

DOI: 10.19650/j.cnki.cjsi.J2513742

基于双目相机深度估计的相机-LiDAR 端到端外参标定方法

刘秋骅, 徐晓苏

(东南大学仪器科学与工程学院 南京 210096)

摘 要:准确可靠的传感器外参标定方法是相机-激光雷达融合系统实现高精度定位与导航的关键。然而,现有的端到端相机-激光雷达外参标定方法仍存在诸多不足,如模型参数量大、模态特征相关性计算不匹配等问题。为此,提出了一种基于双目相机估计深度图与激光雷达初始投影深度图的相机-激光雷达外参联合标定方法。该方法首先采用 SGBM 算法对双目图像进行立体匹配,获得高精度深度估计图;然后将其与激光雷达初始投影深度图一同输入设计的轻量级深度学习网络,进行多模态特征提取,有效缓解模态不一致问题;接着通过相关性匹配层对两种特征进行相关性计算,并且引入两组自注意力机制分别关注旋转外参和平移外参;最后,通过迭代细化的网络训练策略实现高精度的外参估计。在 KITTI Odometry 数据集上的实验结果表明,所提算法可以分别取得 0.67 cm 的平均平移误差和 0.09° 平均角度误差,较当前主流的方法 LCCNet 分别降低了 59.64% 和 72.73%,并且具有更少的模型参数数量。此外,所提算法在实车测试中也展现了精确的端到端标定效果,以所提算法标定结果为初始外参的 LVI-SAM 算法绝对轨迹均方根误差相较 LCCNet 降低了 5.18%,验证了该方法在标定准确性和工程实用性方面的优势。

关键词: 激光雷达; 相机; 深度学习; 端到端; 外参标定

中图分类号: TH74 **文献标识码:** A **国家标准学科分类代码:** 460.40

End-to-end camera-LiDAR extrinsic calibration method based on stereo camera depth estimation

Liu Qiuhua, Xu Xiaosu

(School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China)

Abstract: Accurate and reliable extrinsic calibration of sensors is essential for achieving high-precision localization and navigation in camera-LiDAR fusion systems. However, existing end-to-end camera-LiDAR calibration methods suffer from various limitations, such as large model parameter sizes and mismatched cross-modal feature correlation computation. To address these issues, this article proposes a novel joint calibration method based on stereo camera-estimated depth maps and initial LiDAR-projected depth maps. Specifically, the SGBM algorithm is used to perform stereo matching and generate high-accuracy depth estimation maps. These maps, along with the initial LiDAR depth projections, are fed into a lightweight deep neural network designed for multi-modal feature fusion, effectively mitigating modality inconsistency. A correlation matching layer is then utilized to compute feature-level correspondences, and two separate self-attention mechanisms are introduced to independently model rotational and translational extrinsic. Finally, an iterative refinement training strategy is adopted to enhance calibration accuracy. Compared with the state-of-the-art method LCCNet, experimental results on the KITTI Odometry dataset show that the proposed method achieves an average translation error of 0.67 cm and an average rotation error of 0.09°, representing reductions of 59.64% and 72.73%, respectively. And it requires fewer model parameters. In addition, real-world vehicle tests further demonstrate the effectiveness of the proposed method. When used as the initial extrinsic calibration in the LVI-SAM system, the absolute trajectory root mean square error is reduced by 5.18% compared with LCCNet, validating the accuracy and practical applicability of the method.

Keywords: LiDAR; camera; deep learning; end-to-end; extrinsic calibration

0 引言

多传感器融合技术在移动机器人自主导航领域应用广泛^[1]。该领域中相机和激光雷达是两种最常见的传感器。其中,相机可以在透视视角下提供丰富的颜色和纹理信息,激光雷达可以提供物体的详细三维信息,通过将二者之间的感知数据融合,可以获得更丰富和精准的环境信息^[2]。相机与激光雷达的外参标定^[3]是实现二者有效融合的先决条件,也是同时定位与建图(simultaneous localization and mapping, SLAM)技术^[4]和机器人感知技术^[5]的基础。

传统的相机-激光雷达外参标定借助于人工标定,大致分为基于目标标定的方法和无目标标定方法。Geiger等^[6]通过在墙壁和地板上放置多个棋盘图案来实现外参标定,但该方法无法得到更精确的相机和激光雷达对应关系。为了解决该问题,一些方法^[7]引入具有特殊外观的标注板,例如多边形^[8]、空心圆^[9]和球形^[10]等。多边形或球形的表面轮廓可以分别在三维 LiDAR (light detection and ranging) 点云和二维图像数据上精确检测到,从而实现快速、稳定的外参标定。Ying等^[11]提出了一种不需要目标的标定方法,利用运动和边缘匹配两阶段校准得到较精确的外参。吴军等^[12]利用无标定板的背景点云与有标定板的目标点云间空间域的密度差异性,自适应获取两类点云间的差异系数,聚类完成三维点云提取,提高了激光雷达和相机外参标定的准确性。基于人工标定的方法能取得较高的标定精度,缺点是需要人为选取特征或者使用特殊的标记物,过度依赖人为操作,标定过程较为繁琐^[13]。

目前,深度学习技术的飞速发展,促进了相机和激光雷达外参标定的端到端实现。2017年 Schneider等^[14]首次将神经网络引入相机和激光雷达外参标定任务,将校准的3个常规步骤(特征提取、特征匹配和全局回归)转换为只需通过卷积神经网络(convolution neural network, CNN)一步实现。但是该方法没有利用 LiDAR 深度图和 RGB(red green blue)图像的相关性,只是将其看作一个黑盒问题,因此取得的标定精度不高。2019年 Cattaneo等^[15]在上述方法的基础上引入了相关性计算方法,利用深度图和 RGB 图的对应关系,进一步的提升了外参标定的精度。2021年 Lyu等^[16]在此基础上提出了一种快速的迭代雷达-相机校准算法(LiDAR-camera calibration network, LCCNet),引入点云距离损失促进模型的训练,实现对激光雷达-相机外参的实时预测,取得了较好的标定效果。2022年 Jing等^[17]提出了一种名为 DXQ-Net 的新型端到端外部校准方法,利用了卷积神经网络强大的特征提取能力,并且通过可微分的姿态估计模块来实现泛化。

针对上述问题,本研究提出一种新颖的利用双目相机估计的深度图与激光雷达深度图进行相机外参标定的方法。该方法通过将激光雷达和相机数据统一至深度图模式,缓解了传统 RGB 图像与 LiDAR 点云在模态特性、纹理表现等方面存在的间隔差异。在深度图中,双目相机与激光雷达所表达的几何结构具有一致的深度纹理特征,对其进行图像相关性计算是合理且高效的。该方法有助于提升外参标定的准确性与鲁棒性,降低跨模态特征匹配所带来的难度与不确定性。此外,引入双自注意力机制,使模型分别关注旋转外参矩阵和平移外参矩阵的特征信息,有效抑制不同类型误差间的耦合干扰,提高外参回归任务中的精度与收敛效果。

1 端到端相机-激光雷达外参标定模型概述

图1展示了目前端到端相机-激光雷达外参标定方法的对比图,所提算法主要利用双目相机估计的深度图和 LiDAR 深度图作为输入,相较于相机 RGB 图像的三通道数据,深度图只有单通道数据,具有更低的计算复杂度。

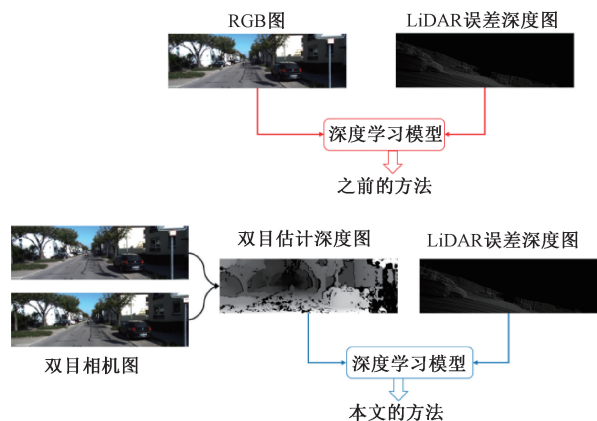


图1 端到端相机-LiDAR 外参标定方法对比

Fig. 1 End-to-end camera-LiDAR external parameter calibration comparison plot

1.1 LiDAR 深度图的生成

给定一组相机内参 $\mathbf{K}$ 和初始外参 $\mathbf{T}_{init}$, 可以将 LiDAR 扫描得到的每个三维点云 $P_i = [X_i, Y_i, Z_i] \in \mathbb{R}^3$ 投影至二维像素空间 $p_i = [u_i, v_i] \in \mathbb{R}^2$ 上, 得到深度图像。投影过程公式如式(1)所示。

$$\begin{cases} Z_i^{init} \cdot \hat{p}_i = Z_i^{init} [u_i, v_i, 1]^T = \mathbf{K} [\mathbf{R}_{init} | \mathbf{t}_{init}] \hat{P}_i = \\ \mathbf{K} [\mathbf{R}_{init} | \mathbf{t}_{init}] [X_i, Y_i, Z_i, 1]^T \\ \mathbf{T}_{init} = \begin{bmatrix} \mathbf{R}_{init} & \mathbf{t}_{init} \\ 0 & 1 \end{bmatrix} \end{cases} \quad (1)$$

其中, $\hat{P}_i$ 和 $\hat{p}_i$ 分别表示 P_i 和 p_i 的齐次坐标, $\mathbf{R}_{init}$ 和

t_{init} 分别表示 T_{init} 的旋转矩阵和平移向量。然后通过使 Z 轴压缩的方法,得到每个像素点 (u_i, v_i) 对应的 3D 点 P_i 上的深度值 Z_i^{init} ,最终得到初始深度图像 $D_{init} \in \mathbb{R}^2$ 。

但是实际应用中,并不是所有投影的点都是有效的,因为初始外参可能存在误差,导致一部分投影点无法准确投影到有效的 RGB 图像预期位置内,因此需要根据实际图像大小去除一部分投影点。具体地说,所有满足 $0 < u_i < W$ 和 $0 < v_i < H$ 的每个投影 2D 点 (u_i, v_i) 都认为有效,其中 W 和 H 分别表示图像的宽度和长度。

1.2 SGBM 深度估计原理

基于半全局立体匹配算法 (semi-global block matching, SGBM) 双目相机深度估计原理如图 2 所示。

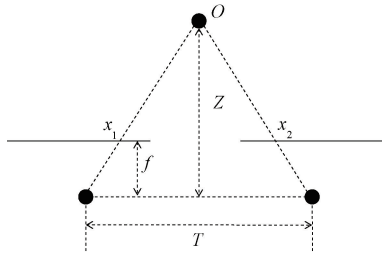


图 2 双目相机深度估计原理

Fig. 2 Depth estimation schematic of the binocular camera

将两个相机平行放置且保证二者焦距相等,假设 O 是三维空间中待测物体的一点,其在左右相机的投影点分别为 O_1 和 O_2 ,且其横坐标分别为 x_1 和 x_2 ,视差为 $d = x_1 - x_2$,则根据相似三角形可计算深度 Z ,计算公式如式(2)所示。

$$\begin{cases} \frac{T - (x_1 - x_2)}{Z - f} = \frac{T}{Z} \\ Z = \frac{fT}{x_1 - x_2} \end{cases} \quad (2)$$

由式(2)可知, f 和 T 都已知,因此双目相机深度估计的关键在于得到视差,这就需要立体匹配算法。立体匹配算法主要包括:全局立体匹配、局部立体匹配和半全局立体匹配算法。其中,全局立体匹配算法的精度最高但是速度慢,局部立体匹配算法速度快但是精度低。因此所提算法采用速度和精度权衡的半全局立体匹配算法,其主要的处理过程为:

1) 预处理

利用 Sobel 算子预处理左右相机图像,具体的公式如式(3)所示。

$$\text{Sobel}(u, v) = 2[P(u+1, v) - P(u-1, v)] + P(u+1, v-1) - P(u-1, v-1) + P(u+1, v+1) - P(u-1, v+1) \quad (3)$$

通过映射公式如式(4)所示生成新图像。

$$P_{new} = \begin{cases} 0, & P < -\delta \\ P + \delta, & -\delta \leq P \leq \delta \\ 2 \times \delta, & P > \delta \end{cases} \quad (4)$$

其中, P 和 P_{new} 分别表示原像素值和新像素值; δ 为一个常数。通过以上方式可以得到图像的梯度信息。

2) 匹配计算

利用函数 $d(u_i, v_i, I_L, I_R)$ 计算左右两幅图像同一像点的可能性大小,具体公式如式(5)所示。

$$d(u_i, v_i, I_L, I_R) = \min_{v_i - \frac{1}{2} \leq v \leq v_i + \frac{1}{2}} |I_L(u_i) - I_R(v)| \quad (5)$$

其中, $I_L(u_i)$ 和 $I_R(v_i)$ 分别表示左扫描线上点的灰度和右扫描线上点经过上采样后的灰度。

3) 路径设置与约束匹配计算

以 P 点为起点,在其周围均匀选取 8 个路径(以 45° 为间隔),并计算最小代价路径 $L_i(p, d)$ 。

4) 计算视差

视差值的计算是通过迭代过程来实现的。具体而言,该过程首先利用随机生成的视差图对右侧图像进行校正,并随后执行匹配步骤,以生成新的视差图。此外,还可以采用一种分级计算的方法,即首先在半分辨率级别上计算出初始的视差图,然后以此为基础,通过递归的方式进一步细化视差图的计算。

5) 剔除误匹配

在完成左右影像中同名点的匹配之后,会利用右影像中的已匹配点去反向匹配左影像中的对应同名点。若这两次匹配过程所得出的视差值不一致,则该匹配对被视为无效匹配。

2 端到端外参标定模型详细结构

第 1 章给出了端到端标定模型的主要输入,该部分将主要介绍深度学习模型的详细结构,主要包括模型整体结构概述、特征提取网络、特征增强网络和全连接预测网络。

2.1 深度学习模型整体结构概述

图 3 展示了相机-LiDAR 端到端标定模型的整体结构图。深度学习模型的输入由两部分组成:LiDAR 点云通过相机内参 K 和初始外参 T_{init} 投影的深度图和双目相机估计的深度图。二者在同一模态下进行融合,为后续特征匹配提供了一致性的几何表达。

模型首先通过特征提取网络对输入深度图进行特征编码,随后利用特征增强网络完成相关性计算,并通过自注意力机制对旋转和平移外参特征进行进一步增强。接着利用全连接预测网络可以回归出预测的外参矩阵 T_{pred} ,最后利用回归得到的外参矩阵对 LiDAR 误差深度图进行迭代修正,实现逐步优化的标定效果。

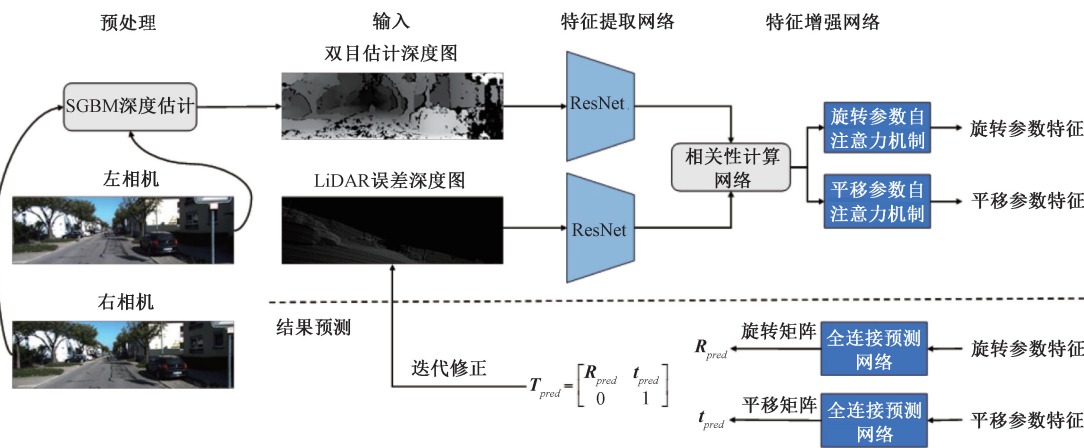


图 3 深度学习模型整体结构

Fig. 3 The overall architecture of the deep learning model

为了迭代优化外参矩阵,在平移扰动为 $[-1.5\text{ m}, +1.5\text{ m}]$ 和角度扰动为 $[-20^\circ, +20^\circ]$ 范围内,按照扰动从小到大的方式挑选了 5 组扰动作为训练条件,表 1 为具体的扰动输入参数。

表 1 5 组旋转和平移扰动

Table 1 Five sets of rotational and translational perturbation		
组号	旋转扰动	平移扰动
1	$[-20^\circ, +20^\circ]$	$[-1.5\text{ m}, +1.5\text{ m}]$
2	$[-10^\circ, +10^\circ]$	$[-1.0\text{ m}, +1.0\text{ m}]$
3	$[-5^\circ, +5^\circ]$	$[-0.5\text{ m}, +0.5\text{ m}]$
4	$[-2^\circ, +2^\circ]$	$[-0.2\text{ m}, +0.2\text{ m}]$
5	$[-1^\circ, +1^\circ]$	$[-0.1\text{ m}, +0.1\text{ m}]$

具体地说,将第 1 次预测结果 T_{pred} 记为 T_0 ,然后根据 $T_0^{-1} \cdot T_{init}$ 得到新的外参矩阵并且重新投影 LiDAR 点云,得到修正后的 LiDAR 误差深度图,继续送入网络后得到 T_1 。上述过程迭代 5 次后,得到最终的外参校准矩阵 T_{LC} ,具体公式如式(6)所示。

$$T_{LC} = (T_0 \cdot T_1 \cdot T_2 \cdot T_3 \cdot T_4 \cdot T_5)^{-1} \cdot T_{init} \tag{6}$$

2.2 特征提取网络

在特征提取模块中,选用计算复杂度较低的 ResNet-18 作为骨干网络,对 LiDAR 误差深度图与双目相机估计的深度图进行特征提取。相比于常用的 ResNet-34、ResNet-50 等大型网络结构,ResNet-18 在保持足够特征表达能力的同时,具备更高的计算效率,能够有效保障端到端外参标定任务的实时性,表 2 给出了具体的网络结构。

表 2 ResNet-18 的详细结构

Table 2 Detailed architecture of ResNet-18		
层名称	输出尺寸	结构
conv1	112×112	7×7, 64, stride 2
		3×3 max pool, stride 2
conv2_x	56×56	$\begin{bmatrix} 3\times 3, 64 \\ 3\times 3, 64 \end{bmatrix} \times 2$
conv3_x	28×28	$\begin{bmatrix} 3\times 3, 128 \\ 3\times 3, 128 \end{bmatrix} \times 2$
conv4_x	14×14	$\begin{bmatrix} 3\times 3, 256 \\ 3\times 3, 256 \end{bmatrix} \times 2$
conv5_x	7×7	$\begin{bmatrix} 3\times 3, 512 \\ 3\times 3, 512 \end{bmatrix} \times 2$

此外,要先对提取到的两种深度图原始特征进行相关性计算,而池化后的特征信息过少不利于相关性计算。因此,对 ResNet-18 的原始结构进行优化,去除了 conv5_x 层输出后的全局池化与全连接操作,保留高分辨率的特征图,避免信息在下采样过程中过度损失,便于后续双深度图之间的细粒度相关性计算。这一改动使得模型在保持轻量化的同时,兼顾了匹配精度与表达能力。

另一方面,采用双目深度图与激光雷达深度图作为输入,摒弃了传统方法中 RGB-深度混合输入的方式,从数据源层面实现了模态一致性,进一步降低了模型的参数规模与计算负担。该设计不仅提升了网络在资源受限环境下的适用性,也从根本上缓解了跨模态特征对齐的难题。

2.3 特征增强网络

为了进一步提升特征表达能力,在特征提取网络的

基础上,设计了特征增强网络,对双目深度图与 LiDAR 深度图之间的匹配信息进行精细建模。针对原始特征在空间对齐与姿态感知方面存在信息量不足,难以有效利用深度图之间的几何关系,导致外参回归精度受限的问题,所提算法结合端到端标定任务的特点,引入了相关性计算模块和自注意力机制,针对旋转和平移外参特征进行增强,提高了模型的适应性与泛化性。

正如上文所述,现有的端到端外参标定方法通常采用相关性计算网络来评估相机 RGB 图像与 LiDAR 深度图之间的匹配度。然而该方法存在着模态间隔和纹理间隔,导致匹配结果的精度受影响。为解决这一问题,提出了一种新的方法,在同一模态和纹理(即深度图)下进行匹配度计算。该方法能够更加便捷和高效率地评估误差深度图与真实深度图之间的匹配度,从而提高了匹配计算的速度与精度,图 4 给出了相关性计算网络示意图。



图 4 相关性计算网络示意图

Fig. 4 Diagram of the correlation calculation network

相关性计算网络本质上是一次特殊的卷积运算,使用双目相机估计深度图特征对 LiDAR 误差深度图特征做卷积处理。此外,该卷积操作不需要训练参数,简化了训练流程。通过该方式可以获得二者的相关度 $c(x_1, x_2)$, 具体的计算如式(7)所示。

$$\begin{cases} c(x_1, x_2) = \sum \langle L_1(x_1 + o), L_2(x_2 + o) \rangle \\ O \in [-k, k] \times [-k, k] \end{cases} \quad (7)$$

其中, L_1 和 L_2 分别表示双目相机估计深度图和 LiDAR 误差深度图; x_1 和 x_2 分别表示 L_1 和 L_2 上的像素点; O 表示相关性计算的延展区域,分别以 x_1 和 x_2 点为基准,上下延展 k 个像素,因此长度为 $2k + 1$ 。

在相机-LiDAR 外参标定任务中,除了特征匹配外,如何从相关性矩阵中有效提取分别有利于平移与旋转参数回归的特征信息也十分重要。传统方法多依赖固定的特征选择策略,无法根据任务需求动态调整关注重点,容易导致特征冗余与信息干扰,影响外参回归的准确性。受自注意力机制的启发,提出基于双自注意力机制的特征增强方法,实现对平移与旋转参数特征的解耦建模与有针对性的特征提取。

自注意力机制能够在模型中实现对重要特征的动态聚焦,使其能够自适应地识别和加权影响最终回归结果

的特征,避免了传统方法中的静态特征选择问题。图 5 给出了自注意力机制的示意图。

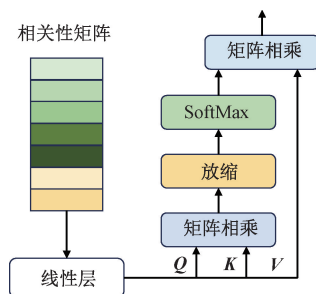


图 5 自注意力机制示意图

Fig. 5 Diagram of the self-attention mechanism

每组自注意力机制通过独立的线性变换层,对相关性矩阵进行不同的投影和加权,分别为平移和旋转回归任务提供特征信息,提升了标定任务中参数估计的精度。自注意力的计算公式如式(8)所示。

$$Attention(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right) V \quad (8)$$

其中, Q 、 K 和 V 由相关性矩阵经过线性层得到; d 为放缩系数。引入两组自注意力机制,可以分别得到对平移参数和旋转参数矩阵有用的特征 M_r 和 M_l 。

2.4 全连接预测网络

在经过特征提取网络和特征增强网络后,该部分主要利用全连接预测网络对旋转矩阵参数 M_r 和平移矩阵参数 M_l 进行旋转和平移矩阵预测,得到预测的旋转参数 R_{pred} 和平移参数 t_{pred} , 图 6 展示了具体的全连接预测网络结构图。搭建多组全连接神经元,采用层层递进的回归方法对平移和旋转矩阵进行精细化的回归。此外,考虑到深层神经网络在标定任务中易出现过拟合的问题,在预测网络中引入了 Dropout 正则化策略,通过在训练过程中随机屏蔽部分隐藏层神经元的激活值,打破特征之间的依赖性,有效抑制了模型在训练阶段的过拟合风险,使模型更加稳定和鲁棒。

3 实验验证分析

为了验证模型的有效性,首先在 KITTI Odometry 数据集上进行了实验,验证该方法在外参标定方面的性能。该数据集共有 22 个序列,涵盖了市区、农村和开阔道路等多种常见场景,同时还存在一定程度的遮挡,适用于评估基于深度学习的相机-LiDAR 外参标定方法。在本章中,将详细介绍实验设置参数、评估指标和实验结果。

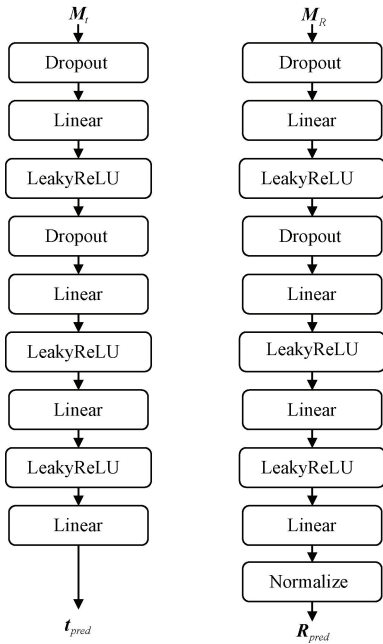


图 6 全连接预测网络结构

Fig. 6 Diagram of the architecture of a fully connected prediction network

3.1 实验参数设置

为了保证实验前后的一致性,所有实验都是在一台 Ubuntu 18.04 操作系统,配备 NVIDIA GeForce RTX 3060 显卡的电脑上进行。采用 Pytorch 深度环境进行模型搭建, batchsize 大小为 16, 初始学习率为 0.000 1, 训练 120 轮, 并且采用 Adam 优化器。

使用 KITTI Odometry 数据集来验证提出的端到端相机-LiDAR 标定算法。具体来说,使用序列 01 到 20 进行训练和验证(39 011 帧),序列 00 进行测试(4 541 帧)。

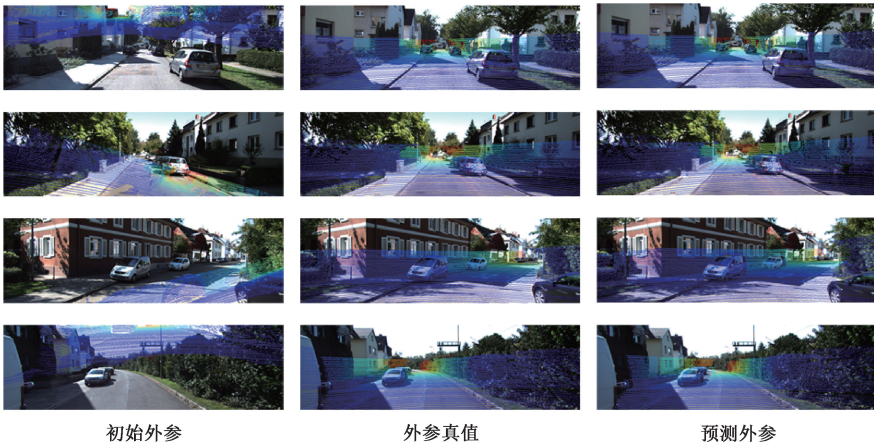


图 7 提出的外参标定方法的可视化结果

Fig. 7 Visualization results of the proposed external parameter calibration method

3.2 评估指标

对于相机-LiDAR 外参的旋转和平移参数误差评价标准分别采用两种方式。平移外参矩阵的误差可以通过计算预测的平移外参矩阵 t_{pred} 和真实平移外参矩阵 t_{gt} 之间的欧几里得距离来计算,公式如式(9)所示。

$$E_t = |t_{pred} - t_{gt}| \quad (9)$$

按照以上方式,分别测试了沿 X、Y、Z 轴方向的绝对平移误差,可以得到误差平均值 $\bar{t} = (E_x + E_y + E_z)/3$ 。

对于旋转矩阵的评估,首先将旋转矩阵变换为欧拉角,然后根据 E_{Roll} 、 E_{Pitch} 和 E_{Yaw} 计算角度误差,其误差平均值为 $\bar{R} = (E_{Roll} + E_{Pitch} + E_{Yaw})/3$, 计算公式如式(10)所示。

$$\begin{cases} R_{pred}^{-1} \cdot R_{gt} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \\ E_{Yaw} = \theta_z = \text{atan2}(r_{21}, r_{11}) \\ E_{Pitch} = \theta_y = \text{atan2}(-r_{33}, \sqrt{r_{31}^2 + r_{33}^2}) \\ E_{Roll} = \theta_x = \text{atan2}(r_{32}, r_{33}) \end{cases} \quad (10)$$

3.3 实验结果展示

图 7 展示了端到端外参标定方法的可视化结果图,图 7 中的预测结果是经过 5 次迭代后的误差标定结果。在给定不同扰动的初始外参的情况下,端到端外参标定方法可以相对正确的回归出准确的相机-LiDAR 外参,并且仅利用两种深度图的方式具有更少的参数量,方便于模型部署。

表 3 给出了所提算法在 KITTI Odometry 数据集序列 0 上的详细测试结果。注意到,所提算法在扰动为 $(\pm 20^\circ, \pm 1.5 \text{ m})$ 的情况下,经过 5 次迭代后,平均平移和角度误差分别为:0.67 cm 和 0.09°。

表 3 本研究算法在 KITTI Odometry 数据集上的实验结果

Table 3 Experimental results of the proposed method on the KITTI Odometry dataset

输入扰动	误差	平移/cm				旋转/(°)			
		E_t	X	Y	Z	E_R	Roll	Pitch	Yaw
$(\pm 20^\circ, \pm 1.5\text{ m})$	平均值	19.67	4.61	7.08	21.74	1.15	0.44	0.34	0.33
	中值	15.22	2.45	8.80	24.92	0.88	0.51	0.30	0.31
	标准差	17.75	4.79	4.12	9.77	1.19	0.21	0.27	0.19
$(\pm 10^\circ, \pm 1.0\text{ m})$	平均值	4.86	2.23	2.06	2.16	0.24	0.01	0.04	0.10
	中值	4.29	1.46	2.08	2.23	0.18	0.01	0.01	0.10
	标准差	4.34	1.92	1.83	1.67	0.64	0.02	0.07	0.08
$(\pm 5^\circ, \pm 0.5\text{ m})$	平均值	2.13	1.16	1.11	1.94	0.19	0.07	0.09	0.08
	中值	1.91	1.52	1.12	2.03	0.15	0.06	0.07	0.10
	标准差	2.83	0.69	0.12	0.70	0.60	0.01	0.06	0.06
$(\pm 2^\circ, \pm 0.2\text{ m})$	平均值	0.99	0.27	0.19	0.57	0.12	0.01	0.03	0.03
	中值	0.82	0.11	0.19	0.76	0.08	0.01	0.01	0.04
	标准差	2.55	0.29	0.17	0.41	0.60	0.01	0.04	0.01
$(\pm 1^\circ, \pm 0.1\text{ m})$	平均值	0.67	0.23	0.26	0.07	0.09	0.03	0.01	0.04
	中值	0.51	0.14	0.30	0.09	0.06	0.03	0.02	0.03
	标准差	2.52	0.19	0.19	0.07	0.59	0.02	0.01	0.03

图 8 展示了在初始扰动为 $(\pm 20^\circ, \pm 1.5\text{ m})$ 的情况下,经过 5 次迭代后的平移误差与旋转误差平均值变化曲线,左纵轴为平移误差,右纵轴为旋转误差,图 8 中圆点实线代表平移误差 E_t ,方框虚线代表旋转误差 E_R 。由图 8 可以看出,尽管在大范围扰动下,初始外参误差较大,但所提算法仍能在 5 次迭代内实现较为鲁棒的标定效果。

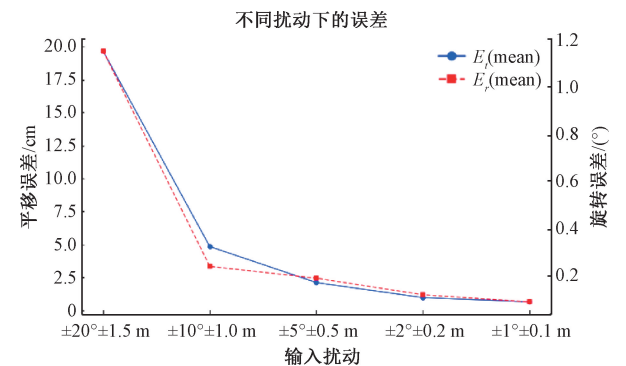


图 8 端到端外参标定方法误差收敛性曲线

Fig. 8 Error convergence curves of end-to-end extrinsic parameters calibration method

为了直观地体现所提算法的先进性,图 9 展示了提出的方法与 CMRNet、LCCNet 和 DXQ-Net 的对比优势。

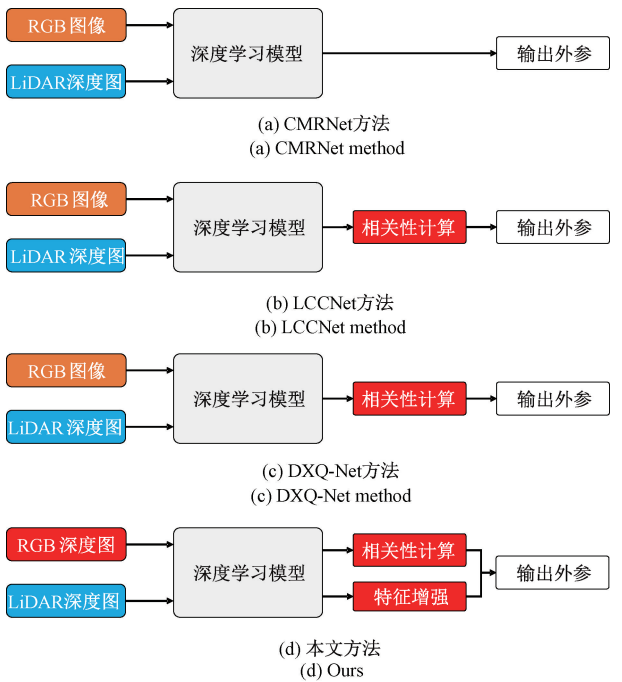


图 9 不同模型方法对比

Fig. 9 Comparison chart of different modeling methods

其中,CMRNet 作为激光雷达-相机在线标定的开山之作,未考虑相关性计算部分,将标定视为黑盒问题,缺乏更精细化的处理。尽管 LCCNet 和 DXQ-Net 引入了相

关性计算,但它们采用 RGB 图像与雷达深度图作为输入,存在模态不匹配问题,且未对后级特征进行增强,导致拟合能力不足。为了更全面地验证所提算法的优势,表 4 给出了各方法在 KITTI Odometry 数据集序列 0 上的测试结果对比。

表 4 KITTI Odometry 序列 0 下不同方法误差及参数量分析

Table 4 Analysis of errors and parameter sizes of different methods under KITTI Odometry sequence 0

方法	E_t/cm	$E_R/(^\circ)$	参数量/M
CMRNet	2.38	0.59	11.11
LCCNet	1.66	0.33	57.11
本文算法	0.67	0.09	23.68

由于 DXQ-Net 未开源代码,未对其进行对比实验。为确保实验公平性,所有测试均在相同机器上进行。CMRNet 取得了最少的参数量(11.11 M),但其平移和旋转误差最大,这是由于 CMRNet 未采用相关性计算模块,仅依靠深度学习的黑盒属性无法精确检测激光雷达点云特征与相机像素特征的对应关系。LCCNet 通过加入相关性计算模块,在性能上优于 CMRNet,但仍弱于所提算法。这是由于直接采用 RGB 图像增加了模型输入负担(RGB 图像为 3 通道,而深度图仅为 1 通道),同时模态不匹配问题和缺乏特征增强方法也是 LCCNet 效果一般的原因。具体而言,在相同扰动条件($\pm 20^\circ, \pm 1.5\text{ m}$)下,LCCNet 的平移误差和旋转误差分别为 1.66 cm 和 0.33 $^\circ$,分别高出所提算法 0.99 cm 和 0.24 $^\circ$ 。由此可见,通过计算双目相机估计深度图与 LiDAR 误差深度图的相关性,能够有效提升模型的回归准确性。此外,LCCNet 的参数量为 57.11 M,而所提算法的参数量仅为 23.68 M,不到 LCCNet 的一半。这得益于仅利用深度图数据,相较于 LCCNet 采用的相机 RGB 图像,具有更少的模型参数量。总的来说,本研究提出的相机-LiDAR 端到端外参标定方法在性能和参数量方面均表现出显著优势。接下来将进一步验证所提出方法各模块的有效性。

3.4 特征提取网络消融实验

为验证所提出的特征提取模块在保持足够特征表达能力的同时,具备更高的计算效率与良好的任务适应性,设计消融实验,分析优化后的 ResNet-18 网络在外参标定任务中的性能表现。表 5 给出了不同网络结构在特征提取阶段的消融实验结果。

从表 5 可以看出,优化后的 ResNet-18 在外参标定精度上与标准 ResNet-18 保持了接近的水平,同时在推理时间与参数量方面实现了显著优化,前向推理耗时降低约 30.79%,参数量减少 13.28%,在保障特征表达能力的前

表 5 特征提取网络消融实验结果对比

Table 5 Comparison of ablation experiment results of feature extraction networks

方法	E_t/cm	$E_R/(^\circ)$	推理时间/ms	参数量/M
原 ResNet-18	0.69	0.12	35.66	13.48
改进 ResNet-18	0.67	0.09	23.68	11.69
ResNet-50	0.62	0.07	77.51	25.51
MobileNet	2.13	0.41	9.42	4.37

提下提升了计算效率,验证了高分辨率特征提取策略的有效性;ResNet-50 虽在精度方面略有提升,但计算开销较大,推理时间增加至 77.51 ms。而 MobileNet 尽管具备更小的模型规模与更快的推理速度,但由于特征表达能力不足,导致标定精度显著下降。综上所述,提出的优化版 ResNet-18 在精度与效率之间取得了良好平衡,验证了其在端到端外参标定任务中的适用性与优越性。

3.5 特征增强网络消融实验

为全面验证特征增强网络中相关性计算模块与自注意力机制在外参回归任务中的性能贡献,设计了针对两者的消融实验,对各模块的独立作用及其组合效果进行了系统性测试。表 6 给出了具体的测试结果。

表 6 特征增强网络消融实验结果对比

Table 6 Comparison of ablation experiment results of feature enhancement networks

方法	E_t/cm	$E_R/(^\circ)$	参数量/M
自注意力机制	1.24	0.23	23.66
相关性计算	1.11	0.19	22.58
本研究算法	0.67	0.09	23.68

从表 6 可以看出,在自注意力机制单独引入时,模型的参数量为 23.66 M,平移误差为 1.24 cm,旋转误差为 0.23 $^\circ$,表明尽管自注意力机制能够有效捕捉全局特征之间的长距离依赖关系,提升了特征表征能力,但由于其对大规模特征图进行加权聚合的计算特性,带来了相对较高的计算复杂度与模型参数量。相关性计算模块单独使用时,参数量为 22.58 M,平移误差为 1.11 cm,旋转误差为 0.19 $^\circ$,该模块主要用于计算不同特征之间的相关性,能够有效提升模型的局部特征提取能力。所提算法结合了自注意力机制和相关性计算模块,参数量为 23.68 M,平移误差为 0.67 cm,旋转误差为 0.09 $^\circ$,与单独使用自注意力机制或相关性计算模块相比,在平移误差和旋转误差上均取得了显著提升。这表明自注意力机制和相关性计算模块的结合能够有效互补,进一步提升模型的整体性能。

尽管所提算法在性能上取得了显著提升,但参数量仍然较高,未来可以考虑进一步优化模型结构,减少参数量,同时保持或提升性能。此外,还可以探索其他模块或技术的引入,以进一步提升模型的泛化能力和鲁棒性。

4 实车验证分析

4.1 实验设置与评估指标

为了更加全面的验证本研究算法,在实车平台上进行了验证,图 10 展示了实验平台车辆图片。该无人车实验平台搭载了 Intel RealSense D435 型号相机获取图像信息,RS-Helios-16P 型号激光雷达获取点云信息。无人车底盘为 SCOUT MINI,选择 8 核 ARMv8.2 64 位处理器作为无人车传感器数据处理平台,测试环境在东南大学无锡国际校区两江院室外。

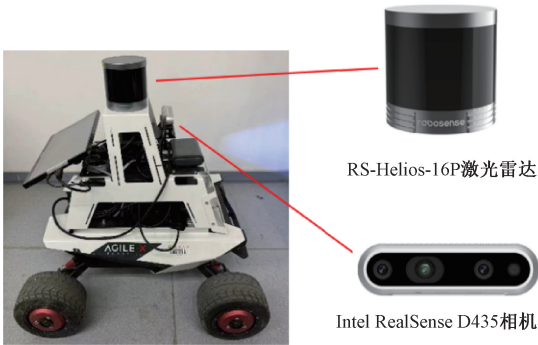


图 10 实车测试平台
Fig. 10 Real vehicle test bench

为了模拟平移和角度误差扰动,人为的给初始外参加入扰动:($\pm 20^\circ, \pm 1.5\text{ m}$)。为了细化实测效果评价指标,对实测数据进行下采样,随机选取了其中的 300 帧数据进行了测试,并且采用与 3.2 节一致的评价指标。

4.2 实车外参标定实验

图 11 展示了实车实验结果,可以看出即使初始外参差距较大的情况下,端到端外参标定方法仍然可以有效的矫正外参。

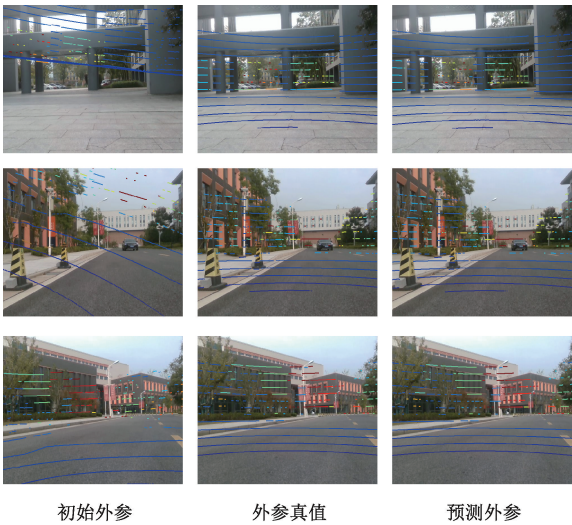


图 11 实车测试效果
Fig. 11 Diagram of real vehicle test effect

表 7 为具体的测试结果,在扰动为($\pm 20^\circ, \pm 1.5\text{ m}$)的情况下,经过 5 次迭代,所提算法在实车测试中取得

表 7 实车验证结果
Table 7 Results of the real-vehicle evaluation

输入扰动	误差	平移/cm				旋转/(°)			
		E_t	X	Y	Z	E_R	$Roll$	$Pitch$	Yaw
($\pm 20^\circ, \pm 1.5\text{ m}$)	平均值	38.46	12.07	32.16	8.86	1.77	0.91	1.38	0.66
	中值	32.39	6.22	36.39	9.95	0.76	0.74	0.86	0.59
	标准差	24.47	14.73	10.43	7.03	0.84	0.52	1.23	0.28
($\pm 10^\circ, \pm 1.0\text{ m}$)	平均值	10.40	3.52	3.26	5.35	0.55	0.35	1.19	0.22
	中值	9.35	2.12	2.65	2.69	0.45	0.28	1.03	0.18
	标准差	7.83	2.64	3.40	4.72	0.50	0.21	0.92	0.20
($\pm 5^\circ, \pm 0.5\text{ m}$)	平均值	4.15	2.09	4.65	2.75	0.31	0.23	0.24	0.10
	中值	3.71	2.29	6.13	2.75	0.25	0.12	0.11	0.10
	标准差	2.99	0.58	3.53	2.64	0.27	0.20	0.30	0.07
($\pm 2^\circ, \pm 0.2\text{ m}$)	平均值	2.31	0.85	1.34	1.04	0.19	0.11	0.10	0.08
	中值	2.01	0.18	1.12	1.17	0.17	0.16	0.10	0.11
	标准差	1.38	1.17	1.18	0.70	0.14	0.0852	0.08	0.06
($\pm 1^\circ, \pm 0.1\text{ m}$)	平均值	1.43	0.76	1.02	0.65	0.12	0.06	0.09	0.03
	中值	1.31	1.02	0.27	0.26	0.10	0.06	0.02	0.04
	标准差	0.99	0.67	1.34	0.78	0.09	0.01	0.13	0.02

1.43 cm 平移误差和 0.12° 旋转误差。对于实车测试而言,取得效果符合实际工程要求。

图 12 展示了在初始扰动为 ($\pm 20^\circ, \pm 1.5\text{ m}$) 的情况下,经过 5 次迭代后的平移误差与旋转误差平均值变化曲线,左纵轴为平移误差,右纵轴为旋转误差,图 12 中圆点实线代表平移误差 E_t , 方框虚线代表旋转误差 E_R 。

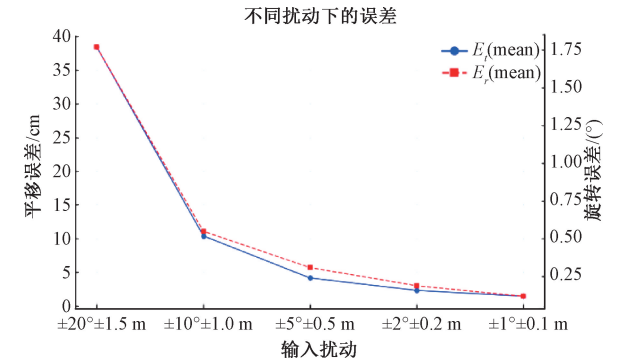


图 12 端到端外参标定方法误差收敛性曲线
Fig. 12 Error convergence curves of end-to-end extrinsic parameters calibration method

由图 12 可以看出,即使在较大扰动下,所提算法依然能够在有限次迭代内实现有效收敛,最终将外参误差控制在较低水平,验证了所提算法的鲁棒性与标定精度。

接下来在相同的机器上测试了主流的 LCCNet 方法和 CMRNet 方法,表 8 给出了各算法具体的对比结果。可以看出,在同样的扰动条件 ($\pm 20^\circ, \pm 1.5\text{ m}$) 下, LCCNet 方法和 CMRNet 方法的平移误差和旋转误差均高出所提算法。

表 8 实车验证结果

Table 8 Results of the real-vehicle evaluation

方法	输入扰动	E_t/cm	$E_R/(\text{^\circ})$
本研究算法	($\pm 20^\circ, \pm 1.5\text{ m}$)	1.43	0.12
LCCNet	($\pm 20^\circ, \pm 1.5\text{ m}$)	2.58	0.34
CMRNet	($\pm 20^\circ, \pm 1.5\text{ m}$)	4.17	0.63

4.3 端到端外参标定方法有效性验证实验

为验证端到端外参标定方法的实用性和有效性,将所提算法标定得到的激光雷达与相机间的外参与 LCCNet 方法、CMRNet 方法得到的外参,分别作为紧耦合激光雷达/视觉/惯性系统 LVI-SAM^[18] 的初始外参,并关闭 LVI-SAM 中的外参在线估计模块,定量分析算法定位精度^[19]。为了获得高精度地面真值,实验采用激光投线仪与激光测距仪结合锚点法进行真值标定。引入绝对轨迹误差 (absolute trajectory error, ATE) 的均方根误差 (root mean square error, RMSE)、最大误差 (max)、平均误差

(mean) 作为定位精度评估的关键指标^[20]。RMSE 的计算公式如式 (11) 所示。

$$E_{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N \| \mathbf{T}_{gt,i}^{-1} \mathbf{T}_{esti,i} \|^2}$$

(11)

其中, $\mathbf{T}_{esti,i}$ 、 $\mathbf{T}_{gt,i}^{-1}$ 分别表示 i 时刻的估计位姿和真实位姿, $i = 1, \dots, N$, 表示关键帧的个数。

控制移动机器人沿着预设锚点前进,得到如图 13 所示的轨迹对比与三轴分量图。

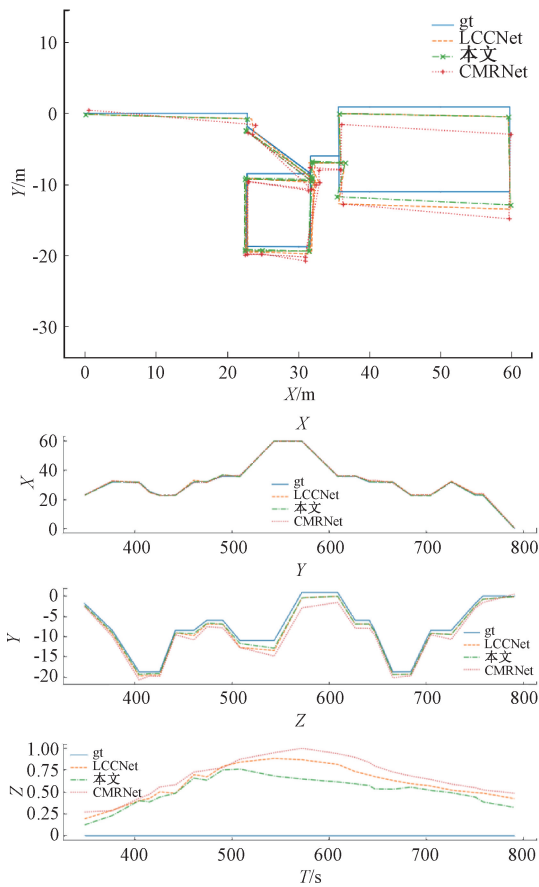


图 13 基于不同标定算法外参的 LVI-SAM 轨迹对比
Fig. 13 LVI-SAM trajectory comparison based on extrinsic parameters of different calibration algorithms

从图 13 中可以观察到,所提方法标定后的 LVI-SAM 输出轨迹与真值轨迹高度吻合,明显优于 LCCNet 与 CMRNet 两种方法。后两者由于外参精度偏差,导致轨迹在全局范围内存在一定偏移,尤其在转弯与快速移动场景下误差更加突出。

表 9 中进一步给出了 3 种算法输出外参的具体定位精度对比。可以看出,使用所提算法标定外参的 LVI-SAM 输出轨迹的均方根误差相比于 LCCNet 与 CMRNet 分别减少了 5.18% 和 8.53%,验证了所提外参标定算法的实用性和有效性。

表 9 算法绝对轨迹误差对比

Table 9 Comparison of absolute trajectory errors among algorithms (m)

对比项	CMRNet	LCCNet	本文
最大误差	1.034	0.927	0.819
平均误差	0.313	0.280	0.269
均方根误差	0.340	0.328	0.311

5 结 论

本研究提出了一种新颖的端到端相机-LiDAR 外参标定方法,主要用于解决模型参数量大和相关性计算不匹配的问题。不同于之前需要将相机图像输入模型的方法,所提算法利用双目相机估计的深度图和 LiDAR 初始深度图作为输入,解决了之前模态不匹配问题;然后进行自注意力机制增强对平移参数和旋转参数矩阵的特征提取,利用全连接预测网络得到预测的相机-LiDAR 外参,最后再进行 5 次迭代细化后得到最终的外参结果。分别在 KITTI Odometry 数据集和真实测试场景下进行了测试,证明了所提算法的有效性。并与当前主流的 LCCNet,CMRNet 方法进行对比,所提算法在平移误差和旋转误差均低于 LCCNet,CMRNet 方法。

参考文献

[1] WEI P, CAGLE L, REZA T, et al. LiDAR and camera detection fusion in a real-time industrial multi-sensor collision avoidance system [J]. Electronics, 2018, 7(6) :84-115.

[2] 顾彩凤. 不同传感器在机器人物体检测中的性能分析[J]. 电子技术,2024,53(4) :290-291.

GU C F. Performance analysis of different sensors in robot object detection[J]. Electronic Technology, 2024, 53(4) : 290-291.

[3] 王庆,谭镭轩,冯悠扬,等. 基于三维标定板的相机-激光雷达联合标定方法 [J]. 中国惯性技术学报, 2023,31(1) :100-106.

WANG Q, TAN R X, FENG Y Y, et al. Joint calibration method of camera and LiDAR based on 3D calibration plate[J]. Journal of Chinese Inertial Technology, 2023, 31(1) : 100-106.

[4] HAMADI A, LATOUI A. An accurate smartphone-based indoor pedestrian localization system using ORB-SLAM camera and PDR inertial sensors fusion approach [J]. Measurement, 2025, 240(2) :115642.

[5] GHAFFARI M, ZHANG R, ZHU M H, et al. Progress in symmetry preserving robot perception and control

through geometry and learning[J]. Frontiers in Robotics and AI, 2022, 9:969380.

[6] GEIGER A, MOOSMANN F, CAR Ö, et al. Automatic camera and range sensor calibration using a single shot[C]. 2012 IEEE International Conference on Robotics and Automation, 2012: 3936-3943.

[7] MISHRA S, PANDEY G, SARIPALLI S. Extrinsic calibration of a 3D-LiDAR and a camera[C]. 2020 IEEE Intelligent Vehicles Symposium, 2020: 1765-1770.

[8] PARK Y, YUN S, WON C S, et al. Calibration between color camera and 3D LiDAR instruments with a polygonal planar board[J]. Sensors, 2014,14(3) :5333-5353.

[9] GUINDEL C, BELTRÁN J, MARTÍN D, et al. Automatic extrinsic calibration for LiDAR-stereo vehicle sensor setups[C]. 2017 IEEE 20th International Conference on Intelligent Transportation Systems, 2017: 1-6.

[10] TÓTH T, PUSZTAI Z, HAJDER L. Automatic LiDAR-camera calibration of extrinsic parameters using a spherical target[C]. 2020 IEEE International Conference on Robotics and Automation, 2020: 8580-8586.

[11] YING Y, SONG F, LI L, et al. Automatic targetless monocular camera and LiDAR external parameter calibration method for mobile robots [J]. Remote Sensing, 2023, 15(23) : 5560.

[12] 吴军,袁少博,祝玉恒,等. 采用自适应背景聚类的激光雷达与相机外参标定优化方法 [J]. 仪器仪表学报,2023,44(2) :230-237.

WU J, YUAN SH B, ZHU Y H, et al. Optimization method for external parameters calibration of LiDAR and camera using adaptive background clustering [J]. Chinese Journal of Scientific Instrument, 2023, 44(2) : 230-237.

[13] 刘万里,刘扬,张学亮. 融合互信息和多特征约束的激光雷达与相机外参标定方法 [J]. 仪器仪表学报, 2025,46(1) :29-41.

LIU W L, LIU Y, ZHANG X L. Extrinsic calibration of LiDAR and camera through mutual information integrated multi-feature constraints[J]. Chinese Journal of Scientific Instrument, 2025,46(1) :29-41.

[14] SCHNEIDER N, PIEWAK F, STILLER C, et al. RegNet: Multimodal sensor registration using deep neural networks[C]. 2017 IEEE Intelligent Vehicles Symposium, 2017: 1803-1810.

[15] CATTANEO D, VAGHI M, BALLARDINI A L, et al. CMRNet: Camera to LiDAR-map registration[C]. 2019 IEEE Intelligent Transportation Systems Conference, 2019: 1283-1289.

[16] LYU X D, WANG B Y, DOU Z W, et al. LCCNet: LiDAR and camera self-calibration using cost volume network [C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2021:2888-2895.

[17] JING X, DING X Q, XIONG R, et al. DXQ-Net: Differentiable LiDAR-camera extrinsic calibration using quality-aware flow [C]. 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2022: 4022-4028.

[18] SHAN T X, ENGLOT B, RATTI C, et al. LVI-SAM: Tightly-coupled LiDAR-visual-inertial odometry via smoothing and mapping [C]. 2021 IEEE International Conference on Robotics and Automation, 2021: 5692-5698.

[19] 任家卫,徐晓苏. 基于平面特征的地面机器人雷达-惯性里程计外参标定方法 [J]. 仪器仪表学报,2025, 46(2):344-354.

REN J W, XU X S. Ground robot LiDAR-inertial odometry calibration based on plane constraint [J]. Chinese Journal of Scientific Instrument, 2025, 46(2): 344-354.

[20] 诸葛晶昌,高宏,罗其俊,等. 激光雷达 IMU 紧耦合 SLAM 算法研究 [J]. 仪器仪表学报,2024, 45(11): 243-251.

ZHUGE J CH, GAO H, LUO Q J, et al. Research on tightly coupled LiDAR-IMU SLAM algorithm [J].

Chinese Journal of Scientific Instrument, 2024, 45(11): 243-251.

作者简介



刘秋骅,2020 年于合肥工业大学获得学士学位,现为东南大学硕士研究生,主要研究方向为移动机器人定位与路径规划。
E-mail:1244920968@ qq. com

Liu Qiuhua received his B. Sc. degree from Hefei University of Technology in 2020. He is currently a master student at Southeast University. His main research interests include mobile robot localization and path planning.



徐晓苏(通信作者),1982 年于东南大学获得学士学位,1985 年于东南大学获得硕士学位,1991 年于东南大学获得博士学位。现为东南大学仪器科学与工程学院教授,主要研究方向为惯性导航、组合导航、卫星导航、视觉导航、智能滤波技术。

E-mail:xxs@ seu. edu. cn

Xu Xiaosu(Corresponding author) received his B. Sc. degree from Southeast University in 1982, received his M. Sc. degree from Southeast University in 1985, and received his Ph. D. degree from Southeast University in 1991. He is currently a professor with the School of Instrument Science and Engineering at Southeast University. His main research interests include inertial navigation, integrated navigation, satellite navigation, visual navigation, and intelligent filtering technologies.