DST-Calib：双路径、自监督、无目标的激光雷达-相机外参标定网络

Zhiwei Huang、Yanwei Fu、Yi Zhou、Xieyuanli Chen、Qijun Chen 和 Rui Fan ✉

arXiv:2601.01188v1 [cs.RO] 3 Jan 2026

摘要—激光雷达-相机外参标定对于机器人感知系统中的多模态数据融合至关重要。然而，现有方法通常依赖人工制作的标定靶标（例如棋盘格）或特定的静态场景类型，因此在真实世界的自主系统和机器人应用中的适应性与部署能力受到限制。本文提出首个能够以在线方式运行且无需特定标定靶标的自监督激光雷达-相机外参标定网络。我们首先发现，现有方法存在显著的泛化性能下降问题，其原因在于传统的单侧数据增强策略。为克服这一局限，我们提出一种新颖的双侧数据增强技术，利用估计深度图生成多视角相机视图，从而提高训练过程中的鲁棒性和多样性。基于这一数据增强策略，我们设计了一种双路径自监督标定框架，降低对高精度真值标签的依赖，并支持完全自适应的在线标定。此外，为了改善跨模态特征关联，我们以差异图构建过程取代传统的双分支特征提取设计，明确关联激光雷达特征与相机特征。这不仅提高了标定精度，还降低了模型复杂度。在五个公开基准数据集以及我们自行采集的数据集上进行的大量实验表明，所提出的方法在泛化能力方面显著优于现有方法。

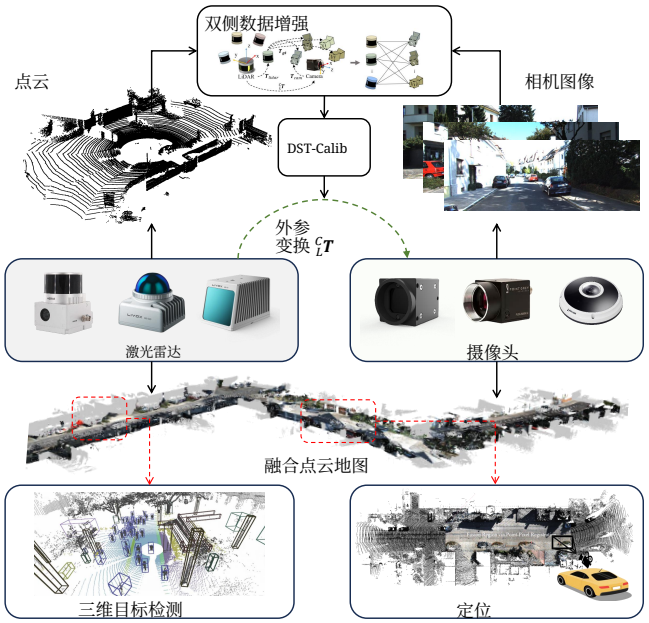

图 1. 我们提出的 DST-Calib 可在自然环境中估计激光雷达扫描与相机图像之间具有六个自由度（6-DoF）的外参变换。它可直接用于目标检测、里程计和定位等机器人任务。

一、引言

具有人类般智能的机器人长期以来一直是机器人学研究的核心愿景 [1]。如今，随着多模态传感器融合系统的快速发展，这一愿景正日益成为现实。通过整合来自各类传感器的互补信息，这些系统显著增强了机器人感知能力，使机器人能够在真实环境中可靠地执行复杂任务[2]。最广泛使用的传感器包括激光雷达和摄像头：前者提供精确的几何与空间测量，而后者捕捉丰富的语义和纹理细节 [3]。融合后，这些模态的互补优势使机器人能够实现稳健的环境理解 [4]。这种增强的感知能力对于支持多种基础机器人任务的可靠运行至关重要，包括里程计 [5], 目标识别 [6],和定位 [7]。

如图所示，1，激光雷达-相机外参标定（LCEC）通过估计两个传感器之间的外参变换，是实现有效数据融合的核心基础过程。现有的 LCEC 方法通常根据算法是否依赖预先定义的标定靶标，分为基于靶标和无目标两类。在存在标定靶标（例如棋盘格或定制三维图案板）的静态、受控环境中，基于靶标的标定方法能够提供稳定可靠的性能 [8], [9]。然而，随着动态应用需求的不断增长（例如自主移动机器人和多机器人协同感知），外参可能因振动、机械冲击或传感器主动移动而发生变化，基于靶标的方法因此变得不切实际，且往往无法维持有效标定

Zhiwei Huang 就职于控制科学与工程系，同济大学电子与信息工程学院，中国上海市201804（电子邮件：zhiwei.huang@outlook.com）。Yanwei Fu 任职于同济大学机械工程系、机械工程学院，中国上海市201804（电子邮件：2350435@tongji.edu.cn）。Yi Zhou 任职于湖南大学机器人学院，中国长沙市410082（电子邮件：eeyzhou@hnu.edu.cn）。Xieyuanli Chen 任职于国防科技大学智能科学与技术学院，中国长沙市410073（电子邮件：chenxieyuanli@hotmail.com）。Rui Fan 任职于同济大学控制科学与工程系、电子与信息工程学院、上海智能自主系统研究院、智能自主系统国家重点实验室及智能自主系统前沿科学中心，中国上海市201804；同时任职于西安交通大学人工智能与机器人研究所、人机混合增强智能全国重点实验室，中国陕西省西安市710049（电子邮件：rui.fan@ieee.org）。✉ 通讯作者：Rui Fan。

以实现可靠的数据融合。尽管已有多种传统的无目标 LCEC 方法被提出，通过利用跨模态几何纹理或互信息来对齐激光雷达点云与相机图像，从而摆脱对特定标定靶标的依赖 [10]–[12], 但它们仍难以适应复杂且非结构化的真实场景，且往往精度不足，仍有很大的改进空间。

深度学习技术的进步推动了研究者对提升传统无目标算法准确性和效率的深入探索。研究人员尝试设计端到端深度神经网络（DNN） [13]–[18] ，利用传感器输入直接回归六自由度外参。这些基于深度神经网络的方法易于训练，并且在使用训练期间相同的相机拍摄的图像进行评估时，通常能够达到较高的精度。然而，这些方法有一个主要缺点：依赖特定相机的参数。在这种情况下，针对单个激光雷达—相机对训练的端到端 DNN 模型往往会过拟合训练数据集中特定的内外参，即使在同一数据集中换用另一台相机进行评估，也无法泛化。此外，我们观察到一种固有的数据不平衡现象，导致网络在外参预测时几乎完全依赖激光雷达点云。在传统的训练数据准备过程中，相机位置通常是固定的，且以往工作采用单侧数据增强策略，仅围绕相机位姿引入扰动。如图2所示，每个投影点云恰好对应一张相机图像，从而形成多对一映射。在这种映射关系下，DNN 的相机图像分支在训练过程中会逐渐退化。固定的相对位姿关系使网络过拟合于激光雷达投影中常见的简单冗余特征，导致其记忆外参，而非学习有意义且稳健的跨模态相关性。因此，训练后的模型会高度依赖训练期间使用的特定相机，无法泛化至不同的传感器配置；一旦更换相机，就必须重新训练。

在本文中，我们旨在解决这些重大问题
存在于以往基于深度神经网络的LCEC方法中，并提升了当前无目标方法的泛化能力。首先，我们证明，如果通过双侧数据增强使相机位置多样化（确保单个点云投影映射到多张相机图像，形成多对多关系），那么相同的网络架构也能显著提升泛化能力。为构建这种双侧数据增强，我们采用单目深度估计，从原始相机图像重建相机深度点云。我们设计了一种高效的深度校正算法，**深度锚点细化**（DAR），利用激光雷达点云提供的稀疏引导校正初始深度估计。将校正后的相机深度图转换为三维深度云后，我们便能从激光雷达侧和相机侧的不同透视视图生成高质量深度投影，从而实现均衡的双侧数据增强。此外，为了解决对真实外参的高度依赖，

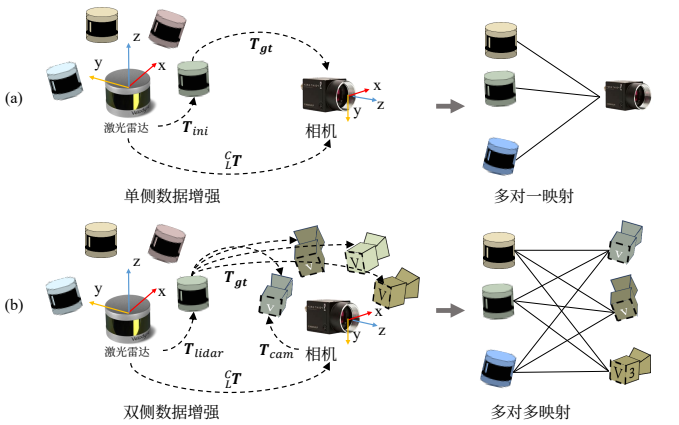

图 2. LCEC网络训练过程中不同数据增强方法的映射关系：(a) 具有多对一映射的经典单侧数据增强；(b) 我们提出的具有多对多映射的双侧方法。

我们提出了DST-Calib，这是首个支持自监督学习的无目标LCEC网络。与以往网络不同，DST-Calib包含双路径：全监督路径和自监督路径。当真实值不可用时，DST-Calib可通过在线自监督学习执行粗略的六自由度外参标定或高精度仅旋转标定。当真实标签可用时，DST-Calib可通过全监督训练实现准确且稳健的标定，并能轻松泛化到未见过的环境。此外，我们证明，先前工作采用的经典双分支架构（由分别处理激光雷达和相机输入的两个独立特征提取分支组成）并非实现高标定精度的必要条件。通过构建激光雷达与相机深度投影之间的差异图，我们显式关联激光雷达数据与相机数据，从而省去一个特征提取分支，构建单分支架构。尽管模型参数更少，这种单分支方法的表现仍优于传统双分支架构。此外，针对五个公开数据集开展的大量实验涵盖总计23种不同的传感器配置，证明了DST-Calib的优异性能。仅在3种传感器配置上训练的模型，无需微调即可直接泛化到所有其他传感器配置，相较于以往的无目标LCEC，泛化能力大幅提升。此外，现有公开数据集大多采用机械旋转式激光雷达，而很少有数据集提供由固态激光雷达通过可重复扫描采集的稠密四维点云。为推动这一方向的研究，我们构建了一个新的公开数据集，其中包含大量稠密点云与相机图像对，涵盖在多样的室内和室外环境中采集的5个序列。

总而言之，我们的创新贡献如下：

- 我们重新审视了以往基于深度神经网络的 LCEC 网络，并发现不平衡的单侧训练数据增强会导致泛化性能大幅下降。为解决这一泛化问题，我们提出了一种新颖的双侧训练数据增强策略。

- DAR 旨在借助激光雷达点云的引导，对单目相机图像的深度进行优化。
- DST-Calib是首个支持自监督学习且不依赖任何传感器特定参数的LCEC网络。
- 在五个公开数据集上进行的大量实验表明，单侧数据增强存在泛化问题，验证了我们提出的双侧数据增强的必要性，并展现了DST-Calib最先进（SoTA）的性能。此外，我们记录了一个新的真实世界数据集LCScenes，用于支持在固态激光雷达采集的密集点云上对LCEC网络进行训练和评估。

II. 相关工作

A. 传统无目标方法

为了消除对特定标定靶标的需求，以及离线数据采集和优化所带来的开销，研究人员开始开发无目标方法，使其能够在传感器运行环境中进行外参标定。在早期尝试中，研究 [10], [12], [19] 通过匹配从激光雷达点云和相机图像中提取的跨模态边缘或互信息来估计相对位姿。尽管这些方法在几何特征和纹理特征丰富的某些场景中效果良好，但它们高度依赖均匀分布的线条和丰富的纹理细节，从而显著降低了鲁棒性。近年来深度学习的发展推动了大量旨在改进无目标标定方法的研究。许多工作 [11], [20]–[26] 将基于学习的工具融入标定流程，以提升鲁棒性和效率。例如， [21]通过匹配经语义分割提取的车道线标记和类似杆状的结构来实现标定。类似地， [23] 将停车标志视为语义地标，并利用卡尔曼滤波方案逐步优化标定参数。最近，研究 [25] 提出了直接视觉激光雷达标定（DVL），这是一种点级方法，利用SuperGlue [27]直接在激光雷达与相机模态之间构建三维—二维对应关系。

B. 基于*DNN*的无目标方法

近年来，人们提出了端到端的基于深度神经网络的方法，用于估计激光雷达与相机传感器之间的外参。此前基于深度神经网络的LCEC网络 [28]主要沿袭了RegNet的开创性工作。如图3（a）所示，这些方法采用了类似的双分支架构，由两个特征提取分支、一个特征相关模块和一个位姿回归模块组成。后续研究，如CalibNet [17],RGGNet [18], CalibDNN [16],LCCNet [14], 和CalibDepth[29],均沿用这种双分支架构，并提出不同的网络模块以提升性能。

在训练过程中，这些基于深度神经网络的方法采用类似的单侧数据增强方法。生成相机侧周围不同视角下的激光雷达投影，以模拟标定初始估计 [14]。单侧数据增强策略的基本假设是：通过投影相机中心周围足够大的失准范围内的深度图像，网络可以学习激光雷达点云与相机图像之间的对应关系，从而准确估计外参。然而，这种单侧数据增强在输入的激光雷达投影与相机图像之间建立了多对一映射关系。尽管激光雷达投影是从相对于相机的随机视角采集的，但每个投影仅对应一个相对位姿。因此，模型很容易仅凭激光雷达投影的特征推断外参变换，而无需利用相机数据中的任何信息。然而，对于标定任务，相机位置是不确定的。激光雷达投影与相机图像之间应建立多对多映射。换言之，每个激光雷达投影可能对应无数种可能的相机图像。我们认为，只有在这一假设下训练的模型才可能具备适应其他传感器配置的泛化能力。

此外，我们还可视化了采用传统单侧数据增强策略训练的双分支网络所生成的特征图。如图3所示，相机分支生成的特征图几乎全黑，几乎不包含任何有效信息。为进一步验证这一点，我们在训练时将RGB输入替换为全零图像。不出所料，训练后的模型仍能在训练数据集内准确预测外参。这进一步证明，当前训练策略存在根本性误导，因为相机分支并未对外参估计作出贡献。矛盾的是，这一现象也表明，对于外参矩阵回归，单个分支就足以取得有竞争力的结果。事实上，双分支架构甚至可能妨碍有效特征的学习，因为激光雷达点云与相机图像之间显著的模态差异增加了特征相关性的难度。

总而言之，现有的LCEC网络主要存在两个关键问题：

1）采用具有误导性的单侧数据增强方法训练的模型，只能估计特定单一激光雷达-相机配置的外参，且无法泛化到其他配置。2）双分支架构不仅对特征提取和相关性计算而言并非必需，还增加了跨模态特征关联的难度。

为解决这两个关键问题，我们提出了一种新颖的双侧数据增强策略和一种支持自监督学习的双路径网络架构。与以往基于深度神经网络的方法不同，我们的DST-Calib不再受限于特定的传感器参数，可直接应用于各种具有挑战性的真实场景，实现在线无靶标标定。

III. 方法论

本节首先介绍一种基于双侧数据增强的新型训练策略。接下来，我们

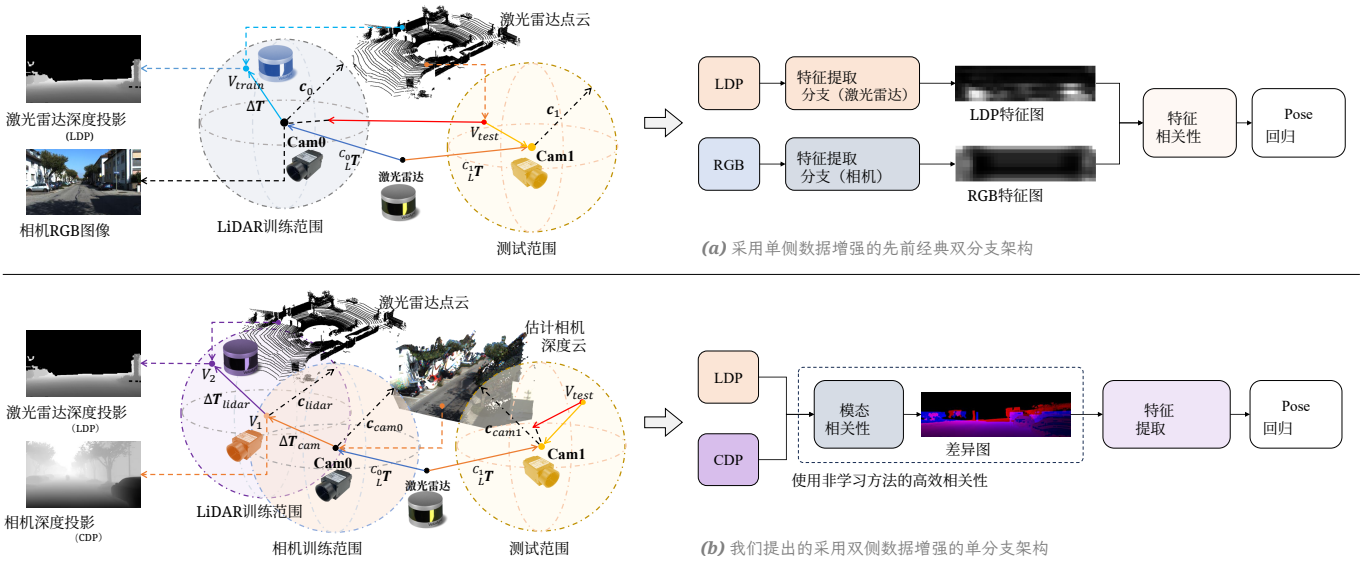

图3. 基于深度神经网络的激光雷达-相机外参标定流程对比：(a) 采用单侧数据增强的经典双分支架构；(b) 我们提出的采用双侧数据增强的单分支架构。

介绍所提出的新型标定框架，该框架同时支持全监督和自监督学习。最后，我们详细说明 DST-Calib 的创新设计，包括一种单分支网络架构。该架构通过分析由激光雷达点云和相机深度点云构建的差异图，直接估计 6 自由度外参变换。

A. 双侧数据增强

LCEC 的目标是估计

$$ {}^{C}_{L}\boldsymbol{T} = \begin{pmatrix} {}^{C}_{L}\boldsymbol{R} & {}^{C}_{L}\boldsymbol{t} \\ \boldsymbol{0}^{\top} & 1 \end{pmatrix} \in SE(3), \tag{1} $$

激光雷达与相机之间的外参变换。 ${}^{C}_{L}\boldsymbol{R} \in SO(3)$ 表示旋转矩阵， ${}^{C}_{L}\boldsymbol{t}$ 表示平移向量，$\boldsymbol{0}$表示零列向量。

以往基于深度神经网络的 LCEC 网络采用经典的单侧数据增强策略（如图2（a）和图3（a）所示）来构造训练样本。具体而言，它们使用 $\Delta\boldsymbol{T}$在相机周围有限的扰动范围内生成未对齐点云投影。对于每个 $\Delta\boldsymbol{T}$，分别存在一个对应的初始外参变换

$$ \boldsymbol{T}_{ini} = \Delta\boldsymbol{T}({}^{C}_{L}\boldsymbol{T}) \tag{2} $$

以确定激光雷达深度投影（LDP）图像的投影位置。每张生成的 LDP 图像都直接对应于从同一相机视角拍摄的 RGB（或灰度）图像，从而形成 LDP-RGB训练对。这种多对一的数据映射关系会误导网络，使其只依赖激光雷达分支的输入，因为仅凭激光雷达输入就足以“猜出”两个传感器之间的外参变换。采用这种数据增强策略训练的模型，在应用于与训练数据集不同的其他传感器配置时，泛化性能会大幅下降。

为从根本上解决这一泛化问题，我们提出了一种新颖的双侧增强框架（如图3(b)所示）。与文中（2）定义的单侧方法不同，后者将增强限制在激光雷达一侧，而这种双侧方法需要确保每个激光雷达投影都对应多份相机数据，从而建立多对多映射。为实现这种多对多映射，我们利用单目深度估计（Depth Anything V2[30] 和 MoGe2 [31]）从原始相机图像 ${}^{C}_{D}\boldsymbol{I} \in \mathbb{R}^{H\times W}$ 生成相机深度投影（CDP）图像 ${}^{C}_{I}\boldsymbol{I} \in \mathbb{R}^{3\times H\times W}$ ，并进一步设计深度细化算法，以获得可靠的深度点云 $\mathcal{Q} = \{[x_i, y_i, z_i]^{\top}\}_{i=1}^{M}$ ，用于生成任意位姿下的投影。我们为激光雷达侧和相机侧分别配置了一系列初始外参变换：

$$ \begin{aligned} \boldsymbol{T}_{cam} &= \Delta\boldsymbol{T}_{cam}({}^{C}_{L}\boldsymbol{T}), \\ \boldsymbol{T}_{lidar} &= \Delta\boldsymbol{T}_{lidar}(\boldsymbol{T}_{cam}) \end{aligned} \tag{3} $$

分别位于两个独立的数据增强中心。CDP图像将在以相机中心为起点的失准范围 $\boldsymbol{c}_{cam}$ 内，于 $\boldsymbol{T}_{cam}$ 处采集。激光雷达深度投影图像（LDP） ${}^{L}_{D}\boldsymbol{I}$ 将在以生成的 CDP 图像位置为起点的偏差范围 $\boldsymbol{c}_{lidar}$ 内，使用位姿 $\boldsymbol{T}_{lidar}$ 生成。$\Delta\boldsymbol{T}_{cam}$和 $\Delta\boldsymbol{T}_{lidar}$在由失准范围 $\boldsymbol{c}_{cam}$ 和 $\boldsymbol{c}_{lidar}$确定的区域内随机采样。标定网络的目标是估计位于$\boldsymbol{T}_{lidar}$处的虚拟激光雷达（记为 $lidar$）与位于 $\boldsymbol{T}_{cam}$处的虚拟相机（记为 cam）之间的相对外参变换 $\boldsymbol{T}_{gt}$，其定义如下：

$$ \boldsymbol{T}_{gt} = \boldsymbol{T}_{cam}(\boldsymbol{T}_{lidar})^{-1}. \tag{4} $$

该方法建立了真正的多对多映射，并通过确保以下两点大幅提升数据多样性：(1) 每个激光雷达点云投影对应多个相机视角；(2) 每张相机图像都与多种不同的激光雷达采样配置配对。

在这种增强方案下，网络无法仅凭单一模态推断外参，从而促使其学习有效的跨模态上下文关系。

B. 基于深度锚点的高效深度细化

单目深度估计在许多场景下通常无法提供准确的度量深度。为获得高质量的深度图并确保双侧数据增强，我们提出了 DAR，一种利用从稀疏激光雷达点云中提取的深度锚点来校正由原始相机图像估计得到的初始深度图的深度细化方法。该过程重建出具有可靠度量精度的稠密深度图。DAR 的详细过程如下所述：

1) 通过锚点进行分段线性深度细化：对于每一对 LDP 和 CDP，我们可以提取一组 K深度锚点

$$\mathcal{A} = \{(d_0^C, d_0^L), (d_1^C, d_1^L), \ldots, (d_{K-1}^C, d_{K-1}^L)\}, \tag{5}$$

其中，$d_i^C \in [0,1]$ 表示归一化单目相机深度，$d_i^L \in \mathbb{R}_+$ 表示具有准确度量值的对应激光雷达深度。我们通过连接相邻锚点与直线段，对整个估计深度图 ${}_D^C\hat{\boldsymbol{I}}_0$ 执行简单但有效的分段线性重映射 $f\colon [0,1] \to \mathbb{R}_+$

$$f(d^C) = \begin{cases} d_0^L, & d^C \le d_0^C, \\ d_{i-1}^L + \dfrac{d_i^L - d_{i-1}^L}{d_i^C - d_{i-1}^C}\left(d^C - d_{i-1}^C\right), & d_{i-1}^C < d^C \le d_i^C, \\ d_{K-1}^L, & d^C > d_{K-1}^C. \end{cases} \tag{6}$$

其中 $i = 1, \ldots, K-1$。实际应用中，如有需要，首尾线段会被投影到锚点范围之外的常数值。

在建立式（6中的分段线性深度细化策略后，给定初始估计深度图在像素 ${}_D^C\hat{\boldsymbol{I}}_0(u,v)$ 处的深度估计值 $\boldsymbol{p}$，校正后的深度可按如下方式获得：

$${}_D^C\hat{\boldsymbol{I}}(u,v) = f\left({}_D^C\hat{\boldsymbol{I}}_0(u,v)\right). \tag{7}$$

该方法在保持深度单调顺序的同时，使其局部适应锚点导出的激光雷达深度。这种简单的映射在锚点分箱内进行线性校准，同时确保全局深度顺序一致性，因此计算效率高且易于实现。它尤其适用于实时应用，或作为后处理细化步骤。尽管在某些孤立的前景物体（例如树干或行人）处，深度估计并不总是像其他区域那样保持相同的单调分段线性特性，但我们认为，DAR 整体的深度细化质量足以满足多视图数据增强的需求。

2) 单调近线性锚点选择：（6），如果深度锚点足够准确，DAR 很可能会获得比单目深度估计得到的初始深度图可靠得多的结果。因此，关键问题是如何获得可靠且准确的深度锚点。为解决这一问题，如图所示。

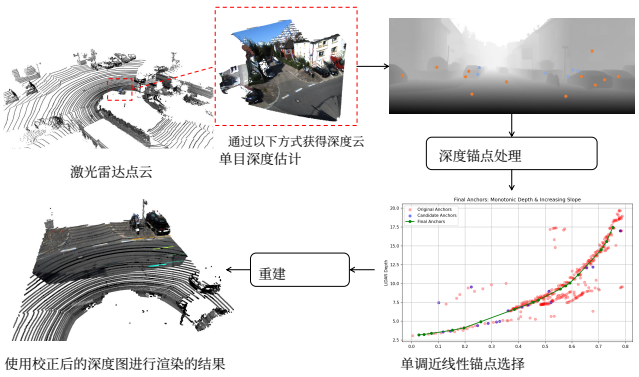

图4. 通过激光雷达-相机深度锚点重建进行深度细化。



图5. 不同数据集上初始估计深度云与经过我们的 DAR 校正后的深度云之间的定性比较：(a)–(d) 分别为 RGB图像、初始估计深度云、校正后的深度云，以及由激光雷达扫描得到的对应点云。

4，DAR 首先将估计相机深度和激光雷达深度按从小到大的顺序排序，并从两个深度投影中选取坐标相同位置的深度值，作为初始深度锚点。实际上，这些深度锚点是直接根据深度投影位置选取的对应关系。考虑到初始深度锚点的可靠性不确定，我们采用单调近线性锚点选择来筛除不可靠的锚点，并利用更可靠的锚点重建更好的深度图。

在此次深度锚点选择中，给定一组含噪深度锚点对应关系 $\mathcal{A} = \{(d_i^C,\ d_i^L)\}_{i=1}^N$，其中 $d_i^C \in [0,1]$ 为从相机图像估计得到的归一化深度，$d_i^L \in \mathbb{R}_+$ 为激光雷达深度。令 $\tilde{\mathcal{A}} = \{\ d_{(i)}^C\ d_{(i)}^L\ \}_{i=1}^N$ $\mathcal{A}$ d^C $d_{(1)}^C \le (,\ )$ $\cdots \le d_{(N)}^C$ 表示按 排序后的结果，其中 。目标是选择一个锚点的有序子序列 $\mathcal{S} = \{(d_k^C,\ d_k^L)\}_{k=1}^K$，使 K 尽可能大，并满足：(1) 在 d^C 和 d^L 上均单调不减；(2) 其局部斜率单调不减，以避免振荡，从而得到近似线性且形状稳定的校准曲线。

对于 $\mathcal{S}$ 中的连续点，其割线斜率定义为

如下：

$$s_k := \frac{d^L_{k+1} - d^L_k}{d^C_{k+1} - d^C_k}, \quad k = 1, \ldots, K-1. \tag{8}$$

我们将离散凸性（斜率递增）条件与单调性约束结合起来：

$$\begin{aligned} & s_k \le s_{k+1}, \quad k = 1, \ldots, K-2, \\ & d^C_1 < d^C_2 < \cdots < d^C_K, \\ & d^L_1 \le d^L_2 \le \cdots \le d^L_K, \end{aligned} \tag{9}$$

以确保满足凸性的离散类比（对应于非负二阶差分），从而有效抑制局部极值和振荡性（“之字形”）行为，同时仍允许斜率平滑、渐进地变化。

最后，我们将锚点选择表述为一个组合优化问题：

$$\max_{\mathcal{S} \subseteq \bar{\mathcal{A}}} |\mathcal{S}|, \quad \text{s.t. (9) hold for } \mathcal{S}. \tag{10}$$

问题（10）旨在寻找最长的单调凸子序列，该子序列表示具有非递减 d^C、非递减 d^L以及非递减割线斜率的最长序列。在实现中，我们引入二次动态规划来识别最长序列，该方法对离群值具有鲁棒性，并在保持 d^C覆盖范围的同时促进近似线性。伪代码见补充材料中的算法1 。

C. DST-Calib 的网络架构

如图 6所示，DST-Calib 采用了一种新颖的框架，由两条路径组成：全监督路径和自监督路径。在全监督路径中，DST-Calib 使用评估模块推断输入激光雷达点云与相机深度点云之间的相似度，并粗略估计外参。该路径的监督信号由标定真值生成。在自监督路径中，DST-Calib 使用位姿估计器推断六自由度外参，其依据包括点云相似度、外参的初始猜测，以及全监督路径中评估网络产生的先验结果（可选）。这两条不同的路径为具体应用提供了以下三种选项，确保其能够在各种环境和传感器配置下实现稳定的性能：

- **仅全监督路径**：仅启用全监督路径时，DST-Calib 使用标定真值来监督网络训练。外参通过评估模块直接回归，并可估计激光雷达点云与相机深度点云之间的相似度。当外参的初始估计与真值位置非常接近时，这

该方法高效、精度高，并能适用于各种具有挑战性的条件。

- **仅使用自监督路径**：仅启用自监督路径时，DST-Calib 利用位姿估计器，在点云相似度、外参初始猜测以及

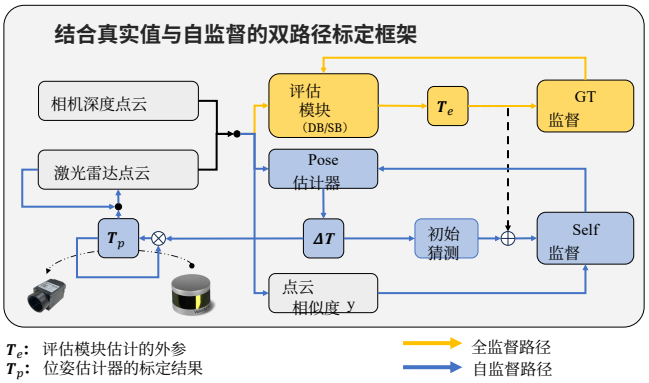

图 6. 提出了一种新颖的、基于学习的标定框架，结合真实值和自监督。该架构有两条可选路径。

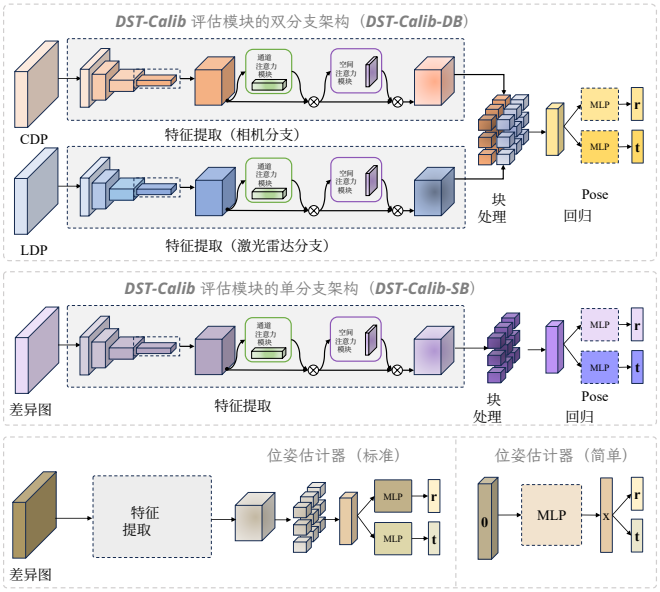

图 7. 所提出的 DST-Calib 的主要组件：DB、SB 和位姿估计器。

外参的监督下，并在条件允许时利用评估模块给出的先验来推导外参。这种应用方式尤其适用于仅旋转的标定任务。

- **两条路径均启用**：当两条路径均启用时，即使失准范围非常大，DST-Calib 仍能实现稳健且高精度的标定，并具备良好的泛化能力。如果有评估模块的训练数据，强烈推荐采用此方法，以获得出色的性能。

如图7所示，DST-Calib的主要组成部分包括评估模块和位姿估计器（PE）。对于评估模块，我们设计了两种不同的架构，即双分支（DB）和单分支（SB），以探究经典双分支结构与我们提出的单分支架构之间的本质差异。具体而言，DB为激光雷达输入和相机输入分别设置了两个独立分支，而SB仅包含一个分支，用于接收由原始激光雷达点云和相机深度点云构建的差异图。DB和

SB架构均引入了CBAM模块，以增强特征提取和跨模态相关性。此外，我们还加入了分块处理，通过结合激光雷达输入和相机输入的不同区域，进一步提升特征相关性。在以下小节中，我们将介绍最重要部分的细节：差异图生成、特征提取与相关性、位姿回归以及位姿估计器。

1) 差异图构建：差异图构建的主要目标是有效关联激光雷达和摄像头的数据，在保留必要信息的同时展现两者之间的差异。构建差异图的方法有很多。对于标定任务，纳入深度信息以及激光雷达点云与相机深度云之间的深度差异至关重要。本文提出了一种简洁而有效的差异图生成策略，用于融合激光雷达和相机的深度投影。该方法保留原始深度测量值，同时放大两种模态之间的差异，从而增强视差线索。具体步骤如下。

给定激光雷达深度投影图像 ${}^{L}_{D}\boldsymbol{I}$ 及其对应的校正后的相机深度图 ${}^{C}_{D}\boldsymbol{I}$ （大小为 $H \times W$的单通道深度图），我们旨在基于阈值$e_{tar} > 0 \in \mathbb{R}$ 构建三通道差异图像 $\boldsymbol{D}$（e_{tar} 为满足最低要求的目标平移标定误差）。具体而言，我们将逐像素差异定义如下：

$$\Delta(u,v) = {}^{L}_{D}\boldsymbol{I}(u,v) - {}^{C}_{D}\boldsymbol{I}(u,v), u = 1,\dots,H;\ v = 1,\dots,W. \tag{11}$$

然后，差异图构建函数定义如下：

$$\boldsymbol{D}(u,v) = \left({}^{L}_{D}\boldsymbol{I}(u,v),\ [\Delta(u,v)]^{e_{tar}}_{+},\ [\Delta(u,v)]^{e_{tar}}_{-}\right), \tag{12}$$

其中

$$[\Delta]^{e_{tar}}_{+} = \begin{cases} \Delta, & |\Delta| > e_{tar}, \\ 0, & \text{otherwise}, \end{cases} \quad [\Delta]^{e_{tar}}_{-} = \begin{cases} \Delta, & |\Delta| \le e_{tar}, \\ 0, & \text{otherwise}. \end{cases}$$

所得差异图 $\boldsymbol{D}$ 包含标定所需的信息，包括激光雷达深度、激光雷达深度与相机估计深度之差，以及投影后的深度坐标。

2) 特征提取： 我们采用 ResNet 作为特征提取的骨干网络，双分支（DB）和单分支（SB）架构均采用相同的提取流程。输入数据首先被变换为特征图$\boldsymbol{F_0}$，随后通过 CBAM 模块进行处理，以生成具有空间与通道注意力的增强特征图 $\boldsymbol{F}$。

在 DB 架构中，需要两个独立分支，分别从 CDP 和 LDP 输入中提取特征图。相比之下，SB 架构只需一个提取分支，因为构建的差异图显式关联了激光雷达数据和相机数据。经典的 DB 架构从两种不同模态中提取特征，并依赖缺乏理论依据且不确定的特征相关性过程；而 SB 设计则显式地将激光雷达点云与相机深度点云关联在一个统一的分支中。这不仅减少了模型参数的数量，也提高了可解释性。

3) 分块处理与位姿回归： 不同于大多数以往直接拼接特征图的方法，我们采用基于分块的处理策略，以增强提取特征图中的区域关系。具体而言，完成特征提取后，特征图 $\boldsymbol{F}$被划分为 $n \times n$个网格块 $\boldsymbol{F_i}$。每个块均经过卷积模块处理，并压缩为更高层次的特征向量 $\boldsymbol{B_i}$。随后，将所得块特征进行拼接并展开为线性特征向量 $\boldsymbol{F}_p$，再通过一系列全连接层进行全局特征聚合。最后，我们将旋转和平移的输出流解耦，以应对潜在的模态特定差异。除非另有说明，网络的输出定义为位姿向量 $\xi = \begin{bmatrix}\boldsymbol{r}^{\top} & \boldsymbol{t}^{\top}\end{bmatrix} \in \mathbb{R}^{1\times 6}$，其中 $\boldsymbol{r}$ 为估计的旋转向量， $\boldsymbol{t}$为估计的平移向量。 $\boldsymbol{T}$ 为对应的外参变换矩阵。

4）位姿估计器：如图7所示，DST-Calib 中的位姿估计器是一个自监督模块，利用点云相似度来指导外参的估计。我们为该模块设计了两种可能的实现方式。第一种采用与评估模块类似的标准架构，从而能够对六自由度位姿进行基于特征的回归。第二种采用轻量级结构，由一个输入为恒定零向量的多层感知机构成，充当自动优化器。在这种情况下，外参变换利用点云相似度的反馈以及评估模块提供的标定误差进行迭代更新。借助该位姿估计器，DST-Calib 可以在完全自监督的情况下执行标定，无需任何真值标签或重新训练。此外，评估模块仅专注于学习点云相似度，该相似度与配置无关，并为自监督位姿估计器提供稳健反馈，以指导其适应不同的传感器配置。

D. 监督方法

与其他基于深度神经网络的激光雷达-相机外参标定方法不同，训练数据集的外参真值 ${}^{C}_{L}\boldsymbol{T}$ 并非监督所必需。我们采用激光雷达点云与估计相机深度点云之间的相似度来监督网络。不过，如果提供了外参真值，我们也可以通过预训练评估模块，并使用预训练模型强化相似度计算，从而提升整体标定精度。在本节中，我们将详细介绍监督方法，包括对评估模块的完全监督，以及对位姿估计器的自监督。

1) 评估模块的监督：对于评估模块，我们利用激光雷达到相机的真实外参来监督训练过程。损失函数由三个部分组成：旋转误差、平移误差和点云距离误差。

旋转误差定义如下：

$$\mathcal{L}_{r_{gt}} = \left\| \boldsymbol{R}_{cam}(\boldsymbol{R}\ \boldsymbol{R}_{lidar})^{-1} - \boldsymbol{I} \right\|_{1,1}. \tag{13}$$

平移误差定义如下：

$$\mathcal{L}_{t_{gt}} = \left\| \boldsymbol{t}_{cam} - (\boldsymbol{t}_{lidar} + \boldsymbol{t}) \right\|_2. \tag{14}$$

点云距离损失定义如下：

$$\begin{aligned}\mathcal{L}_{cloud} = \sum_{\boldsymbol{p}\in\mathcal{P}} \Big\| & \boldsymbol{R}(\boldsymbol{R}_{lidar}\boldsymbol{p} + \boldsymbol{t}_{lidar}) + \boldsymbol{t} \\ & -(\boldsymbol{R}_{cam}\boldsymbol{p} + \boldsymbol{t}_{cam}) \Big\|_2,\end{aligned} \tag{15}$$

它表示标定点云与使用真实值变换后的点云之间的距离。$\boldsymbol{p} = [x, y, z]^\top$ 是输入激光雷达点云 $\mathcal{P}$中的一个点。

最后，总损失函数是三个分量之和：

$$\mathcal{L}_{eva} = \mathcal{L}_{r_{gt}} + \mathcal{L}_{t_{gt}} + \mathcal{L}_{cloud}. \tag{16}$$

通过真实值监督，评估模块可以学习激光雷达点云与估计相机深度云之间相似度的不确定性分量。这有助于它在相对较小的范围内标定外参时表现良好，并在失准范围大得多时更好地引导位姿估计器。

2）位姿估计器的监督：位姿估计器的监督包含三个分量。第一个分量是 $\mathcal{L}_{t_{ini}}$，其方式与 $\mathcal{L}_{t_{gt}}$相同。不过，$\boldsymbol{t}_{gt}$ 被外参的初始猜测所取代。

第二个组成部分是变换后的激光雷达点云与估计相机深度云之间的倒角距离，其定义如下：

$$\mathcal{L}_{CD} = \frac{\alpha}{|\hat{\mathcal{P}}|}\sum_{\hat{\boldsymbol{p}}\in\hat{\mathcal{P}}}\min_{\boldsymbol{q}\in\mathcal{Q}}\|\hat{\boldsymbol{p}} - \boldsymbol{q}\|_2^2 \ + \ \frac{\beta}{|\mathcal{Q}|}\sum_{\boldsymbol{q}\in\mathcal{Q}}\min_{\hat{\boldsymbol{p}}\in\hat{\mathcal{P}}}\|\boldsymbol{q} - \hat{\boldsymbol{p}}\|_2^2, \tag{17}$$

ˆ其中，$\mathcal{P}$ 是使用估计的外参矩阵变换后的激光雷达点云，$\mathcal{Q}$ 是相机深度点云。α 和 β 是倒角距离两个组成部分的权重，默认均设为 0.5。

第三个组成部分是位姿估计器输出的外参矩阵 $\boldsymbol{T}$ 的外参评估分数，并以评估模块回归得到的位姿 $\boldsymbol{T}_{eva}$ 作为参考：

$$\mathcal{L}'_{eva} = a\left\|\boldsymbol{e}_{eva} - \boldsymbol{e}\right\|_2 + \left\|\boldsymbol{t}_{eva} - \boldsymbol{t}\right\|_2, \tag{18}$$

其中，$\boldsymbol{e}$是旋转分量 $\boldsymbol{R}$ 对应的欧拉角。a 是用于平衡旋转误差和平移误差的尺度参数。考虑到旋转误差在 1° 以内且平移误差在 0.1m 以内的外参标定结果可视为良好结果，我们将尺度参数 a 设为$a = 0.1/1 = 0.1$。最终损失 $\mathcal{L}_{pe}$ 也是这三个分量之和：

$$\mathcal{L}_{pe} = \mathcal{L}_{t_{ini}} + \mathcal{L}_{CD} + \mathcal{L}'_{eva}. \tag{19}$$

如果没有外参的初始猜测，$\mathcal{L}_{t_{ini}}$ 应为 0。如果无法在应用中对评估模块进行预训练，$\mathcal{L}'_{eva}$ 也应为 0。因此，即使未提供先验标定结果，且不具备预训练评估模块的条件，DST-Calib 仍可通过点云倒角距离的监督实现自监督学习。

E. 多帧优化

除了直接从每一帧推断外参外，我们还开发了一种多帧优化方法，利用相同传感器配置下从多个场景采集的数据所提供的约束，共同优化标定结果。这种多帧优化可以利用多帧激光雷达点云和相机图像，共同优化外参标定结果。设$\mathcal{T} = \{\boldsymbol{T}_1, \boldsymbol{T}_2, \ldots, \boldsymbol{T}_n\}$为 DST-Calib 多帧输出得到的一系列外参矩阵，并设 $h : SE(3) \to \mathbb{R}$ 为一个为每个外参矩阵分配质量分数的评分函数。给定选择比例 $x \in (0, 1],$，我们的目标是根据分数计算排名前 x 百分比外参的平均值。

在多帧优化中，使用评分函数评估每个 $\boldsymbol{T}_i$ 的可靠性。在全监督路径中，评分函数定义如下：

$$\begin{aligned}s_i &= h_f(\boldsymbol{T}_i, \boldsymbol{T}'_i) \\ &= \exp\Big(-\left(a\left\|\boldsymbol{e}'_i - \boldsymbol{e}\right\|_2 + \left\|\boldsymbol{t}'_i - \boldsymbol{t}_i\right\|_2\right)\Big), \quad i = 1, 2, \ldots, n,\end{aligned} \tag{20}$$

其中，$\boldsymbol{e}_i$ 分别是 $\boldsymbol{T}_i$的旋转分量 $\boldsymbol{R}_i$ 对应的欧拉角，$\boldsymbol{t}_i$ 是 $\boldsymbol{T}_i$的平移分量。$\boldsymbol{T}'_i$是评估模块在位置 $\boldsymbol{T}_i$处使用输入数据所得的评估结果。

在自监督路径中，评分函数由变换后的激光雷达点云与相机深度点云之间的倒角距离构造，定义如下：

$$s_i = h_s(\boldsymbol{T}_i) = \exp\big(-\mathcal{L}_{CD}(\hat{\mathcal{P}}_i, \mathcal{Q}_i)\big), \quad i = 1, 2, \ldots, n, \tag{21}$$

ˆ其中，$\mathcal{P}_i$ 是使用外参矩阵 $\boldsymbol{T}_i$ 变换后的激光雷达点云，$\mathcal{Q}_i$ 是 i帧中对应的相机深度点云。

计算每个估计外参矩阵的分数后，初始矩阵 $\mathcal{T}$ 将按分数降序排列。令 π：$\{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$为一个排列，使得 $s_{\pi(1)} \geq s_{\pi(2)} \geq \cdots \geq s_{\pi(n)}$，则要选取的外参矩阵数量由以下公式确定：

$$k = \lceil x \cdot n \rceil, \tag{22}$$

其中 $\lceil\cdot\rceil$ 表示向上取整函数，确保当 $x > 0$时至少选取一个向量。选定的子集 $\mathcal{T}_{\text{top}}$包含得分最高的 k 个向量 $\mathcal{T}_{\text{top}} = \{\boldsymbol{T}_{\pi(1)}\ \boldsymbol{T}_{\pi(2)}, \ldots, \boldsymbol{T}_{\pi(k)}\}$。随后，可以通过均匀加权或基于分数的加权来计算所选向量的平均值。在

基于分数的加权中，得分较高的矩阵对平均值的贡献更大。权重 w_j:

$$w_j = \frac{s_{\pi(j)}}{\sum_{i=1}^{k} s_{\pi(i)}}, \quad j = 1, 2, \ldots, k \tag{23}$$

每个选定矩阵的权重与其得分成正比。利用 $\mathcal{T}_{\text{top}}$ 中选定的外参矩阵，我们计算平移分量和旋转分量的联合平均值，$\boldsymbol{R}^*$ 和$\boldsymbol{t}^*$，从而得到最终优化后的外参 $\boldsymbol{T}^*$。

IV. 实验

本节将详细介绍并分析我们为验证新提出的 DST-Calib 而开展的大量实验评估。首先，我们介绍用于训练和评估该方法的数据集，然后说明评估指标和实现细节。其次，我们报告以往方法采用单侧数据增强所导致的泛化性能下降，并分析训练策略对泛化能力的影响。随后，我们报告与其他最先进的无靶标LCEC方法相比的定性和定量结果，并评估所提出的深度细化方法 DAR 的性能。最后，我们展示 DST-Calib 的跨域泛化结果，并开展大量消融研究，以验证各方法组件的有效性。

A. 数据集

我们在公开数据集 KITTI 里程计 [32], KITTI-360 [33], MIAS-LCEC[26] （包括 TF70 和 TF360）、nuScenes [34], Argoverse2[35], 以及我们新采集的数据集 LCScenes 上进行了大量实验。这些真实世界数据集在不同国家采集，采用不同的传感器配置，并包含不同类型的标定场景。表I 总结了我们用于训练和评估的数据集的主要特征。图 9展示了训练和测试传感器配置的详细信息。

LCScenes 使用我们定制的平台采集，如图 8所示。激光雷达和相机安装在置于三脚架上的可调平台上。该数据集包含大量具有不同外参、在不同室内外场景中采集的四维点云与二维相机图像对。该数据集分为五个序列，每个序列沿一条特定路线采集。激光雷达传感器为 Livox-Mid70。相机传感器的分辨率为 1200 × 800 和 2400 × 1200。

B. 评估指标

为全面评估 LCEC 方法的性能，我们遵循以往研究 [14], [25] ，使用幅值 e_r 作为欧拉角误差的幅值以及 e_t平移误差，表达式如下:

$$\begin{aligned} e_r &= \left\| \boldsymbol{e}^* - \boldsymbol{e}_{gt} \right\|_2, \\ e_t &= \left\| \boldsymbol{t}^* - \boldsymbol{t}_{gt} \right\|_2, \end{aligned} \tag{24}$$

用于量化标定误差。在（24）中， $\boldsymbol{e}^*$ 和 $\boldsymbol{e}_{gt}$ 分别表示估计的和真实的欧拉角向量，计算方式为

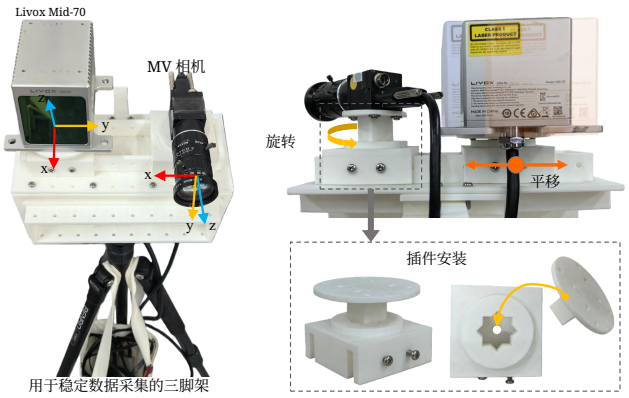

图 8. LCScenes 的数据采集平台。使用一台 Livox 固态激光雷达和一台 MindVision 相机采集稠密四维点云和 RGB图像。

表 I 评估数据集中的传感器配置。

数据集	激光雷达	摄像头	相机分辨率	包含的传感器中
KITTI 里程计	HDL-64E	2	1226 × 370/ 1241 × 376	1 × 激光雷达 1 × 相机
KITTI-360	HDL-64E	2	1408 × 376	✗
Argoverse2	2 ×VLP-32	9	2048 × 1550 / 1550 × 2048	1 × 激光雷达 2 × 相机
MIAS-LCEC-TF70	Livox-Mid70	1	1200 × 800	✗
MIAS-LCEC-TF360	Livox-Mid360	1	1200 × 800	✗
nuScenes	HDL-32E	6	1600 × 900/ 1600 × 1200	✗
LCScenes	Livox-Mid70	2	1200 × 800/ 2400 × 1200	✗

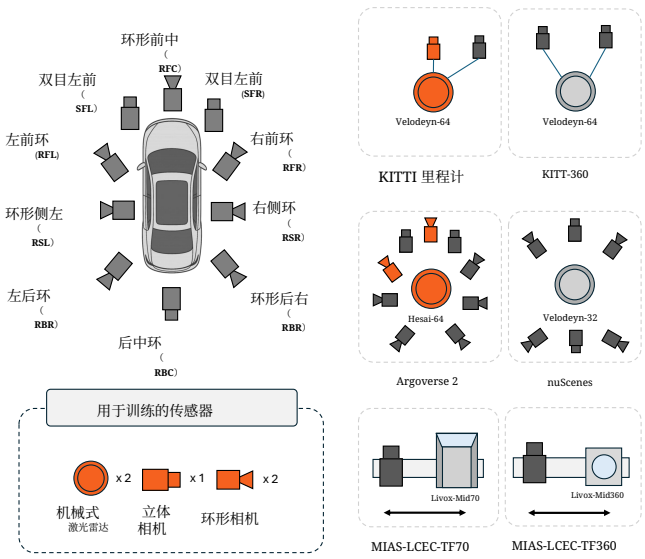

图 9. 训练数据集和测试数据集的不同传感器配置。橙色传感器用于训练，其余传感器用于测试。值得注意的是，为确保公平比较，与基于深度神经网络的方法相比，我们仅使用 KITTI 08序列的左摄像头进行训练。

分别来自旋转矩阵 $\boldsymbol{R}^*$ 和 $\boldsymbol{R}_{gt}$。类似地，$\boldsymbol{t}^*$ 和 $\boldsymbol{t}_{gt}$ 分别表示估计平移和真实平移

表II KITTI 里程计左摄像头和右摄像头上最先进的基于深度神经网络的方法的定量结果。最佳结果以粗体显示，次佳结果以下划线标出。

方法	初始范围	左摄像头							右摄像头						
		幅值		旋转误差（°）			平移误差（m）			幅值		旋转误差（°）			平移误差（m）		
		e_r（°）	e_t（m）	Yaw	俯仰	Roll	X	Y	Z	e_r（°）	e_t(米)	Yaw	俯仰	Roll	X	Y	Z
CalibDNN	[±10°, ±1.0m]	1.172	0.098	0.402	0.998	0.180	0.072	0.025	0.045	1.238	0.405	0.480	1.010	0.195	0.396	0.026	0.050
CalibDNN（无 RGB）	[±10°, ±1.0米]	1.169	0.096	0.405	0.994	0.171	0.070	0.024	0.045	1.218	0.407	0.473	0.999	0.184	0.397	0.026	0.050
CalibDNN	[±5°, ±0.5米]	**0.585**	0.059	**0.201**	0.493	0.106	0.042	**0.015**	**0.027**	0.692	0.328	**0.330**	0.512	0.141	0.323	0.017	0.033
CalibDNN（无 RGB）	[±5°, ±0.5m]	0.605	0.058	0.202	0.506	0.128	0.041	0.016	**0.027**	0.685	0.317	0.331	0.493	0.158	0.312	0.018	**0.032**
CalibDepth	[±10°, ±1.0m]	0.996	0.075	0.332	0.848	0.141	0.046	0.023	0.038	1.140	0.406	0.469	0.915	0.166	0.392	0.029	0.049
CalibDepth（无 RGB）	[±10°, ±1.0m]	1.106	0.073	0.332	0.971	0.134	0.045	0.023	0.038	1.181	0.409	0.466	0.966	0.161	0.396	0.027	0.048
CalibDepth	[±5°, ±0.5m]	0.638	**0.052**	0.268	0.498	0.137	0.025	0.017	0.036	0.904	**0.303**	0.520	0.507	0.332	**0.290**	0.025	0.046
CalibDepth（无 RGB）	[±5°, ±0.5m]	0.630	**0.052**	0.255	0.501	0.133	**0.024**	0.018	0.036	0.913	0.305	0.526	0.505	0.333	0.293	0.024	0.045
CalibNet	[±10°, ±1.0m]	1.109	0.167	0.333	0.955	0.206	0.115	0.046	0.086	1.182	0.389	0.346	1.018	0.226	0.363	0.048	0.084
CalibNet（无 RGB）	[±10°, ±1.0米]	1.251	0.164	0.575	0.984	0.190	0.117	0.042	0.079	1.238	0.407	0.510	0.993	0.224	0.384	0.050	0.082
CalibNet	[±5°, ±0.5m]	0.666	0.067	0.323	0.515	0.080	0.046	0.017	0.034	**0.665**	0.304	0.356	**0.481**	0.103	0.298	0.016	0.036
CalibNet（无 RGB）	[±5°, ±0.5米]	0.650	0.068	0.336	**0.486**	**0.070**	0.049	0.016	0.033	0.702	0.309	0.411	0.483	**0.089**	0.303	**0.015**	0.037

从激光雷达到相机的平移向量。

在单目深度估计中，沿用先前工作的做法[36]–[38]，使用若干标准指标评估性能。平均绝对误差（平均绝对误差）和均方根误差（均方根误差）衡量预测深度与真实深度之间的绝对差异和平方差异，以评估整体精度并加大对较大偏差的惩罚。绝对相对误差（绝对相对误差）和平方相对误差（平方相对误差）通过根据真实深度值对误差进行归一化，评估预测值与真实值之间的相对偏差，因此能够反映不同距离范围内的精度。此外，阈值准确率指标 δ_1、δ_2和 δ_3 用于计算预测深度落在逐渐放宽的乘性误差界限内的比例，从而反映估计的鲁棒性。

C. 训练详情

除非另有说明，我们均在单台 NVIDIA RTX4090D GPU 上训练 DST-Calib，总批次大小为 8，使用 AdamW优化器，基础学习率为 $5*10^{-4}$，权重衰减为 $1*10^{-4}$。训练过程中使用 OneCycle学习率调度器。对于每个 DST-Calib 实例，评估模块使用式（16）中定义的全监督损失训练 200 个周期。对于自监督路径中的位姿估计器，默认采用简单架构，并将其直接应用于多帧在线标定，假设一个序列中的所有帧外参保持一致。位姿估计器与标定过程并行进行在线自监督训练。在线标定数据序列中的所有帧均用于位姿估计器的自监督学习。如果帧数相对较少，则重复采样数据，以确保学习过程中至少处理 30 个批次。

此前大多数基于深度神经网络的方法仅在 KITTI 数据集上训练和测试其模型 [14], [16], [18]。值得注意的是，除 00 序列外，KITTI 里程计中的其他序列均被纳入这些基于深度神经网络的方法的训练数据集。为确保比较公平，在作者提供代码的情况下，我们复现了序列00上左、右摄像头的标定结果；否则，我们采用其论文中报告的左摄像头结果。对于我们的 DST-Calib，我们在 KITTI 里程计的 08 序列上进行训练，并在其他序列上进行测试。由于大多数基于学习的方法缺少用于自定义数据的应用程序接口，我们与这些方法的比较仅限于 KITTI 里程计的 00 序列。

为进一步提升 DST-Calib 的泛化能力，我们还使用来自 KITTI 里程计和 Argoverse2 的样本对其进行训练。使用多个数据集训练的这些模型用于验证跨域泛化能力。在多数据集训练中，使用 KITTI 里程计第 08 序列的左摄像头、Argoverse2 000 00-000 01 的环形前中相机，以及 Argoverse2 000 00-000 01 的环形前左相机作为训练样本。我们使用在 KITTI 里程计第 08 序列上训练得到的预训练权重作为初始模型参数。

在训练和测试期间，除非另有说明，双侧数据增强中激光雷达和相机两侧在平移和旋转所有轴上的随机偏差范围均设为总计$[\pm5°, \pm0.5\text{m}]$ 。在实验和复现先前工作时，我们发现每个轴上的均匀偏差较大时，网络难以收敛。为使训练过程更容易收敛，我们为每个轴设置一个权重 $[0.6, 0.2, 0.2]$。例如，在偏差 $[\pm5°, \pm0.5m]$下，旋转和平移各轴的具体偏差分别为 $[\pm3°, \pm1°, \pm1°]$ 和 $[\pm0.3\text{m}\ \pm0.1\ \text{m}\ \pm0.1\text{m}]$,。虚拟相机的投影尺寸默认设为 (256, 512)，焦距设为 600。此外，块处理器的块数默认设为 5。

D. 推理

在实验中，我们评估了五种不同类型的 DST-Calib，包括 DB、SB、SB*、PE 和 PE+SB*，它们是在不同的初始失准范围下训练的。DB 是全监督路径中的评估模块，采用双重架构。SB 是采用我们提出的单分支架构的评估模块。SB* 是单分支

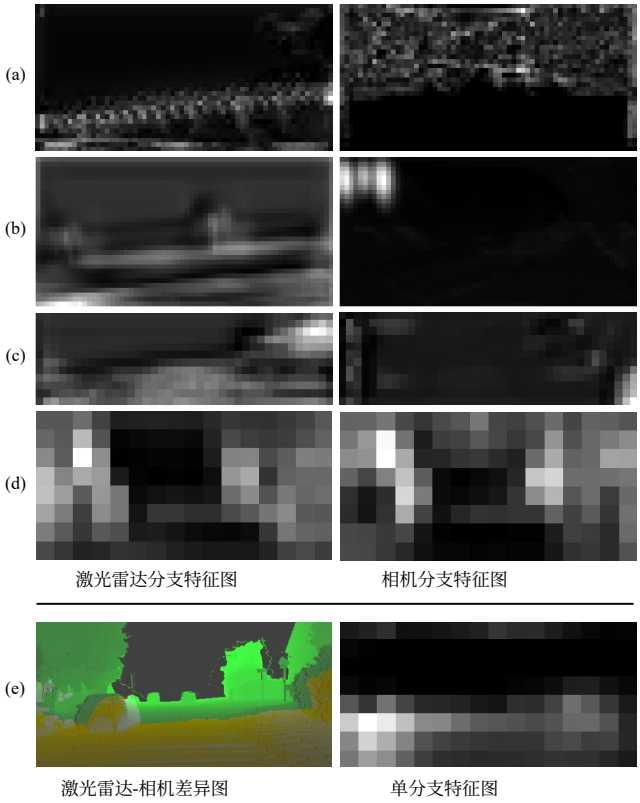

图10. 提取的特征图可视化结果：(a)-(d) CalibDepth、CalibNet 和 Cal- ibDNN 的激光雷达分支与相机分支特征图；(e) 构建的激光雷达-相机差异图及其特征提取结果。在单侧数据增强下，相机分支的特征提取结果有时几乎为空且毫无意义。

分支评估模块，在标定过程中使用多帧优化。PE 是自监督路径中的位姿估计器。PE+SB* 结合了全监督和自监督两种方式。它将 PE 的推理结果作为 SB 的先验，以获得精细的标定结果。值得注意的是，PE 和 PE+SB* 均采用多帧优化，其中假设激光雷达与相机之间的外参在整个序列中保持不变。为测试多帧优化，我们进行了 10 次独立的序列标定（每次使用不同的初始位姿），并以平均值作为总标定误差，以确保与其他实例进行公平比较。

E. 单侧数据增强不平衡导致的泛化性能下降

在本节中，我们进行了大量实验，以展示单侧训练数据增强策略（先前的 LCEC 网络广泛采用该策略）导致的泛化性能下降。定量结果见表 II。定性结果见图10。

1）使用不同传感器配置进行泛化实验：大量先前研究都使用 KITTI 序列中的左摄像头训练和测试其网络。然而，标定任务是一项灵活的任务，应能适应不同传感器位置下的各种相对位姿。我们测试

表 III 不同训练数据增强范围的定量比较（采用双分支架构）。

c_{lidar}	c_{cam}	传感器输入	KITTI Left00 e_r (°)	KITTI Left00 e_t(米)	KITTI Right00 e_r (°)	KITTI Right00 e_t(米)
[±5°, ±0.5米]	[±0°, ±0.0米]	双重	0.301	0.067	0.708	0.192
		无摄像头	0.757	0.091	0.776	0.229
		无激光雷达	2.082	0.221	2.144	0.237
[±5°, ±0.5m]	[±1°, ±0.1m]	双重	**0.168**	**0.042**	0.461	0.137
		无摄像头	1.002	0.169	1.135	0.294
		无激光雷达	2.446	0.228	2.505	0.257
[±5°, ±0.5m]	[±2°, ±0.2米]	双重	0.191	0.046	0.357	0.100
		无相机	1.762	0.154	1.704	0.211
		无激光雷达	2.521	0.250	2.586	0.246
[±5°, ±0.5米]	[±3°, ±0.3米]	双重	0.192	0.043	0.253	0.063
		无摄像头	1.409	0.213	1.376	0.198
		无激光雷达	2.105	0.233	2.056	0.226
[±5°, ±0.5m]	[±4°, ±0.4m]	双重	0.191	0.046	0.232	0.058
		无摄像头	2.162	0.201	2.190	0.227
		无激光雷达	2.678	0.235	2.666	0.274
[±5°, ±0.5米]	[±5°, ±0.5米]	双重	0.186	0.045	**0.218**	**0.054**
		无摄像头	1.748	0.249	1.715	0.276
		无激光雷达	2.377	0.241	2.399	0.248

了若干典型的基于深度神经网络的方法（包括 CalibDepth、Cal- ibDNN 和 CalibNet），在 KITTI 右摄像头上使用其基于左摄像头训练的模型进行测试。实验结果如表II所示，即使只是在同一标定场景中改变相机位置，性能也会显著下降。这表明训练好的模型缺乏泛化能力，只适用于特定的相机位置。

2）*RGB*图像分支中的性能退化现象：我们发现，当数据增强范围不平衡时，RGB分支会出现性能退化。如表II所示，即使从模型的相机分支中移除输入RGB图像，在相同的传感器配置下，模型仍能取得相近的性能。此外，我们还仅使用激光雷达分支的输入训练了模型，它们在训练数据集上也能表现良好。这进一步表明，单侧数据增强策略和双分支模型架构无法有效关联激光雷达点云与相机图像中的特征。在单侧数据增强所导致的多对一映射下，网络可以直接根据激光雷达投影的形状“猜测”相对位姿。此外，我们还对相机输入分支的特征提取结果进行了可视化。如图10所示，相机图像经过特征提取模块后生成的特征图几乎为空，毫无意义，这意味着输入的相机图像对网络估计外参没有贡献。相比之下，采用双侧数据增强策略时，DST-Calib生成的特征图保留了更丰富的上下文信息。

F. 训练策略如何影响模型性能

1）数据增强范围对泛化能力的影响：为进一步探究先前的训练策略为何导致性能不佳，以及数据生成范围如何影响泛化能力，我们设置了不同的

表 IV 有无激光雷达和相机输入时两种不同架构的定量比较。

数据源	模型	传感器输入	KITTI Left00 e_r (°)	KITTI Left00 e_t(米)	KITTI Right00 e_r (°)	KITTI Right00 e_t (米)
LDP-LDP	DST-Calib（DB）	双重	0.186	0.045	0.218	0.054
		无摄像头	1.748	0.249	1.715	0.276
		无激光雷达	2.377	0.241	2.399	0.248
LDP-CDP	DST-Calib（DB）	双重	0.524	0.101	0.664	0.147
		无摄像头	1.554	0.168	1.590	0.207
		无激光雷达	2.171	0.207	2.164	0.207
LDP-LDP	DST-Calib（SB）	双重	**0.136**	**0.033**	**0.140**	**0.035**
		无摄像头	1.865	0.244	1.885	0.242
		无激光雷达	2.066	0.265	1.978	0.264
LDP-CDP	DST-Calib（SB）	双重	0.366	0.078	0.505	0.109
		无摄像头	2.115	0.174	2.086	0.186
		无激光雷达	1.900	0.203	1.837	0.197

失准范围，并在 KITTI 里程计数据集的左、右摄像头上验证结果。为突出失准范围对数据增强的影响，我们使用相机视角下激光雷达点云的深度投影替代 CDP图像。将从不同视角采集的激光雷达投影输入 DST-Calib数据库，以获得其相对位姿。我们将 $\boldsymbol{c}_{lidar}$ 范围设为$[5°, 0.5\text{m}]$,，并将 $\boldsymbol{c}_{cam}$ 范围设为从 $[0°, 0\text{m}]$ 到 $[5°, 0.5\text{m}]$。表III所示结果表明，随着两种数据增强中心处的失准范围增大，泛化能力也会提高。当一侧的失准范围过小时，网络将丧失泛化能力，无法用于真实场景标定任务。当范围 $\boldsymbol{c}_{lidar}$ 和$\boldsymbol{c}_{cam}$ 相近时，准确率和泛化能力都较为理想。这表明，平衡的双侧数据增强对于 LCEC网络获得泛化能力是必要的。

2）模型输入的模态差异的影响：我们还研究了输入数据中的模态差异所带来的影响。在本实验中，我们比较了采用不同输入配置训练的模型性能。如表IV所示，采用 LDP-LDP 输入训练的模型，其标定误差明显低于采用 LDP-CDP 输入训练的模型。这表明，缩小两种传感器输入之间的模态差异能够提升标定性能。由于我们采用单目深度估计生成相机深度云，与先前的方法相比，激光雷达数据与相机数据之间的模态差异大幅缩小，使模型能够学习到更有效的跨模态相关性。此外，如图10 （e）所示，与采用单侧数据增强策略时获得的特征图相比，DB和单分支架构提取的特征图均有明显改善。这进一步验证了所提出的双侧增强方法在提升泛化能力方面的有效性。

G. 与最先进方法的比较

在本节中，我们将所提出的 DST-Calib 与最先进的无靶标 LCEC 方法进行比较，包括基于深度神经网络的方法和非学习方法。如表V[1] 和表VI所示，DST-Calib 实现了领先的精度和强大的泛化能力。值得注意的是，先前的基于深度神经网络的方法在左摄像头上表现良好，但在右摄像头上的性能却大幅下降；而 DST-Calib 在两个摄像头上都取得了高度一致的结果。在右摄像头上，它显著优于所有其他基于深度神经网络的方法。进一步分析发现，当初始失准范围相对较小（例如， $[\pm 5°, 0.5\text{m}]$）时，先前基于深度神经网络的方法在右摄像头上的平移误差约为施加扰动的一半（约为0.3–0.4米）。随着初始失准程度增大，误差也进一步上升至约0.4–0.6米（几乎等于左右摄像头之间的物理距离）。这清楚地表明，这些方法对训练期间见过的特定外参过拟合了。这些基于深度神经网络的方法在KITTI左摄像头上看似较高的精度具有误导性，因为训练后的模型甚至无法泛化到同一数据集、同一场景中的另一个相机位置。

根据 KITTI 里程计测试序列（00–09）上的全部实验结果，所提出的 DST-Calib 相较于其他最先进的无目标 LCEC 方法，性能稳定性显著更高。此外，在自监督路径下，即使初始旋转偏差高达 $30°$ 、平移偏差高达 1 m，DST-Calib PE 仍能实现令人满意的旋转标定精度。两条路径均启用时，PE+SB* 配置在很大的失准范围内，旋转和平移均能持续获得准确结果，这凸显了所提出的基于双侧数据增强的双路径标定框架的有效性。此外，SB架构的整体性能优于 DB架构。基于 SB 的模型更容易收敛，且对未见场景的泛化能力更强。这证实了所提出的单分支设计通过利用显式构建的差异图，增强了特征相关性和位姿回归，从而提高了特征可靠性和跨模态关联。此外，多帧优化显著提升了精度。例如，在序列 01 上，应用多帧优化使 SB 的旋转误差降低约 20.6%–22.7%，平移误差降低约 38.2%–46.0%。我们将这一改进归因于多帧优化所采用的评分机制：该机制通过对各帧的旋转和平移估计值取平均，有效提高了最终外参标定精度。

H. 深度锚点校正的性能

在本节中，我们评估所提出的深度细化方法 DAR 的性能。我们首先在 KITTI 数据集上，将其深度校正精度与具有代表性的单目深度估计方法进行比较。如

[1]与论文中报告的结果相比，LCCNet 的复现结果具有更高的标定误差。

表V 在 KITTI 里程计的 00 序列上与最先进的无靶标LCEC方法进行定量比较。†：这些方法未公开代码，因此无法复现左、右摄像头的结果。

方法	初始范围	左摄像头								右摄像头							
		幅值		旋转误差（°）			平移误差（m）			幅值		旋转误差（°）			平移误差（m）		
		e_r（°）	e_t（m）	Yaw	俯仰	Roll	X	Y	Z	e_r（°）	e_t(米)	Yaw	俯仰	Roll	X	Y	Z
CalibRCNN† [15]	[±10°, ±0.25m]	0.805	0.093	0.446	0.640	0.199	0.062	0.043	0.054	-	-	-	-	-	-	-	-
RegNet† [28]	[±20°, ±1.5m]	0.500	0.108	0.240	0.250	0.360	0.070	0.070	0.040	-	-	-	-	-	-	-	-
LCCNet [14]	[±10°, ±1.0米]	1.418	0.600	0.455	0.835	0.768	0.237	0.333	0.329	1.556	0.718	0.457	1.023	0.763	0.416	0.333	0.337
RGGNet [18]	[±20°, ±0.3m]	1.290	0.114	0.640	0.740	0.350	0.081	0.028	0.040	3.870	0.235	1.480	3.380	0.510	0.180	0.056	0.061
CalibDNN [16]	[±10°, ±1.0米]	1.172	0.098	0.402	0.998	0.180	0.072	0.025	0.045	1.238	0.405	0.480	1.010	0.195	0.396	0.026	0.050
CalibDepth [29]	[±10°, ±1.0米]	0.996	0.075	0.332	0.848	0.141	0.046	**0.023**	0.038	1.140	0.406	0.469	0.915	0.166	0.392	0.029	0.049
CalibNet [17]	[±10°, ±1.0米]	1.109	0.167	0.333	0.955	0.206	0.115	0.046	0.086	1.182	0.389	0.346	1.018	0.226	0.363	0.048	0.084
CRLF [21]	-	0.629	4.118	**0.033**	0.464	0.416	3.648	1.483	0.550	0.633	4.606	**0.039**	0.458	0.424	4.055	1.636	0.644
密歇根大学 [12]	-	4.161	0.321	0.113	3.111	2.138	0.286	0.077	0.086	4.285	0.329	0.108	3.277	2.088	0.290	0.085	0.090
HKU-Mars [10]	-	33.84	6.355	19.89	18.71	19.32	3.353	3.232	2.419	32.89	4.913	18.99	15.77	17.00	2.917	2.564	1.646
DVL [25]	-	122.1	5.129	48.64	87.29	98.15	2.832	2.920	1.881	120.5	4.357	49.60	87.99	96.72	2.086	2.517	1.816
MIAS-LCEC [26]	-	5.385	1.014	1.574	4.029	4.338	0.724	0.383	0.343	7.655	1.342	1.910	5.666	6.154	0.843	0.730	0.358
EdO-LCEC [39]	-	0.295	0.082	0.117	0.176	0.150	0.051	0.044	**0.032**	0.336	0.118	0.216	**0.168**	0.121	0.083	0.067	0.032
DST-Calib（DB）	2*[±5°, ±0.5m]	0.524	0.101	0.292	0.308	0.173	0.077	0.024	0.044	0.664	0.147	0.318	0.440	0.207	0.129	0.023	0.045
DST-Calib（DB）	2*[±10°, ±1.0m]	1.132	0.185	0.634	0.724	0.284	0.136	0.042	0.084	1.196	0.246	0.683	0.740	0.320	0.207	0.046	0.082
DST-Calib（SB）	2*[±5°, ±0.5m]	0.366	0.078	0.195	0.236	0.092	0.054	0.020	0.038	0.505	0.109	0.220	0.378	0.099	0.088	0.022	0.040
DST-Calib（SB）	2*[±10°, ±1.0米]	0.566	0.134	0.314	0.345	0.156	0.086	0.035	0.072	0.878	0.188	0.360	0.677	0.171	0.144	0.040	0.080
DST-Calib (SB*)	2*[±5°, ±0.5m]	**0.257**	**0.063**	0.082	**0.200**	**0.086**	**0.036**	0.026	0.034	**0.295**	**0.064**	0.134	0.215	**0.055**	**0.049**	**0.010**	0.035
DST-Calib (SB*)	2*[±10°, ±1.0m]	0.429	0.115	0.272	0.261	0.120	0.077	0.036	0.061	0.745	0.175	0.282	0.593	0.124	0.134	0.023	0.085
DST-Calib（位姿估计器）	2*[±30°, ±0.5m]	1.392	0.176	0.437	0.705	1.015	0.144	0.057	0.061	1.257	0.148	0.246	0.598	1.016	0.128	0.058	**0.028**
DST-Calib（位姿估计器+SB*）	2*[±30°, ±0.5米]	0.845	0.129	0.110	0.629	0.487	0.106	0.042	0.036	0.804	0.116	0.225	0.494	0.458	0.083	0.044	0.048

表VI 与 KITTI 里程计（01-09 序列）上的最先进 LCEC 方法进行比较。

方法	初始范围	01		02		03		04		05		06		07		09	
		e_r	e_t	e_r	e_t	e_r	e_t	e_r	e_t	e_r	e_t	e_r	e_t	e_r	e_t	e_r	e_t
CRLF [21]	-	0.623	7.363	0.632	3.642	0.845	6.007	0.601	0.372	0.616	5.961	0.615	25.76	0.606	1.807	0.626	5.133
密歇根大学 [12]	-	2.196	0.305	3.733	0.331	3.201	0.316	2.086	0.348	3.526	0.356	2.914	0.353	3.928	0.368	3.117	0.363
HKU-Mars [10]	-	20.73	3.770	32.95	12.70	21.99	3.493	4.943	0.965	34.42	6.505	25.20	7.437	33.10	7.339	20.38	3.459
DVL [25]	-	112.0	2.514	120.6	4.285	124.7	4.711	113.5	4.871	123.9	4.286	128.9	5.408	124.7	5.279	116.7	3.931
MIAS-LCEC [26]	-	0.621	0.300	0.801	0.327	1.140	0.324	0.816	0.369	4.768	0.775	2.685	0.534	11.80	1.344	0.998	0.432
EdO-LCEC [39]	-	2.269	0.459	0.561	0.142	0.737	0.137	1.104	0.339	**0.280**	0.093	0.485	0.124	**0.188**	0.076	0.386	0.120
DST-Calib（DB）	2*[±5°, ±0.5m]	0.706	0.132	0.551	0.108	0.791	0.128	0.576	0.114	0.520	0.103	0.528	0.110	0.528	0.103	0.559	0.109
DST-Calib（DB）	2*[±10°, ±1.0m]	1.343	0.262	1.198	0.201	1.358	0.237	1.263	0.208	1.108	0.191	1.121	0.210	1.152	0.191	1.207	0.202
DST-Calib（SB）	2*[±5°, ±0.5m]	0.651	0.136	0.422	0.088	0.617	0.113	0.424	0.092	0.366	0.083	0.397	0.086	0.360	0.077	0.421	0.088
DST-Calib（SB）	2*[±10°, ±1.0m]	1.126	0.263	0.652	0.159	0.849	0.194	0.671	0.183	0.610	0.146	0.611	0.152	0.558	0.135	0.657	0.160
DST-Calib (SB*)	2*[±5°, ±0.5m]	0.517	**0.084**	**0.297**	**0.051**	**0.439**	**0.075**	**0.382**	**0.068**	0.318	**0.058**	**0.305**	**0.056**	0.285	**0.063**	**0.273**	**0.055**
DST-Calib（SB*）	2*[±10°, ±1.0米]	0.870	0.142	0.440	0.115	0.501	0.137	0.552	0.162	0.510	0.134	0.525	0.123	0.476	0.097	0.548	0.143
DST-Calib（位姿估计器）	2*[±30°, ±0.5m]	1.365	0.200	1.480	0.178	2.508	0.185	2.313	0.182	1.504	0.222	0.939	0.161	1.497	0.162	1.239	0.195
DST-Calib（位姿估计器+SB*）	2*[±30°, ±0.5米]	**0.435**	0.100	0.797	0.133	1.700	0.171	0.912	0.105	0.905	0.147	0.637	0.120	0.684	0.109	0.447	0.092

表 VII：KITTI 数据集上最先进单目深度估计的定量比较。

方法	越低越好				越高越好		
	绝对相对误差	平方相对误差	均方根误差	对数均方根误差	δ_1	δ_2	δ_3
LEGO [40]	0.192	1.352	6.276	0.252	0.783	0.921	0.969
单目深度2 [37]	0.106	0.818	4.750	0.196	0.874	0.957	0.979
融合深度 [38]	**0.063**	0.364	3.291	0.139	0.945	0.978	0.988
DPA2 [30]	0.521	3.220	4.558	0.546	0.313	0.578	0.789
MoGe2 [31]	0.206	0.480	2.240	0.265	0.517	0.963	0.985
DPA2+DAR	0.087	**0.149**	**1.191**	**0.119**	**0.950**	**0.987**	**0.994**
MoGe2+DAR	0.089	0.169	1.281	0.121	0.947	0.986	**0.994**

表 VII所示，我们的方法准确度更高，显著优于 Depth Anything V2 和 MoGe2 生成的初始深度图。此外，细化后的深度结果明显优于 LEGO 等最先进的自监督单目度量深度估计方法 [40],MonoDepth2 [37],和 FusionDepth [38]。这些显著的性能提升充分证明了所提出的深度校正策略的有效性和实用价值。为进一步评估鲁棒性，我们还在 KITTI-360、Argoverse2 和 TF70 上评估了 DAR。表 VIII 中的定量结果表明，我们的方法能够持续改善大型视觉模型生成的原始深度预测，生成的相机深度图也更接近真实的激光雷达测量值。尤其是，MoGe2+DAR 达到了

表 VIII 深度校正的定量比较。

数据集	方法	平均绝对误差	均方根误差	绝对相对误差	平方相对误差	δ_1	δ_2	δ_3
KITTI-360	DPA2	5.352	5.931	0.830	6.059	0.12	0.364	0.666
	DPA2+DAR	0.786	1.353	0.100	0.247	0.908	0.976	0.991
	MoGe2	1.378	1.783	0.174	0.353	0.711	0.962	0.978
	MoGe2+DAR	**0.675**	**1.126**	**0.085**	**0.159**	**0.931**	**0.979**	**0.992**
Argoverse2	DPA2	12.69	12.71	2.569	32.60	0.000	0.001	0.008
	DPA2+DAR	0.950	1.546	**0.078**	0.232	0.928	0.981	**0.995**
	MoGe2	1.265	1.577	0.098	**0.192**	**0.961**	**0.986**	0.994
	MoGe2+DAR	**0.923**	**1.405**	0.088	0.270	0.942	0.979	0.991
nuScenes	DPA2	3.669	3.950	0.369	1.551	0.232	0.887	0.989
	DPA2+DAR	0.766	1.136	0.066	0.110	0.980	**0.998**	**1.000**
	MoGe2	1.725	2.070	0.144	0.310	0.861	0.994	0.997
	MoGe2+DAR	**0.591**	**0.908**	**0.050**	**0.066**	**0.992**	0.996	**1.000**
TF70	DPA2	6.155	6.856	1.502	16.96	0.178	0.393	0.578
	DPA2+DAR	**0.795**	**1.478**	**0.102**	**0.486**	**0.916**	**0.963**	**0.984**
	MoGe2	1.340	1.781	0.184	0.574	0.749	0.939	0.981
	MoGe2+DAR	0.983	1.565	0.134	0.590	0.898	0.958	0.980

四个数据集上的平均绝对误差为 0.793，平均均方根误差为 1.251，表明在多样化环境和传感器配置下，深度校正具有稳定性。

I. 跨域泛化

为进一步评估泛化能力，我们在另外四个公开数据集以及我们的 LCScenes 数据集上测试了所提出的 DST-Calib，涵盖总计 19 种不同的激光雷达-相机配置。由于 DB架构并非我们的主要贡献，且其性能通常劣于 SB（如表V 和表 VI所示），因此我们未将 DB 纳入跨域泛化评估。值得注意的是，DST-Calib 仅使用 KITTI 和 Argoverse 中的前视相机进行训练；其他所有相机、激光雷达和场景域在训练期间均完全未见。因此，所报告的标定结果构成零样本评估，为任何无目标 LCEC 方法实现准确的外参估计提出了极具挑战性的场景。

首先，我们在 MIAS-LCEC-TF70 和 MIAS-LCEC-TF360 上评估 DST-Calib 的跨域泛化能力。这两个数据集包含由不同类型的固态激光雷达采集的点云，其扫描视场各不相同（70° 和 360°）。由于训练过程中从未使用固态激光雷达数据，因此对其外参进行标定颇具挑战。尽管如此，表 IX中的实验结果仍超出了我们的预期。尽管其标定精度略逊于 MIAS-LCEC 和 EdO-LCEC，DST-Calib 的表现仍优于其他专为固态激光雷达生成的稠密点云设计的无目标 LCEC 方法。此外，采用自监督路径时，即使 CRLF、密歇根大学和 HKU-Mars 的评估所用的误标定范围小得多（在 1° 和 0.1 m 以内），DST-Calib 仍能取得更高的标定精度。

其次，我们在 LCScenes 上评估 DST-Calib。各序列的定量结果如图12所示。与 MIAS-LCEC 数据集不同，LCScenes 包含更多室内环境，且摄像头分辨率和传感器配置各不相同。实验结果表明，所有 DST-Calib 架构在五个序列上均保持稳定性能。除序列 01 外，SB* 的表现始终优于 SB，而 PE+SB* 的标定精度始终高于 PE，证明了多帧优化在提升精度和鲁棒性方面的有效性。此外，尽管自监督 PE 的稳定性不如全监督 SB，其平均标定误差仍与之相当。这证实了由双侧数据增强实现的自监督策略既有效又可靠。在初始标定误差范围相对较小时，全监督 DST-Calib 路径的表现优于自监督路径。在外参真值的指导下，模型能够学习点云相似度中自监督无法获取的不确定部分。

最后，我们进一步在 Argoverse2、nuScenes 和 KITTI-360 上验证 DST-Calib；这些数据集的激光雷达点云由机械旋转式激光雷达采集。值得注意的是，Argoverse2 和 nuScenes 包含朝向车辆左侧、右侧和后方的摄像头，这与训练期间使用的前向摄像头存在显著差异。我们评估了总计 15 种传感器组合，以全面考察零样本泛化能力。定量结果汇总于表X中，定性示例则展示于图13中。在这些异构数据集上，DST-Calib 始终保持较低的标定误差，并展现出对多样环境和严苛光照条件的强大适应能力。这些令人鼓舞的结果表明，该网络成功学习了输入投影与其对应外参之间的内在关系。我们认为，这一性能提升归功于所提出的单分支架构及其显式差异图构建；该架构以清晰、确定性的两种模态关联，取代了以往基于深度神经网络的方法中模糊的特征相关过程。

图。11 总结了所有数据集的整体可视化结果。使用校准后的外参参数生成的融合激光雷达-相机投影，在几何结构和纹理边缘处均呈现出清晰的对齐效果。尽管 DST-Calib 在某些室内环境中无法达到同等精度（这些环境中的点云深度范围本身有限），但与初始位姿相比，它仍显著改善了对齐效果，即使初始旋转误差较大亦是如此。

J. 计算开销与运行时间分析

我们的算法在 Intel i7-14700K CPU 和 NVIDIA RTX4090D GPU 上实现。我们评估了标定流程中各主要组件的计算开销。处理单帧所需的总时间约为 0.8 秒，其中包括深度估计、深度校正、模型推理和位姿优化。单帧处理时间应乘以用于标定的帧数，以得出多帧优化的总运行时间。如表 XI 所示，深度校正、模型推理和位姿优化阶段



图 11. DST-Calib 在多个真实世界数据集上的定性结果：(a)-(g) 分别展示了 KITTI 里程计、KITTI-360、MIAS-LCEC-TF360、MIAS-LCEC-TF70、Argoverse2、nuScenes 和 LCScenes 上的激光雷达-相机数据融合可视化结果。为便于可视化，所有激光雷达投影均使用标定后的外参叠加在各自的 RGB图像上。此外，还提供了 DAR 获取的单目深度图。

效率很高，总耗时仅约 0.09 秒。主要瓶颈在于单目深度估计阶段。使用 Depth Anything V2 或 MoGe2 等大型模型时，每帧处理时间会增加至约 0.65 秒。不过，考虑到大多数激光雷达传感器的工作频率为 10–20 Hz，整体计算开销仍足够低，适用于在线标定。

K. 消融研究与分析

1) 特征提取和块大小的影响：DST-Calib 的 DB 和 SB 都包含一个特征提取模块和一个用于位姿回归的块处理器（这不同于以往的双分支网络）。为评估这两个模块的作用，我们在 KITTI 左、右摄像头上，采用不同的特征提取骨干网络测试 DST-Calib 的 SB架构，并测试不同的块大小。表 XII表明，在不同的

表 IX 与最先进的非学习型无目标 LCEC 方法在 MIAS-LCEC 数据集上的定量比较。DST-Calib 的结果为零样本泛化结果。

方法	初始范围	MIAS-LCEC-TF70							MIAS-LCEC-TF360						
		幅值		旋转误差（°）			平移误差（米）			幅值		旋转误差（°）			平移误差（m）	
		e_r（°）	e_t（m）	Yaw	俯仰	Roll	X	Y	Z	e_r（°）	e_t（m）	Yaw	俯仰	Roll	X	Y	Z
CRLF [21]	-	1.683	11.13	0.197	1.625	0.946	10.05	3.036	2.589	1.463	7.418	0.135	1.421	0.317	7.093	0.621	1.885
密歇根大学 [12]	-	4.265	0.333	0.485	1.945	4.272	0.217	0.134	0.115	4.434	0.249	0.324	2.340	3.356	0.137	0.129	0.110
HKU-Mars [10]	-	3.941	1.261	2.140	2.156	1.988	0.806	0.555	0.475	56.41	8.042	33.45	35.59	7.843	3.646	5.931	2.227
DVL [25]	-	0.423	0.100	0.201	0.292	0.104	0.075	0.050	0.026	54.70	1.325	23.98	43.27	34.73	0.335	0.769	0.892
MIAS-LCEC [26]	-	0.298	0.061	0.133	0.196	0.110	**0.040**	0.028	0.019	0.799	0.142	0.349	0.535	0.277	0.095	0.075	0.050
EdO-LCEC [39]	-	**0.255**	**0.055**	**0.117**	**0.166**	**0.096**	**0.040**	**0.024**	**0.016**	**0.504**	**0.107**	**0.229**	0.341	**0.112**	**0.064**	0.066	**0.032**
DST-Calib（SB）	2*[±5°, ±0.5m]	1.200	0.134	0.920	0.424	0.337	0.095	0.049	0.049	1.232	0.134	0.953	0.364	0.392	0.094	0.056	0.047
DST-Calib (SB*)	2*[±5°, ±0.5m]	0.841	0.143	0.518	0.382	0.371	0.113	0.054	0.044	0.667	0.129	0.419	**0.288**	0.274	0.079	0.062	0.058
DST-Calib（位姿估计器）	2*[±30°, ±0.5m]	1.903	0.181	0.513	0.882	1.453	0.145	0.054	0.053	2.303	0.152	2.068	0.510	0.684	0.117	**0.049**	0.053
DST-Calib（位姿估计器+SB*）	2*[±30°, ±0.5m]	1.423	0.149	0.549	0.654	1.021	0.100	0.060	0.071	1.871	0.165	1.656	0.662	0.300	0.122	0.080	0.053

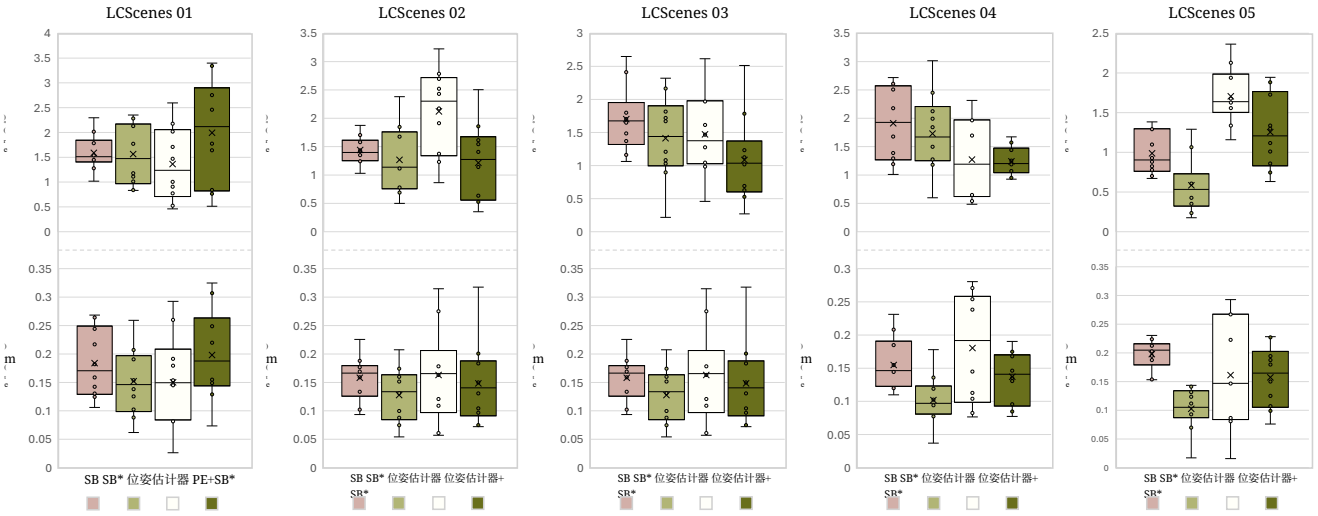

图 12。 定量标定结果 在我们创建的 LCScenes 数据集上的结果。所有架构的初始失准范围均设为 2*[5°, 0.5m] .

表 X 不同传感器配置下三个公开数据集上的零样本泛化结果。

方法	初始范围	KITTI-360 SFL + SFR		Argoverse2 SFL +SFR + RFR		Argoverse2 RSL +RSR		Argoverse2 RBL +RBR		nuScenes RFC + RFL +RFR		nuScenes RBC + RBL + RBR	
		e_r（°）	e_t（米）	e_r（°）	e_t（米）	e_r（°）	e_t（米）	e_r（°）	e_t（米）	e_r（°）	e_t（米）	e_r（°）	e_t（米）
DST-Calib（SB）	2*[±5°, ±0.5米]	0.898	0.137	0.938	0.142	1.336	0.154	0.901	**0.141**	1.332	0.174	1.333	0.165
DST-Calib（SB*）	2*[±5°, ±0.5米]	**0.709**	**0.113**	**0.803**	**0.129**	0.892	**0.151**	0.733	0.148	**0.880**	**0.135**	**0.954**	**0.096**
DST-Calib（SB）	2*[±15°, ±0.5m]	3.980	0.169	3.940	0.187	4.503	0.180	3.965	0.187	4.281	0.188	4.340	0.187
DST-Calib（位姿估计器）	2*[±5°, ±0.5m]	2.173	0.208	1.579	0.164	**0.807**	0.208	1.206	0.179	1.144	0.192	1.161	0.181
DST-Calib（位姿估计器）	2*[±15°, ±0.5m]	2.014	0.169	1.070	0.163	1.099	0.164	1.109	0.205	1.399	0.172	1.599	0.184
DST-Calib（位姿估计器）	2*[±30°, ±0.5米]	2.105	0.153	1.441	0.171	1.421	0.202	1.083	0.174	1.383	0.160	1.266	0.178
DST-Calib（位姿估计器+SB*）	2*[±5°, ±0.5m]	1.804	0.154	1.049	0.189	0.956	0.180	0.885	0.161	1.289	0.148	0.992	0.138
DST-Calib（位姿估计器+SB*）	2*[±30°, ±0.5m]	1.824	0.136	1.230	0.179	0.831	0.166	**0.688**	0.151	1.155	0.142	1.112	0.161

特征提取骨干网络和不同网格尺寸下，均可获得相近的标定精度。这进一步证明，对于直接回归外参的基于深度神经网络的 LCEC 网络而言，具体的网络模块并非必不可少。真正重要的是训练策略和整体的特征相关性架构。正是 DST-Calib 的双侧数据增强和单分支架构

Calib 能显著提升 LCEC 网络的泛化能力，但并非针对特定的特征提取模块或下游位姿回归层。

2) 位姿估计器不同架构的性能比较：我们比较了位姿估计器的简单架构和标准架构（见图7），并进一步考察

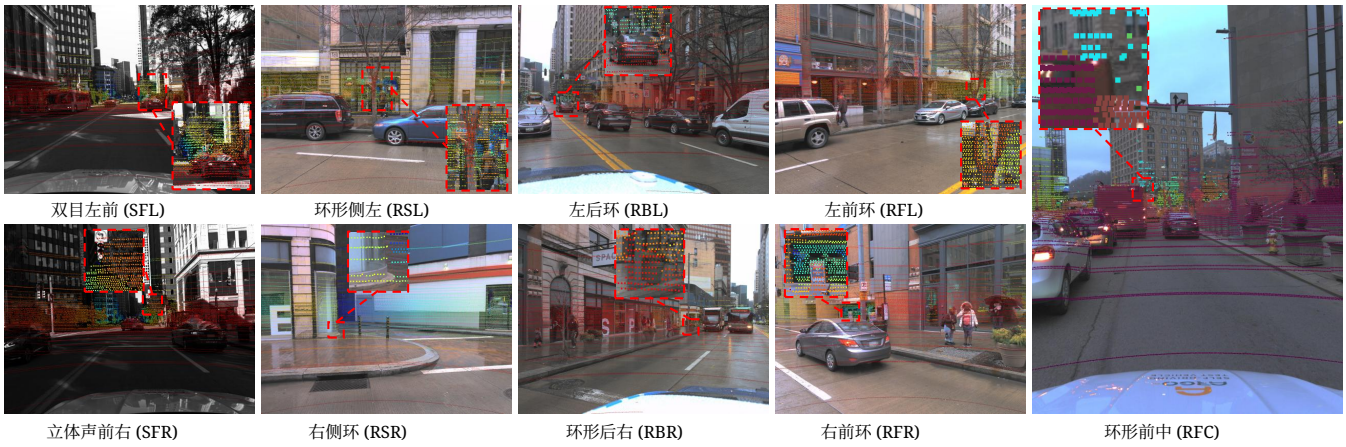

图 13. Argoverse2 数据集所有摄像头上的数据融合结果可视化。使用 DST-Calib 标定的外参，将激光雷达点云投影到各自对应的 RGB图像上。

表 XI DST-Calib 的计算开销和运行时间分析。

方法	深度估计（秒/帧）	深度校正（秒/帧）	模型推理（秒/帧）	Pose 优化（秒/帧）	总计（帧/秒）
DST-Calib（DB）	0.661	0.052	0.009	0.066	0.871
DST-Calib（SB）	0.598	0.053	0.007	0.030	0.772
DST-Calib（PE）	0.640	0.050	0.007	0.027	0.808
DST-Calib（PE+SB*）	0.634	0.051	0.013	0.055	0.869

表XII 不同块大小下不同特征提取骨干网络的定量比较（采用我们提出的 DST-Calib 单分支架构）。

骨干网络	块大小	KITTI Left00 e_r (°)	KITTI Left00 e_t (m)	KITTI Right00 e_r (°)	KITTI Right00 e_t (米)
ResNet 18	1	**0.282**	0.068	**0.488**	0.111
	3	0.351	0.075	0.576	0.123
	5	0.366	0.078	0.505	**0.109**
ResNet 34	1	0.316	0.076	0.544	**0.109**
	3	0.387	0.090	0.524	0.116
	5	0.402	0.089	0.508	0.117
EfficientNet	1	0.386	0.079	0.514	0.120
	3	0.310	0.064	0.544	0.112
	5	0.311	**0.062**	0.591	0.137

自监督路径中各损失分量的作用。如表XIII所示，在所有损失项组合下，两种架构在 KITTI 左、右摄像头上的表现相近，而标准架构（使用 LDP 和 CDPas 输入）的整体精度略高。在所有损失项中，倒角距离最为关键，能够显著降低旋转误差。与初始位姿约束结合时，位姿估计器可以实现良好的仅旋转标定，同时保留所提供外参初始猜测的初始平移。加入 $\mathcal{L}'_{eva}$ 项后，与另外两种损失结合使用时，精度还会进一步提高。

表 XIII 位姿估计器不同架构的性能比较。

架构	$\mathcal{L}_{CD}$	$\mathcal{L}_{t_{ini}}$	$\mathcal{L}'_{eva}$	KITTI Left00 e_r (°)	KITTI Left00 e_t(米)	KITTI Right00 e_r (°)	KITTI Right00 e_t (米)
PE（简单）	✓			1.181	0.581	1.466	0.669
	✓	✓		0.897	0.162	1.197	0.205
	✓		✓	1.333	0.600	1.438	0.694
	✓	✓	✓	0.919	0.182	1.189	0.203
PE（标准）	✓			1.398	0.608	1.672	0.691
	✓	✓		1.187	0.208	1.132	0.191
	✓		✓	**0.849**	0.388	1.260	0.658
	✓	✓	✓	0.890	**0.135**	**1.086**	**0.181**

由于位姿估计器假设整个序列中的外参始终保持一致，因此在实际应用中，简单架构只需较少的模型参数，就足以实现良好的仅旋转标定。位姿估计器与评估模块（DST-Calib PE+SB*）的组合是实现稳健、准确的六自由度外参估计的最佳选择。

3）对不同旋转失准范围的适应性：在这项消融研究中，我们评估了 DST-Calib 对不同旋转失准范围的适应性。我们将外参初始化为不同程度的旋转扰动，从较小的偏差 1° 一直到较大的 30°，并测量由此得到的标定精度，以评估各架构的鲁棒性。如图 14所示，结果符合预期。对于 SB架构，当初始偏差超出训练时见过的范围后，平均旋转误差便会增大；尽管多帧优化有助于降低训练范围内的误差，但在初始失准过大时仍难以奏效。相比之下，启用自监督路径后，PE 和 PE+SB* 在所有失准程度下都能保持稳定、一致的性能。这表明在线自监督学习显著提升了模型对较大旋转偏差的鲁棒性。此外，PE+SB*

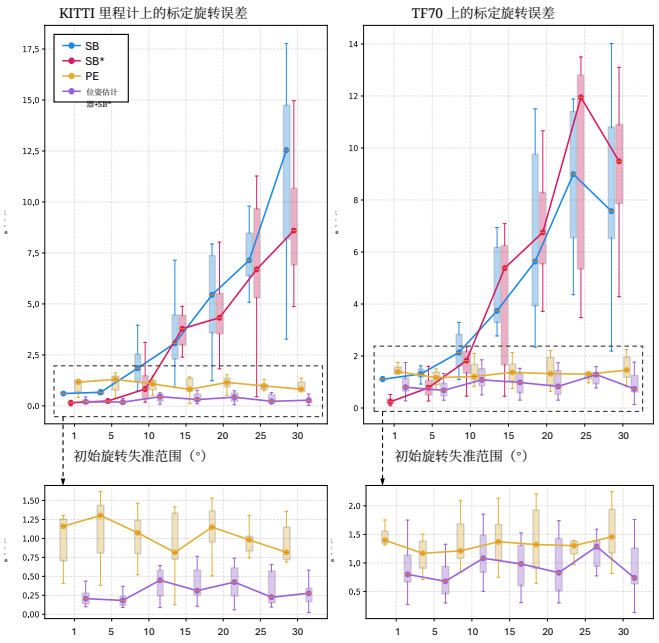

图14. 初始失准范围从 1 到 30 度时的旋转标定误差。

始终优于单独使用 PE，表明结合两条路径可在不同失准范围下实现仅旋转标定的最高精度和最可靠性能。

V. 结论

本文深入探讨了基于深度神经网络的无靶标激光雷达-相机外参标定网络。我们发现，先前方法采用的单侧数据增强策略会导致泛化能力下降，并提出一种新颖的双侧数据增强策略加以解决。与先前方法不同，我们的方法不再受限于特定的传感器配置，能够泛化至多种环境和各种传感器对。为进一步提升鲁棒性和准确性，我们提出了一种双路径标定框架，整合了全监督和自监督路径，实现完全无靶标标定，并大幅降低训练过程中对预标定外参真值的依赖。在多个公开真实世界数据集以及我们新采集的数据集上进行的大量评估表明，我们的方法达到了当前最先进的性能。

参考文献

[1] G. Pagnanelli 等，“通过仿生光学触觉传感器，融合类人阻抗调节与基于模型的方法进行柔顺性辨别”，*IEEE Transactions on Robotics*，第 41 卷，第 857–870 页，2025 年。[2] D. Cattaneo 和 A. Valada，“CMRNext：面向真实环境中定位与外参标定的相机到激光雷达匹配”，*IEEE Transactions on Robotics*，第 41 卷，第 1995–2013 页，2025 年。[3] C. Zheng 等，“FAST-LIVO2：快速直接的激光雷达-惯性-视觉里程计”，*IEEE Transactions on Robotics*，第 41 卷，第 326–346 页，2025 年。

[4] Y. Li等人，“DeepFusion：面向多模态三维目标检测的激光雷达-相机深度融合”，载于IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，2022年，第17182–17191页。[5] Y. Tian等人，“Kimera-multi：面向多机器人系统的鲁棒、分布式、稠密度量-语义同步定位与建图”，《IEEE机器人学汇刊》，第38卷，第4期，2022年。[6] A. Zhang等人，“迈向城市场境中鲁棒的机器人三维感知：UT校园目标数据集”，《IEEE机器人学汇刊》，第40卷，第3322–3340页，2024年。[7] L. Luo等人，“BEVPlace++：面向自主地面车辆的快速、鲁棒且轻量的激光雷达全局定位”，《IEEE机器人学汇刊》，第41卷，第4479–4498页，2025年。[8] J. Cui等人，“ACSC：非重复扫描固态激光雷达与相机系统的自动标定”，arXiv预印本 arXiv:2011.08516，2020年。[9] G. Koo等人，“面向相机与激光雷达精确基于平面性外参标定的解析平面协方差构建”，载于2020年IEEE国际机器人与自动化会议（ICRA）。IEEE，2020年，第6042–6048页。[10] C. Yuan等人，“无目标环境中高分辨率激光雷达与相机的像素级外参自标定”，《IEEE机器人与自动化快报》，第6卷，第4期，第7517–7524页，2021年。[11] Y. Zhu等人，“利用传感器语义信息的相机-激光雷达在线标定”，载于2020年IEEE国际机器人与自动化会议（ICRA）。IEEE，2020年，第4970–4976页。[12] G. Pandey等人，“通过最大化互信息实现视觉与激光雷达的自动外参标定”，《野外机器人学杂志》，第32卷，第5期，第696–722页，2015年。[13] J. Borer等人，“从混沌到标定：一种面向无目标相机-激光雷达外参标定的几何互信息方法”，载于IEEE/CVF计算机视觉应用冬季会议（WACV）论文集，2024年，第8409–8418页。[14] X. Lv等人，“LCCNet：基于代价体网络的激光雷达与相机自标定”，载于IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，2021年，第2894–2901页。[15] J. Shi等人，“CalibRCNN：基于循环卷积神经网络与几何约束的相机和激光雷达标定”，载于2020年IEEE/RSJ智能机器人与系统国际会议（IROS）。IEEE，2020年，第10197–10202页。[16] G. Zhao等人，“CalibDNN：利用深度神经网络进行感知的多模态传感器标定”，载于《信号处理、传感器/信息融合与目标识别XXX》，第11756卷。SPIE，2021年，第324–335页。[17] G. Iyer等人，“CalibNet：利用三维空间变换网络进行几何监督的外参标定”，载于2018年IEEE/RSJ智能机器人与系统国际会议（IROS）。IEEE，2018年，第1110–1117页。[18] K. Yuan等人，“RGGNet：基于几何深度学习与生成模型的容差感知激光雷达-相机在线标定”，《IEEE机器人与自动化快报》，第5卷，第4期，第6956–6963页，2020年。[19] F. Lv和K. Ren，“基于线和点特征的机载激光雷达点云数据与光学影像深度图自动配准”，《红外物理与技术》，第71卷，第457–463页，2015年。[20] J. Li等人，“利用语义特征的全景图像序列与移动激光扫描数据自动配准”，《ISPRS摄影测量与遥感杂志》，第136卷，第41–57页，2018年。[21] T. Ma等人，“CRLF：面向道路场景中激光雷达与相机的基于线特征的自动标定与精化”，arXiv预印本 arXiv:2103.04558，2021年。[22] Y. Wang等人，“基于杆匹配的城市场景点云与全景图像自动配准”，《国际应用地球观测与地理信息杂志》，第115卷，103083，2022年。[23] Y. Han等人，“面向城市自动驾驶应用的基于停车标志的自动标定方法”，载于2021年IEEE国际机器人与自动化会议（ICRA）。IEEE，2021年，第13179–13185页。[24] Y. Liao等人，“SE-Calib：城市场景中基于语义边缘的激光雷达-相机视轴在线标定”，《IEEE地球科学与遥感汇刊》，2023年，DOI: 10.1109/TGRS.2023.3278024。[25] K. Koide等人，“通用、单次、无目标且自动的激光雷达-相机外参标定工具箱”，载于2023年IEEE国际机器人与自动化会议（ICRA），2023年，第11301–11307页。[26] Z. Huang等人，“基于跨模态掩码匹配的在线无目标激光雷达-相机外参标定”，《IEEE智能车辆汇刊》，第1–12页，2024年。

[27] P.-E. Sarlin等人，“SuperGlue：利用图神经网络学习特征匹配”，载于IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，2020年，第4938–4947页。[28] N. Schneider等人，“RegNet：利用深度神经网络的多模态传感器配准”，载于2017年IEEE智能车辆研讨会（IV）。IEEE，2017年，第1803–1810页。[29] J. Zhu等人，“CalibDepth：统一深度图表示以实现迭代式激光雷达-相机在线标定”，载于2023年IEEE国际机器人与自动化会议（ICRA）。IEEE，2023年，第726–733页。[30] L. Yang等人，“Depth Anything V2”，arXiv:2406.09414，2024年。[31] R. Wang等人，“Moge-2：具有度量尺度与清晰细节的精确单目几何”，arXiv预印本 arXiv:2507.02546，2025年。[32] A. Geiger等人，“我们准备好迎接自动驾驶了吗？KITTI视觉基准套件”，载于2012年IEEE计算机视觉与模式识别会议（CVPR）。IEEE，2012年，第3354–3361页。[33] Y. Liao、J. Xie和A. Geiger，“KITTI-360：用于二维与三维城市场景理解的新型数据集与基准”，《IEEE模式分析与机器智能汇刊》，第45卷，第3期，第3292–3310页，2022年。[34] H. Caesar等人，“nuScenes：面向自动驾驶的多模态数据集”，载于IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，2020年，第11621–11631页。[35] B. Wilson等人，“Argoverse 2：面向自动驾驶感知与预测的下一代数据集”，arXiv预印本 arXiv:2301.00493，2023年。[36] D. Eigen等人，“利用多尺度深度网络从单幅图像进行深度图预测”，《神经信息处理系统进展》（NeurIPS），第27卷，2014年。[37] C. Godard等人，“深入探究自监督单目深度估计”，载于IEEE/CVF国际计算机视觉会议（ICCV）论文集，2019年，第3828–3838页。[38] Z. Feng等人，“利用稀疏激光雷达推进自监督单目深度学习”，载于机器人学习会议（CoRL）。PMLR，2022年，第685–694页。[39] Z. Huang等人，“环境驱动的激光雷达-相机在线外参标定”，《IEEE自动化科学与工程汇刊》，第22卷，第24164–24176页，2025年。[40] Z. Yang等人，“LEGO：通过观看视频同时学习边缘与几何”，载于IEEE计算机视觉与模式识别会议（CVPR）论文集，2018年，第225–234页。

DST-Calib：双路径、自监督、无目标的激光雷达-相机外参标定网络

补充材料

VI. DAR 详情

如第 III-B 节所述，在 DAR 的深度锚点选择中，我们采用一种两阶段流程，该流程具有较强的抗离群值能力，同时鼓励近似线性，并保持对 d^C 的覆盖。伪代码见算法 1[2]。具体而言，第一阶段执行 B 少量最小二乘拟合，并对分箱进行一次遍历，计算复杂度为 $O(N+B)$。第二阶段对 $M \le B$ 个候选点执行二次动态规划（DP）优化，时间复杂度为 $O(M^2)$，内存复杂度为 $O(M)$。在实践中，$M \approx 2T$，因此总体计算开销仍然较小。

第一阶段：覆盖感知候选点稀疏化。将x域划分为 $B = 2T$ 个相等的分箱，其中 T 为目标锚点数。对于每个非空分箱 $[b_j, b_{j+1}]$，我们至多选择一个代表点（$\hat{x}, \hat{y}$），使其与该分箱内的局部线性趋势最吻合，例如通过最小化相对于最小二乘直线拟合的分箱残差：

$$(\hat{x},\hat{y}) \in \arg\min_{(x,y)\in\tilde{\mathcal{A}}\cap[b_j,b_{j+1}]} \big|\, y-(\alpha_j x+\beta_j)\,\big|, \qquad (25)$$

其中（α_j，β_j）是对分箱限制点的最小二乘解。所得候选集 $\mathcal{C} = \{(x_i^c, y_i^c)\}_{i=1}^{M}$，$M \le B$ 按 x排序，并且在深度方向上大致均匀分布，这有助于后续生成一条“长”的近线性链。

第二阶段：使用动态规划求最长单调凸子序列。令 $\mathcal{C}$ 按递增的 x 编号。对于每个 i，我们存储以（x_i^c，y_i^c）结尾的最佳有效子序列 DP[i] 。对于 $i = 1, \ldots\ldots, M$:

$$\mathrm{DP}[i] \leftarrow \{(x_i^c, y_i^c)\}; \quad \text{for } j < i \text{ with } y_i^c \ge y_j^c:$$

$$\text{let} \quad s(i,j) = \frac{y_i^c - y_j^c}{x_i^c - x_j^c}.$$

如果 $|\mathrm{DP}[j]| \le 1$，则可以追加（x_i^c，y_i^c）。否则，令（x_a^c，y_a^c）为最后一个点，令（x_b^c，y_b^c）为 DP[j] 的倒数第二个点，并计算最后一个斜率

$$s_{\text{last}} = \frac{y_a^c - y_b^c}{x_a^c - x_b^c}.$$

如果 s（i，j）$\ge s_{\text{last}}$（离散凸性），则更新

$$\mathrm{DP}[i] \leftarrow \arg\max\big\{\, \mathrm{DP}[i],\ \mathrm{DP}[j] \cup \{(x_i^c, y_i^c)\} \,\big\}$$

按基数排序。最后，返回最长链 $\mathcal{S}^\star$ =arg max$_i$ |DP[i]|。如果 $|\mathcal{S}^\star| > T$，则在 x 中对其进行均匀子采样，同时保留端点，以获得恰好 T 个锚点。

[2]由于该算法是一个统一框架，也可用于其他任务，因此我们分别使用 x 和 y 替代 d^C 和 d^L。

算法1 近线性覆盖的两阶段锚点选择

要求： 数据集 $\tilde{\mathcal{A}} = \{(x_i, y_i)\}_{i=1}^{N}$，目标锚点数量 T**确保：** 选定的锚点 $\mathcal{S}$，其中包含 $|\mathcal{S}| = T$

阶段 I：覆盖感知候选点稀疏化

将 x 域划分为 $B = 2T$ 个相等的分箱
初始化候选集 $\mathcal{C} \leftarrow \emptyset$
对于 每个非空分箱 $[b_j, b_{j+1}]$ **执行**
    计算最小二乘直线 (α_j, β_j)，基于 $\tilde{\mathcal{A}} \cap [b_j, b_{j+1}]$
    选择 $(\hat{x}, \hat{y}) \in$ 参数 $\min_{(x,\ y)\in\tilde{\mathcal{A}}\cap[b_j,b_{j+1}]} |y - (\alpha_j x + \beta_j)|$
    $\mathcal{C} \leftarrow \mathcal{C} \cup \{(\hat{x},\ \hat{y})\}$
结束循环
对 $\mathcal{C} = \{(x_i^c, y_i^c)\}_{i=1}^{M}$排序 按 x （递增）

第二阶段：最长单调凸子序列

初始化动态规划数组：动态规划[i] $\leftarrow \{(x_i^c, y_i^c)\}$，其中 $i = 1, \ldots, M$
对于 $i = 1$ 到 M **执行**
    对于 每个满足 $y_i^c \ge y_j^c$ 的 $j < i$ **执行**
        计算斜率 $s(i,j) \leftarrow \dfrac{y_i^c - y_j^c}{x_i^c - x_j^c}$
        如果 |动态规划[j]| ≤ 1 **则**
            候选 $\leftarrow$动态规划[j] $\cup \{(x_i^c, y_i^c)\}$
        else
            令 (x_a^c, y_a^c) = 动态规划的最后一个点[j]
            令 (x_b^c, y_b^c) = 动态规划的倒数第二个点[j]
            $s_{\text{last}} \leftarrow \dfrac{y_a^c - y_b^c}{x_a^c - x_b^c}$
            如果 $s(i,j) \ge s_{\text{last}}$ **则** ▷ 离散凸性条件
                Candidate $\leftarrow$ DP[j] $\cup \{(x_i^c, y_i^c)\}$
            end if
            结束条件判断
        如果 候选项存在且 |候选项| > |动态规划[i]| **则**
            动态规划[i] $\leftarrow$候选项
        结束条件判断
    结束循环
结束循环
$\mathcal{S}^\star \leftarrow$ 求最大值$_i$ |动态规划[i]|

后处理
如果 $|\mathcal{S}^\star| > T$ **则**
    在 x 中均匀子采样 $\mathcal{S}^\star$（保留端点）以获得恰好 T 个锚点
结束 if
返回 $\mathcal{S}^\star$

VII. 相对位姿的平均计算

如第 III-E 节所述，本文采用评分机制，在多帧优化中对外参进行加权平均。本节将

详细介绍这一加权平均过程。

对于平移，平移向量 $\boldsymbol{t}_i$ of $\boldsymbol{T}_i$ 的加权平均值计算如下：

$$\boldsymbol{t}^* = \sum_{j=1}^{k} w_j \cdot \boldsymbol{t}_{\pi(j)}. \tag{26}$$

与平移不同，不能通过简单地对矩阵取平均来计算旋转矩阵的加权平均值。我们使用基于特征值分解的加权四元数平均来计算平均旋转矩阵。首先，可以将 $\boldsymbol{T}_i$ 中的每个旋转分量 $\boldsymbol{R}_i$ 转换为四元数向量 $\boldsymbol{q}_i$。通过对归一化权重开平方来构造加权四元数：

$$\tilde{\boldsymbol{q}}_i = \sqrt{\hat{w}_i} \cdot \boldsymbol{q}_i, \quad i = 1, 2, \ldots, n. \tag{27}$$

通过拼接所有加权四元数 $\boldsymbol{Q} = [\tilde{\boldsymbol{q}}_1, \tilde{\boldsymbol{q}}_2, \ldots, \tilde{\boldsymbol{q}}_n]$，构成一个 $4 \times n$ 四元数矩阵。协方差矩阵计算如下：

$$\boldsymbol{C} = \boldsymbol{Q}\boldsymbol{Q}^\top. \tag{28}$$

对$\mathbf{C}$ 进行特征值分解，得到：

$$\boldsymbol{C}\boldsymbol{e}_j = \lambda_j \boldsymbol{e}_j, \quad j = 1, 2, 3, 4, \tag{29}$$

其中，$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4$ 为特征值，$\boldsymbol{e}_j$ 为对应的特征向量。最后，平均四元数对应于与最大特征值 $\bar{\boldsymbol{q}} = \boldsymbol{e}_1$ 相关联的特征向量。对平均四元数进行归一化，以确保其长度为单位长度：

$$\bar{\boldsymbol{q}} \leftarrow \frac{\bar{\boldsymbol{q}}}{\|\bar{\boldsymbol{q}}\|}. \tag{30}$$

通过以下方式保持符号一致性：

$$\bar{\boldsymbol{q}} \leftarrow \begin{cases} \bar{\boldsymbol{q}}, & \text{if } \bar{\boldsymbol{q}}_0 \geq 0 \\ -\bar{\boldsymbol{q}}, & \text{otherwise.} \end{cases} \tag{31}$$

$\bar{\boldsymbol{q}}$ 对应的旋转矩阵是最终旋转矩阵 $\boldsymbol{R}^*$。这种多帧优化方法可以成功融合相同传感器配置下多个场景的标定结果。加权平均计算过程使最终估计的外参更接近真实值，从而提升整体标定精度。

VIII. 数据集详情

我们在 KITTI 里程计、KITTI360、Argoverse2、MIAS-LCEC-TF70、MIAS-LCEC-TF360、nuScenes 以及我们创建的 LCScenes 上进行了广泛实验。现介绍这些数据集的详情。

1) KITTI 里程计：KITTI 里程计是一个大规模公开数据集，由一辆配备两台 Point Grey Flea 2 RGB 相机和一台 Velodyne HDL-64E 激光雷达的车辆采集。传感器数据以 10 Hz 的频率采集。评估使用前 10 个序列（00-09）。具体而言，使用 08 序列的数据进行训练，其余序列用于测试。

2) KITTI-360：KITTI-360 是一个大规模自动驾驶数据集，旨在推动语义场景理解、三维目标检测和同步定位与建图领域的研究。我们使用前向立体相机拍摄的图像，以及 Velodyne HDL-64E 激光雷达采集的点云。使用 00 序列中前 3000 对激光雷达点云和相机图像，评估 DST-Calib 的零样本泛化性能。

3) Argoverse2：Argoverse2 是一个大规模多模态数据集，旨在推动自动驾驶领域的感知与预测研究。该数据集提供了在美国多个城市环境中采集的高质量传感器数据。采集车辆配备了两台 Velodyne VLP-32 激光雷达、七个环视 RGB 相机（1920×1200帧率为 30 FPS），以及两个前向立体相机（2056×2464帧率为 5 FPS）。由于该数据集包含多种不同的激光雷达-相机传感器配置，因此非常适合用于验证 LCEC 方法的泛化能力。在跨域泛化实验中，前方中央视角和前方左侧视角的环形相机用于 DST-Calib 的训练。

4）*MIAS-LCEC-TF70*：MIAS-LCEC-TF70是一个具有挑战性的多模态数据集，包含60组配对的四维点云和RGB图像。点云记录了三维空间坐标和强度值。数据采集使用Livox Mid-70激光雷达和MindVision SUA202GC相机进行。该数据集涵盖广泛的室内外场景，采集条件多样，包括不同的环境、天气状况和光照变化，因此整体复杂度和多样性更高。

5）*MIAS-LCEC-TF360*：MIAS-LCEC-TF360数据集包含12组成对的四维点云和RGB图像，这些数据是在室内外场景中使用Livox Mid-360激光雷达和MindVision SUA202GC相机采集的。由于Mid-360传感器具备全 360° 扫描能力，其生成的点云比Mid-70激光雷达采集的点云稀疏得多。此外，激光雷达与相机之间显著的视场不匹配，导致两种模态之间的重叠相对有限。这些特点使MIAS-LCEC-TF360特别适用于评估稀疏数据和跨模态重叠最小的场景下的算法性能。

6）*nuScenes*：nuScenes 使用的激光雷达是 32线传感器HDL-32E，其点云明显比 KITTI 和 KITTI-360 所采用的 64 线激光雷达更稀疏，因此标定任务更具挑战性。我们使用官方 nuScenes SDK，从 150 个测试场景中生成图像-点云对，以构建评估集。

7）*LCScenes*：LCScenes 是使用我们的定制设备采集的数据集。激光雷达和相机安装在放置于三脚架上的可调平台上。该数据集包含大量具有不同外参的四维点云和二维相机图像对，采集于不同的室内外场景中。激光雷达传感器为 Livox-Mid70。相机传感器的分辨率为 1200×800 和 2400×1200。该数据集分为五个序列，每个序列均沿一条特定路线采集。