角扫描数据在野外环境的应用价值。但大规模城市环境中,如何对这类点云数据高效表示与检索还需深入研究。

面向城市场景车载激光雷达地点识别任务,本文基于归一化点密度的激光雷达BEV图像,进行平移与旋转不变性特征分析,设计旋转平移等变的BEV图像特征提取网络,在公开数据集和城市场景广角车载激光雷达数据上开展BEV地点识别实验,验证所提方法的有效性与泛化能力。

1 车载激光雷达BEV城市道路场景地点识别与定位流程

基于车载激光雷达BEV的城市道路场景地点识别与定位任务通常分为3个阶段。首先,将车载激光雷达的三维点云转换为BEV图像,构建覆盖车辆行驶路线的BEV数据库。其次,以当前BEV作为查询,在数据库中进行相似度搜索,得到与之最相似的BEV作为结果帧,实现地点识别。最后,通过查询帧与检索匹配帧之间的特征匹配或配准计算,估计查询帧相对于匹配帧的相对位姿,并结合检索匹配帧的已知全局位姿,获得查询帧的全局位置与姿态,从而完成定位。

针对车载激光雷达在城市道路行驶过程中存在的平移和旋转变换,设计旋转平移鲁棒的BEV图像特征提取方法。首先,利用卷积神经网络(Convolutional Neural Network, CNN)的局部感受野和权值共享特性,设计提取平移变换高区分度BEV图像局部特征。其次,通过多角度旋转输入图像,提取对应方向特征,反向旋转对齐及方向最大池化,获得旋转等变的局部特征表示。最后,通过NetVLAD聚合局部特征,构建用于地点检索的旋转平移鲁棒的全局描述符,从而提升城市道路场景车载激光雷达地点识别在车辆位姿变化条件下的准确性。

2 激光雷达BEV图像特征提取

2.1 基于归一化点密度的激光雷达BEV生成方法

使用归一化点密度方法生成BEV图像的过程中,像素值为落入该像素范围内的激光点云相对密度。首先,以车载激光雷达自身为中心,定义一个右手笛卡尔坐标系。其中: X 轴指向车体右侧; Y 轴指向车头方向; Z 轴指向车体上方; XY 平面与地面近似平行。给定一帧激光点云 $\mathcal{P}=\{P_i|i=1,\dots,N_p\}$,其中: $\mathcal{P}$ 表示该帧点云的点集; N_p 表示点云中包含的激光点总数。第 i 个激光点 $P_i=(x_i,y_i,z_i)$,记录了该点在坐标系中的位置。然后,在该坐标系 XY 平面上,取 $[-D_m,D_m]$ 的窗口,仅考虑落入该窗口之内的点云。将该窗口划分成尺寸为 $g\text{ m}$ 的规则格网。之后,使用体素滤波器(体素尺寸为 $g\text{ m}$)对窗口内的点云进行均匀下采样,减少同一体素内重复的点。记 N_m 为归一化因子,表示单个格网单元内允许的最大点数。设 $I_{u,v}$ 表示BEV图像中第 v 行、第 u 列像素的像素值,统计每个单元

格内包含的点数 N_g , 按式(1)计算其归一化点密度的数值作为 BEV 图像像素值。

$$I_{u,v} = \min(N_g, N_m) / N_m \quad (1)$$

由此, 每个网格单元对应 BEV 图像中的一个像素, 生成尺寸为 $\lceil 2D/g \rceil \times \lceil 2D/g \rceil$ 的 BEV 图像 $I_{u,v}$ 。

归一化点密度 BEV 图像中编码点云的空间分布密度, 而非网格内最大高度^[13]。图1展示了不同 BEV 生成方法对比。高度 BEV 易受传感器朝向和扫描距离影响, 而归一化点密度对视角变化更鲁棒, 能够更稳定地反映地面、建筑物及道路设施的几何轮廓。

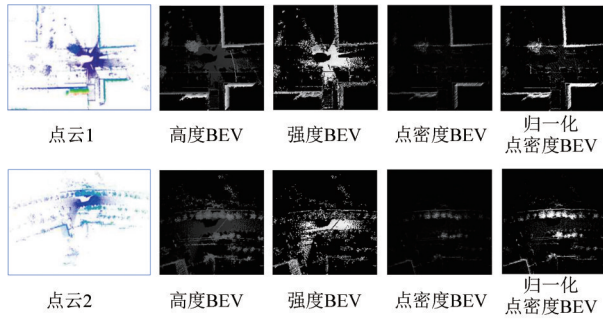


图1 激光雷达 BEV 生成方法对比

2.2 平移与旋转变换下 BEV 图像特征的等变性和不变性分析

车辆在局部区域的运动可近似视为平面移动, 主要是估计位姿中的三自由度参数平移 (x, y) 及偏航角, 其余自由度可由数据库中匹配帧的已有位姿补充获得。

CNN 具备平移等变性^[14]。当车辆发生平移运动时, 对应的 BEV 图像整体发生了空间位移, 但图像里的场景内容保持不变。为验证 CNN 对 BEV 图像特征的平移区分能力, 将 BEV 图像的特征提取过程表示为

$$\mathbf{F} = \phi(\mathbf{I}) \quad (2)$$

式中: $\mathbf{I} \in \mathbb{R}^{H \times W}$, 表示输入的 BEV 图像; H 和 W 分别表示图像的高度和宽度; 特征提取器 ϕ 提取 BEV 图像得到特征 $\mathbf{F} \in \mathbb{R}^{H \times W \times C}$; C 为特征通道数。

设 $f_{u,v}$ 表示特征图中像素位置 (u, v) 处的局部特征, u 和 v 分别表示 BEV 图像的列索引和行索引。位置 (u, v) 与偏移后位置 $(u + \delta u, v + \delta v)$ 处局部特征的欧氏距离为

$$d(f_{u,v}, f_{u+\delta u, v+\delta v}) = \|f_{u,v} - f_{u+\delta u, v+\delta v}\|_2 \quad (3)$$

式中: δu 和 δv 分别表示相对于位置 (u, v) 在图像列方向和行方向上的像素偏移量。

通常情况下, 在 BEV 图像空间中, 位置较近的

关键点具有更相似的局部结构, 因此其特征距离较小; 而位置较远的关键点由于对应的局部场景结构差异更大, 其特征距离通常较大。设 $(\delta u_1, \delta v_1)$ 和 $(\delta u_2, \delta v_2)$ 分别表示相对于位置 (u, v) 的两组像素偏移量。当两组偏移量满足 $\|(\delta u_1, \delta v_1)\| < \|(\delta u_2, \delta v_2)\|$ 时, 其特征距离通常满足:

$$d(f_{u,v}, f_{u+\delta u_1, v+\delta v_1}) < d(f_{u,v}, f_{u+\delta u_2, v+\delta v_2}) \quad (4)$$

式中, $(u + \delta u_1, v + \delta v_1)$ 和 $(u + \delta u_2, v + \delta v_2)$ 分别表示距离位置 (u, v) 较近和较远的两个特征位置。式(4)描述了局部特征相似性随空间距离变化的一般趋势, 即 CNN 提取的局部特征能够在一定程度上刻画不同位置处的局部场景结构, 并保留关键点之间的相对空间关系。这种特性有助于建立稳定的特征对应关系, 从而提高后续位姿估计的准确性。

给定一幅由图像 $\mathbf{I}$ 经过平移变换得到的 BEV 图像 $\mathbf{I}'$ 。设原图像中某一关键点的位置为 (u, v) , 平移后该关键点在图像 $\mathbf{I}'$ 中的对应位置为 (u', v') 。理想情况下, $\mathbf{I}'$ 中的关键点特征 $f'_{u', v'}$ 应当与其在原图像 $\mathbf{I}$ 中对应位置的关键点特征 $f_{u,v}$ 保持一致。同时, 为保证特征对应关系的唯一性, 对应位置之间的特征距离应小于非对应位置之间的特征距离, 即

$$d(f'_{u', v'}, f_{u,v}) < d(f'_{u', v'}, f_{u+\delta u, v+\delta v}), (\|(\delta u, \delta v)\| > 0) \quad (5)$$

当 $\|(\delta u, \delta v)\| > 0$ 时, 位置 $(u + \delta u, v + \delta v)$ 与位置 (u, v) 不重合。式(5)表明, 经过平移变换后, 对应关键点之间的特征距离应最小, 而非对应位置的特征距离应相对较大。因此, 可以通过最近邻搜索建立 BEV 特征对应关系, 并结合随机抽样一致性 (Random Sample Consensus, RANSAC) 算法剔除错误匹配, 进而求解车辆的相对位姿。

考虑到视角旋转会影响 BEV 特征稳定性, 后续可通过随机旋转数据增强^[15], 使网络学习同一点在不同角度下的一致特征表示, 从而提高特征的旋转鲁棒性。

2.3 旋转等变的局部特征提取网络

为增强 BEV 图像特征对旋转变化的鲁棒性, 本文设计旋转等变的局部特征提取网络 (Rotation Equivariant Local Feature Extraction Net, RE-Net)。RE-Net 以 CNN 为基础, 通过输入 BEV 图像进行多角度旋转、权重共享特征提取、逆旋转对齐以及跨角度最大响应聚合, 从而实现旋转等变性。

给定一个输入的 BEV 图像 $\mathbf{I}$, 定义离散旋转角

度集合 $\mathcal{R} = \left\{0, \frac{2\pi}{N}, \dots, \frac{2\pi(N-1)}{N}\right\}$, 对旋转图像分别提取特征, 并将特征图逆旋转回原始方向后进行最大池化, 得到旋转等变局部特征 F 。

$$F = \max_{r \in \mathcal{R}} R_r^{-1} \circ \psi(R_r \circ I) \quad (6)$$

式中: ψ 表示权重共享的残差卷积模块; R_r 表示角度为 r 的旋转变换; R_r^{-1} 为其逆变换。局部特征提取网络如图2所示。

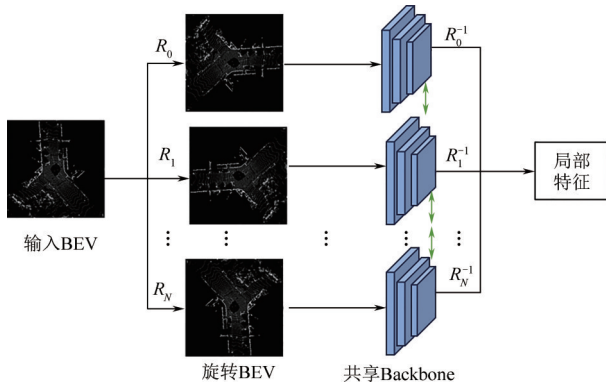


图2 旋转等变的局部特征提取网络

若输入图像发生角度为 α 的旋转, 且 $\alpha \in \mathcal{R}$, R_α 表示旋转角度为 α 的旋转变换。此时, 旋转后图像对应的输出特征图 F' 满足:

$$F' = R_\alpha \circ F \quad (7)$$

式(7)表明, 当输入BEV图像旋转角度 α 时, 输出特征图也会产生相同角度的空间旋转, 即输入图像和输出特征图具有一致的旋转变换关系。因此, RE-Net具有旋转等变性。由于CNN池化和下采样操作对微小空间偏移具有一定鲁棒性, 只需选取有限个旋转角度即可获得较好的近似效果。

为直观验证上述性质, 将RE-Net与ResNet的特征提取结果进行对比。分别使用ResNet和RE-Net从一对存在旋转关系的BEV图像中提取特征, 结果如图3所示。

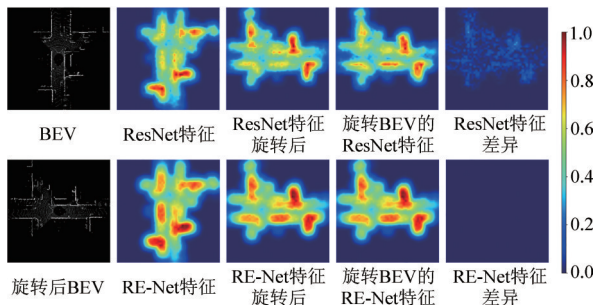


图3 RE-Net与ResNet特征旋转变换比较

旋转输入后, RE-Net提取的特征与原始特征经相同旋转后的结果基本一致, 差异趋近于零, 而

ResNet特征变化较大, 说明RE-Net具有更好的旋转等变性。

2.4 旋转不变的全局特征描述符

为实现稳健的地点识别, 本文在旋转等变局部特征的基础上进一步构建旋转不变的全局描述符。本文将RE-Net与NetVLAD进行级联, 利用RE-Net提供旋转等变的局部特征, 再由NetVLAD将其聚合为旋转不变的全局描述符。

NetVLAD通过聚类中心和软分配机制, 将局部特征残差加权累加为全局特征。设RE-Net输出的特征图为 F , 将其展平为局部特征集合 $\{f_i \in \mathbb{R}^C\}_{i=1}^{H'W'}$, 其中, H' 和 W' 分别表示经过RE-Net下采样后特征图的高度和宽度。设NetVLAD包含 K 个聚类中心, 对于第 k 个聚类中心 c_k 对应的全局特征分量为

$$V_k = \sum_{i=1}^{H'W'} v_k(f_i) = \sum_{i=1}^{H'W'} \frac{\exp(w_k^T f_i + b_k)}{\sum_{k'=1}^K \exp(w_{k'}^T f_i + b_{k'})} (f_i - c_k) \quad (8)$$

式中: w_k 和 b_k 分别为第 k 个聚类中心对应的可学习权重和偏置; k' 是软分配权重分母中的求和索引, $k' \in \{1, 2, \dots, K\}$, 用于遍历所有聚类中心。

设 $\{f_i\}$ 和 $\{f'_i\}$ 分别为原始BEV图像和旋转BEV图像经RE-Net提取的局部特征集合。等变性, 即输入图像旋转后, 其局部特征仅发生空间位置排列变化, 特征集合中的元素保持一致, 仅排列顺序不同。

又由于NetVLAD采用求和形式进行聚合, 具有置换不变性, 因此有:

$$V'_k = v_k(f'_1) + v_k(f'_2) + \dots + v_k(f'_n) = \sum_i v_k(f'_i) = \sum_i v_k(f_i) = V_k \quad (9)$$

该关系对所有 $k = \{1, 2, \dots, K\}$ 成立, 因此 $V'_k = V_k$ 。RE-Net的旋转等变局部特征经过NetVLAD聚合后, 可转化为旋转不变的全局描述符, 从而满足地点识别对旋转不变性的要求。

3 实验与结果分析

3.1 实验设计

采用首位召回率(Recall@1)为地点识别的评估指标。若查询帧与Top-1匹配帧之间的距离在5 m以内, 视为成功定位。为衡量描述符对正确匹配与错误匹配的区分能力, 进一步采用最大F1分数(max F1 score), 通过遍历描述符距离阈值, 统计不同阈值下的精确率(Precision)、召回率

(Recall)和 $F1$ 分数,并取最大 $F1$ 分数作为最终结果。

计算平均平移误差和旋转误差评估位姿估计。采用定位成功率(Success Rate, SR)综合评估整体定位性能,当平均平移误差小于 2 m 且旋转误差低于 5° 时认为定位成功。

选择 RING++^[8]、BEVPlace++^[16]作为对比方法,将点云过滤到相同范围[-40 m, 40 m]。设置 BEV 的网格大小为 200×200 ,分辨率为 $80/200=0.4$ m/px,旋转分辨率为 $360/200=1.8$ ($^\circ$)/px。对比方法采用其开源代码,并按默认设置重新复现结果。实验在 RTX 3090 GPU 设备上运行,共进行 50 轮次的训练。

3.2 数据集

采用 NCLT^[17]和城市道路场景车载广角激光扫描数据集开展测试。城市道路场景车载广角激光扫描数据集采集于郑州市区,涵盖城市主干道、小区道路、环湖道路、非结构化区域等多种场景,空间结构上差异明显。

该城市道路场景车载广角激光雷达数据集包含西三环(XSH)、洞林湖(DLH)、思念果岭(SNG)、中牟(ZM)4 个区域,共 19 个序列,总里程达到 128.31 km,累计生成 57 637 帧点云子图,具体统计信息如表 1 所示。

表 1 城市道路场景车载广角激光雷达数据集				
区域	序列	长度/km	子图帧数	说明
西三环 (XSH)	XSH-01	4.25	1 912	同一区域,
	XSH-02	5.76	2 631	01 与 02 反向
	XSH-03	3.75	1 664	同一区域,
	XSH-04	4.67	2 099	03 与 04 反向
	XSH-05	4.90	2 184	同一区域,
	XSH-06	5.02	2 270	05 与 06 反向
	XSH-07	3.84	1 724	同一区域
	XSH-08	4.91	2 200	
	XSH-09	4.55	1 998	
	XSH-10	10.08	4 502	
洞林湖 (DLH)	DLH-01	13.41	5 892	双向道路+反向
	DLH-02	9.33	4 120	双向道路+反向
思念果岭 (SNG)	SNG-01	4.04	1 800	开阔绿地+反向
	SNG-02	9.27	4 125	开阔绿地+反向
中牟 (ZM)	ZM-01	8.93	4 010	同一区域,
	ZM-02	9.22	4 192	01 与 02 反向
	ZM-03	9.84	4 576	小区+
	ZM-04	9.92	4 531	03 与 04 反向
	ZM-05	2.62	1 207	小区

采集设备为其域灵光 L2Pro 32 线广角激光雷达。该设备集成激光雷达、视觉与惯性测量单元(Inertial Measurement Unit, IMU)传感器,具备实时动态差分(Real-Time Kinematic, RTK)融合功能,可直接输出点云数据。L2Pro 32 线激光雷达的点频每秒 64 万点,最大探测距离 300 m,采取绕 Z 轴以约 0.5 Hz 的速度旋转,视场角达到 $360^\circ \times 270^\circ$,能够有效获取周围环境的侧立面信息。采集时,将设备安装在车辆顶部,以 20 km/h 的平均速度在郑州市内进行数据采集。

3.3 地点识别实验结果分析

3.3.1 NCLT 地点识别

地点识别实验选取 NCLT 数据集中“2012-01-15”序列作为数据库,使用序列“2012-02-04”(序列 1)、“2012-03-17”(序列 2)、“2012-06-15”(序列 3)、“2012-09-28”(序列 4)、“2012-11-16”(序列 5)以及“2013-02-23”(序列 6)作为查询序列进行测试。实验结果如表 2 所示。

表 2 在 NCLT 数据集不同序列上的 Recall@1 性能比较 单位:%							
方法	序列1	序列2	序列3	序列4	序列5	序列6	平均
RING++	69.00	80.15	56.22	74.66	66.56	59.06	67.61
BEVPlace++	95.05	93.42	91.35	88.40	89.70	86.60	90.75
本文方法	97.89	96.86	96.94	94.37	95.80	92.80	95.78

实验结果表明,本文方法在所有测试序列上均取得最优性能,尤其在跨季节变化显著的序列 6 中依然保持超过 92% 的首位召回率,具有较强的环境适应能力。图 4 为 BEVPlace++ 与本文方法地点识别结果对比,正确匹配结果用绿色标注,错误匹配结果用红色标注。BEVPlace++ 方法在转弯区域与路段交叉处出现连续性红色错误匹配点。相比之下,本文方法的错误匹配点数量明显减少,绿色正确匹配点覆盖了绝大部分轨迹。

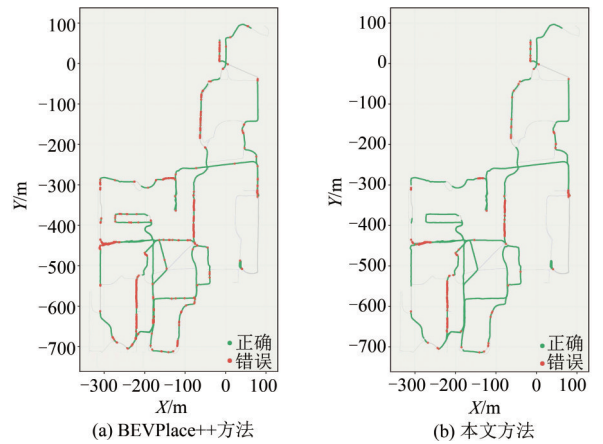


图 4 BEVPlace++与本文方法地点识别结果可视化

3.3.2 广角 LiDAR 地点识别

在自采集的城市道路场景车载广角激光扫描数据集上开展地点识别实验。实验采用分区域训练与测试的策略:1)在 XSH 区域,以序列 XSH-01 作为训练数据,在 XSH 其余序列上测试;2)在 DLH 区域,以 DLH-01 序列的前 3 500 帧作为训练数据,并在 DLH-01、DLH-02、SNG-01 和 SNG-02 序列上进行测试,测试时,分别选取 DLH-01、DLH-02、SNG-01 和 SNG-02 序列的前 3 500 帧、前 2 000 帧、前 1 000 帧和前 2 200 帧构建地点数据库,并将各序列剩余的帧作为查询;3)在 ZM 区域,以 ZM-01 作为训练数据,在 ZM 其他序列上测试。训练所用的数据样本数量基本一致。

表 3 展示了本文方法在自采集数据集各子序列上的 Recall@1 和最大 F1 分数性能。表中的“查询→数据库”表示地点识别时查询序列与数据库序列的对应关系。对于单序列测试,则以前若干帧作为数据库,其余帧作为查询数据。

表 3 广角 LiDAR 数据集上 Recall@1 与最大 F1 分数 单位: %			
训练数据	查询→数据库	Recall@1	最大 F1 分数
XSH-01	XSH-01→XSH-02	90.67	84.40
	XSH-03→XSH-04	90.41	87.97
	XSH-05→XSH-06	99.40	98.42
	XSH-07→XSH-08	91.88	85.52
	XSH-09→XSH-10	97.34	95.56
DLH-01 前 3 500 帧	DLH-01, 前 3 500 帧作为库	94.30	83.39
	DLH-02, 前 2 000 帧作为库	92.88	87.12
	SNG-01, 前 1 000 帧作为库	94.65	82.82
	SNG-02, 前 2 200 帧作为库	97.95	89.06
ZM-01	ZM-01→ZM-02	90.04	78.39
	ZM-03→ZM-04	99.30	96.50
	ZM-05, 前 600 帧作为库	99.49	99.49

从表 3 可以看出,本文方法在所有测试子序列上的 Recall@1 都超过了 90%,在真实城市道路场景中具有较好的地点识别稳定性。部分场景虽然 Recall@1 较高,但最大 F1 分数相对较低,主要原因是这些区域几何结构相似或稳定局部特征较少,导致错误候选帧与查询帧的描述符距离较近,产生一定假阳性匹配。总体来看,本文方法对车辆视角变化和场景结构差异具有良好的鲁棒性。

3.3.3 训练数据组成对地点识别性能的影响

为分析训练数据的规模与组成对地点识别性能的影响,本文在自采集数据集 DLH 与 SNG 区域开展对比实验。DLH-01 为洞林湖环湖双向道路序

列,车辆约在第 2 800 帧处掉头返回,前后轨迹方向相反;DLH-02 为洞林湖内部道路序列,SNG 区域则包含植被、自然道路和低层建筑等非结构化场景。

根据 DLH-01 轨迹特点,本文设计了 4 种训练数据划分方式:1)前 2 800 帧,仅包含单向行驶轨迹;2)第 700~3 500 帧,数据规模与前 2 800 帧接近,但包含往返轨迹;3)第 2 100~3 500 帧,数据量最少,但涵盖明显的方向变化;4)前 3 500 帧,训练所用数据量最大,同时包含单向与双向数据。测试时,各序列的数据库与查询集划分保持一致。实验结果如表 4 所示。

表 4 不同 DLH-01 训练区间下的 Recall@1 对比 单位: %					
训练区间	DLH-01	DLH-02	SNG-01	SNG-02	平均
0~2 800	87.70	89.32	91.71	97.75	91.62
700~3 500	93.79	93.27	94.39	97.85	94.83
2 100~3 500	92.18	90.68	91.71	96.99	92.89
0~3 500	94.30	92.88	94.65	97.95	94.95

由表 4 可知,训练数据的视角多样性对地点识别性能具有重要影响。仅包含单向轨迹的 0~2 800 帧平均 Recall@1 为 91.62%,而数据规模相近但包含往返轨迹的 700~3 500 帧提升至 94.83%,这说明了多视角覆盖比单纯增加数据量更有利于特征学习。虽然 2 100~3 500 帧数据量最少,但其平均召回率仍高于 0~2 800 帧,进一步说明方向变化有助于提升模型鲁棒性。0~3 500 帧取得最高平均 Recall@1 为 94.95%,但相比 700~3 500 帧仅提升 0.12 个百分点,说明当训练数据已覆盖主要视角变化后,继续增加相似场景数据带来的性能增益有限。

3.3.4 特征提取网络的泛化能力

为评估特征提取网络的跨场景泛化能力,本文设计了 5 种训练配置,前 3 种为单序列训练,分别使用 XSH-01、DLH-01 前 3 500 帧、ZM-01 序列作为训练数据;第 4 种为同场景多序列联合训练,采用 XSH 区域的 01、03 和 05 序列进行联合训练,记为 XSH(3);第 5 种为跨场景联合训练,采用 XSH(3)+DLH-01 序列共同训练。完成训练后,分别在 XSH(01→02)、DLH-01 和 ZM(01→02)这 3 组代表性测试数据上进行地点识别实验,以比较不同训练配置下模型的同场景及跨场景泛化性能,结果如表 5 所示。

由表 5 可见,单序列训练的跨场景泛化能力与训练数据本身的场景多样性密切相关。仅使用 XSH-01 训练配置时,模型在本场景上 Recall@1 为

90.67%,但在DLH场景下降至78.65%,这说明单一场景训练难以适应结构差异较大的新环境。相比之下,DLH-01训练配置在3个测试场景上均取得较高Recall@1,这与DLH-01包含往返轨迹和多种道路结构有关。虽然XSH(3)增加了同区域训练数据,但平均Recall@1仅为88.86%,这表明仅增加同类场景数据对泛化能力的提升有限。跨区域联合训练取得最优结果,平均Recall@1达到97.17%,也说明了包含不同空间结构和行驶方向的训练数据能够显著增强特征提取网络的泛化能力。

表5 不同训练配置下的跨场景地点识别Recall@1比较 单位:%

训练配置	XSH(01→02)	DLH-01	ZM(01→02)	平均
XSH-01	90.67	78.65	85.56	84.96
DLH-01	93.06	94.30	92.58	93.31
ZM-01	91.23	86.93	90.04	89.40
XSH(3)	92.01	84.43	90.15	88.86
XSH(3)+ DLH-01	96.48	98.29	96.73	97.17

综上,训练数据的场景多样性是影响模型泛化能力的关键因素。跨场景、多方向的数据能够帮助网络学习更鲁棒的全局描述符,从而提升其在未知环境中的地点识别性能。

3.3.5 性能分析

在地点识别的基础上,本文进一步从位姿估计精度与在线定位效率两个方面评估所提方法的性能。位姿估计实验沿用与表3地点识别实验相同的分区域训练与测试策略,评估指标包括定位成功率(SR)、平均平移误差 $\hat{e}_t$ 和平均旋转误差 $\hat{e}_r$ 。结果如表6所示。

表6 广角激光雷达数据集位姿估计性能

特征提取网络 训练数据	查询帧→数据库	SR/%	$\hat{e}_t/\text{m}$	$\hat{e}_r/(\text{^\circ})$
XSH-01	XSH-01→XSH-02	95.67	0.93	0.69
	XSH-03→XSH-04	97.69	0.45	0.37
	XSH-05→XSH-06	98.43	0.37	0.32
	XSH-07→XSH-08	99.15	0.23	0.23
	XSH-09→XSH-10	98.59	0.29	0.25
DLH-01 前3 500帧	DLH-01,前3 500帧作为库	99.04	0.63	0.42
	DLH-02,前2 000帧作为库	95.15	0.67	0.45
	SNG-01,前1 000帧作为库	89.30	0.83	0.81
	SNG-02,前2 200帧作为库	99.68	0.69	0.52

总体来看,本文方法在大多数测试序列上均取得了95%以上的定位成功率,平均平移误差 $\hat{e}_t$ 均低于1 m,平均旋转误差 $\hat{e}_r$ 低于1°。在多个场景中

的位姿估计精度保持在可接受范围内,在真实城市环境可以进行稳定、有效的全局定位。

在RTX 3090 GPU平台上测试在线定位耗时。以NCLT数据集“2012-01-15”序列为地图库、“2013-02-23”序列为查询序列,重复推理500次取均值。结果显示,描述符生成耗时19.917 ms,Top-1检索耗时0.541 ms,单帧总定位耗时20.554 ms,满足实时定位需求,本文方法具有较好的工程实用性。

4 结束语

面向城市道路场景车载激光雷达地点识别与全局定位任务,本文提出一种基于归一化点密度BEV的旋转等变特征提取网络,建立了从点云投影、特征提取、检索匹配到位姿估计的完整定位流程。实验结果表明,本文方法在NCLT数据集上优于现有主流方法,且训练数据的场景与视角多样性对模型泛化能力的提升作用显著强于数据规模,为激光雷达地点识别的模型训练与数据构建提供了实用参考。

此外,本文构建的郑州城市场景广角激光雷达数据集可为相关研究提供真实场景支撑。本文方法仍存在一定局限性。实验结果表明,在环湖道路、开阔绿地等几何结构相似、稳定局部特征较少的场景中,部分正确匹配与错误匹配在描述符空间中的区分度仍然有限。本文方法提高了在旋转变换、反向行驶、跨季节与跨场景条件下地点识别与定位的鲁棒性,下一步工作将结合几何结构增强、多尺度特征融合与自监督学习进一步提升在复杂城市道路场景中的鲁棒性与泛化能力。

参考文献:

- [1] Uy M A, Lee G H. PointNetVLAD: deep point cloud based retrieval for large-scale place recognition[C]//Proceedings of the 2018 IEEE/CVF conference on computer vision and pattern recognition. Piscataway, USA: IEEE, 2018:4470-4479.
- [2] Yin H, Xu X C, Lu S, et al. A survey on global LiDAR localization: challenges, advances and open problems[J]. International Journal of Computer Vision, 2024, 132(8):3139-3171.
- [3] Luo K, Yu H S, Chen X, et al. 3D point cloud-based place recognition: a survey[J]. Artificial Intelligence Review, 2024, 57(4):83.
- [4] Xu X C, Yin H, Chen Z X, et al. DiSCO: differentiable scan context with orientation[J]. IEEE Robotics and Auto-

- mation Letters, 2021, 6(2):2791-2798.
- [5] Luo L, Cao S Y, Han B, et al. BVMatch: LiDAR-based place recognition using bird's-eye view images [J]. IEEE Robotics and Automation Letters, 2021, 6(3):6076-6083.
- [6] Luo L, Zheng S H, Li Y X, et al. BEVPlace: learning LiDAR-based place recognition using bird's eye view images [C]//Proceedings of the 2023 IEEE/CVF international conference on computer vision. Piscataway, USA: IEEE, 2023: 8666-8675.
- [7] Arandjelovic R, Gronat P, Torii A, et al. NetVLAD: CNN architecture for weakly supervised place recognition [C]//Proceedings of the 2016 IEEE conference on computer vision and pattern recognition. Piscataway, USA: IEEE, 2016:5297-5307.
- [8] Xu X C, Lu S, Wu J, et al. RING: roto-translation invariant gram for global localization on a sparse scan map [J]. IEEE Transactions on Robotics, 2023, 39(6):4616-4635.
- [9] Shen Y Q, Tuna T, Hutter M, et al. ForestLPR: LiDAR place recognition in forests attentioning multiple BEV density images [C]//Proceedings of the 2025 IEEE/CVF conference on computer vision and pattern recognition. Piscataway, USA: IEEE, 2025:6659-6669.
- [10] Wang Z Y, Shi C H, Wang N, et al. BEVDiffLoc: end-to-end LiDAR global localization in BEV view based on diffusion model [C]//Proceedings of the 2025 IEEE/RSJ international conference on intelligent robots and systems. Piscataway, USA: IEEE, 2025:16240-16246.
- [11] Zhang C H, Luo L, Cao S Y, et al. S-BEVLLoc: BEV-based self-supervised framework for large-scale LiDAR global localization [J]. IEEE Robotics and Automation Letters, 2025, 10(10):9614-9621.
- [12] Knights J, Vidanapathirana K, Ramezani M, et al. Wildplaces: a large-scale dataset for LiDAR place recognition in unstructured natural environments [C]//Proceedings of the 2023 IEEE international conference on robotics and automation. Piscataway, USA: IEEE, 2023:11322-11328.
- [13] Li Y H, Li H. LiDAR-based initial global localization using two-dimensional (2D) submap projection image (SPI) [C]//Proceedings of the 2021 IEEE international conference on robotics and automation. Piscataway, USA: IEEE, 2021:5063-5068.
- [14] Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86(11):2278-2324.
- [15] 张志超, 徐友春, 陈晋生, 等. 基于航空图像与地面点云的跨模态地点识别方法 [J]. 兵工学报, 2026, 47(2): 44-55.
- [16] Luo L, Cao S Y, Li X R, et al. BEVPlace++: fast, robust, and lightweight LiDAR global localization for autonomous ground vehicles [J]. IEEE Transactions on Robotics, 2025, 41:4479-4498.
- [17] Carlevaris-Bianco N, Ushani A K, Eustice R M. University of Michigan North Campus long-term vision and lidar dataset [J]. The International Journal of Robotics Research, 2016, 35(9):1023-1035.

(编辑: 邵 华)

(上接第 390 页)

- [55] Zhang Y R, Arenas A, Yağan O. Cascading failures in interdependent systems under a flow redistribution model [J]. Physical Review E, 2018, 97(2):022307.
- [56] Wu J J, Gao Z Y, Sun H J. Effects of the cascading failures on scale-free traffic networks [J]. Physica A Statistical Mechanics & Its Applications, 2007, 378(2):505-511.
- [57] Pei J X, Liu Y, Wang W, et al. Cascading failures in multiplex network under flow redistribution [J]. Physica A: Statistical Mechanics and Its Applications, 2021, 583: 126340.
- [58] Jia C X, Liu R R. Cascading dynamics in double-layer hypergraphs with higher-order inter-layer interdependencies [J]. Reliability Engineering & System Safety, 2025, 257: 110841.
- [59] Qian C, Zhao D D, Zhong M, et al. Modeling and analysis of cascading failures in multilayer higher-order networks [J]. Reliability Engineering & System Safety, 2025, 253:110497.
- [60] Zheng K X, Liu Y, Wang Y, et al. K-core percolation on interdependent and interconnected multiplex networks [J]. Europhysics Letters, 2021, 133(4):48003.
- [61] Hu H F, Liu Z N, Wu F. RL-PMO: a reinforcement learning-based optimization algorithm for parallel SFC migration [J]. Sensors, 2026, 26(1):242.
- [62] Maurer A, Tixeuil S, Défago X. Reliable communication in a dynamic network in the presence of Byzantine faults [PP/OL]. arxiv (2015-02-16) [2026-05-22]. <https://arxiv.org/abs/1402.0121>.
- [63] Liu L, Zhang L, Liao S, et al. A generalized approach to solve perfect Bayesian Nash equilibrium for practical network attack and defense [J]. Information Sciences, 2021, 577:245-264.

(编辑: 冯 春)