arXiv:2603.01839v1 [cs.CV] 2 Mar 2026

LEAR：用于事件到激光雷达定位的边缘感知表示学习

Kuangyi Chen Jun Zhang Yuxi Hu YiZhou Friedrich Fraundorfer

摘要——**事件相机能够以高时间分辨率进行感知，即使在高速运动和光照条件严苛的情况下也能保持可靠，因此有望用于全球定位系统拒止和视觉质量下降环境中的激光雷达点云定位。然而，将稀疏异步事件与稠密激光雷达地图对齐在本质上是一个不适定问题，因为直接进行对应关系估计会受到模态差异的影响。我们提出 LEAR，这是一种双任务学习框架，可联合估计边缘结构和稠密事件—深度流场，以弥合感知模态之间的差异。LEAR 不将边缘仅仅视为事后辅助，而是通过跨模态融合机制和迭代细化策略将边缘与光流估计相结合：前者将模态不变的几何线索注入运动表示，后者则在多个更新步骤中确保两个任务互相一致。这种协同作用生成边缘感知、深度对齐的流场，从而通过透视-n点（PnP）求解器实现更稳健、更准确的位姿恢复。在多个广受认可且具有挑战性的数据集上，LEAR 的表现优于此前的最佳方法。源代码、训练好的模型和演示视频现已在网上公开**[1]。

I. 引言

视觉定位旨在估计相机在参考地图中的六自由度（6-DoF）姿态。由于相机轻便且成本低廉，这是一种前景广阔且应用广泛的方法，尤其适合部署在无人机（UAV）等紧凑型机器人平台上。然而，标准相机容易受到运动模糊和光照变化的影响。事件相机是一类新型仿生传感器，具备高动态范围、低延迟和异步运行等显著优势，因此能够有效应对这些挑战。因此，事件相机作为标准相机的替代方案，在视觉定位领域日益受到关注。

基于事件的定位面临的一项核心挑战，是在事件与参考地图之间建立对应关系；二者的信号或数据表示存在显著差异：事件通常稀疏但包含纹理信息，而点云通常稠密且具有丰富的几何信息。传统方法通常依赖特定任务参考地图，例如带地理标记的事件

Kuangyi Chen、Jun Zhang、Yuxi Hu 和 Friedrich Fraundorfer 就职于奥地利格拉茨，邮编 8010，格拉茨工业大学视觉计算研究所。电子邮箱：{kuangyi.chen, jun.zhang,yuxi.hu, friedrich.fraundorfer}@tugraz.atYi Zhou 就职于中国长沙 410012，湖南大学人工智能与机器人学院神经形态自动化与智能实验室（NAIL）。电子邮箱：eeyzhou@hnu.edu.cn[1]

https://github.com/EasonChen99/LEAR

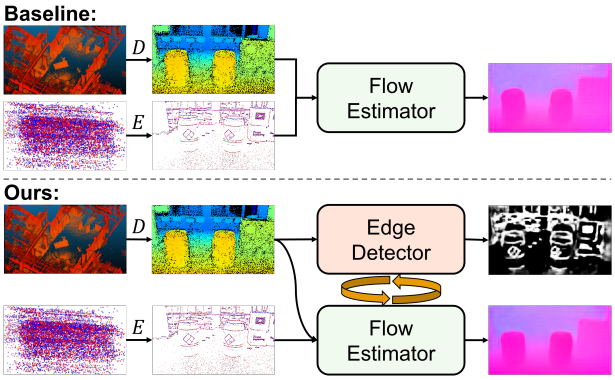


图1. 我们的方法（**下方**）将光流估计器与边缘检测器整合到一个相互促进的循环中，与仅使用光流的基线方法*EVLoc*[19]（**上方**）相比，可实现更准确的二维—三维匹配（即事件—深度光流）。

帧 [1]–[3]，手工特征数据库 [4]，[5]，或学习型描述子 [6]–[8]。虽然这些方法在地点识别方面效果良好，但通常只能实现较粗略的定位，而且往往需要投入大量精力来构建和维护。为实现更精确的定位，其他方法会构建三维线地图 [9]或半稠密重建 [10]–[12]，将事件与投影后的地图边缘对齐。然而，这些方法在存在噪声、遮挡或纹理较弱的情况下性能会受到影响。基于光度地图的方法 [13], [14] 试图将预测的强度变化与观测到的事件进行匹配，但缺乏视角相关外观建模，因此在动态或复杂环境中表现不佳。直接位姿回归方法[15]–[18]通过在网络参数中隐式编码场景几何，提供了高效的替代方案。然而，这类方法对新环境的泛化能力有限。

与这些特定任务的表示方式不同，激光雷达点云作为参考地图提供了一种颇具吸引力的替代方案：它们具有精确的几何信息、对光照变化稳健，并且通常可从第三方提供商处轻松获取。然而，由于事件与激光雷达本质上存在差异，直接将二者对齐仍然颇具挑战。据我们所知，EVLoc [19]，首次尝试弥合这一差距，将该问题建模为二维空间中事件与投影后的激光雷达点之间的稠密光流估计。随后利用光流估计值确定用于相机位姿恢复的2D-3D对应关系。尽管该方法展现了令人鼓舞的结果，但其运作很大程度上仍像一个黑箱，对底层机制的可解释性有限。

我们的核心洞见是，边缘几何构成了事件与激光雷达之间不受模态影响的桥梁：事件会

在强度梯度处触发，而激光雷达中的深度不连续性则捕捉结构边界。将这种共享结构融入表征，可以减轻跨模态特征编码的负担，并使对应关系估计更加稳健。在我们的实验中，使用事件衍生边缘图对激光雷达点云进行掩蔽（如图4所示，真实边缘图），与仅使用原始点云相比，性能显著提升；表 I 中的*Oracle* 设置也证明了这一点。然而，生成这种对齐的边缘掩码需要知道相对位姿，也就是我们要估计的相关对应关系。这凸显了一种相互依赖关系：准确的边缘提取依赖于已知的对应关系，而高质量的对应关系则需要对齐良好的边缘感知编码特征。

受这一观察的启发，我们提出了 LEAR，如图1所示。这是一种利用激光雷达地图进行基于事件的定位的双任务框架。给定一个初始位姿估计（通过视觉地点识别等粗略的全局定位方法获得），LEAR 会将其细化为精确的六自由度相机位姿。LEAR 将稠密的事件—深度光流估计器与边缘检测器结合在一个相互促进的架构中。跨任务特征融合（CFF）模块在特征层面将模态不变边缘线索嵌入运动估计流程，从而提高跨模态对应关系的可靠性。随后，迭代特征细化（IFR）模块循环更新边缘特征和光流特征，确保几何一致性并逐步提升预测质量。这种协同作用生成边缘感知、深度对齐的光流场，通过基于随机抽样一致性的 PnP 求解器实现更稳定、准确的位姿恢复。总体而言，我们的主要贡献总结如下：

- 我们提出了一种新颖的双任务学习框架，用于基于激光雷达地图的事件视觉定位。该框架融合了光流估计器和边缘检测器，能够生成边缘感知表征，从而增强跨模态一致性，促进事件-激光雷达数据对齐。
- 我们引入 CFF 和 IFR 模块，以相互促进的方式促进两个任务分支的学习，从而提升定位和边缘预测性能。
- 我们在涵盖不同场景的公共数据集上进行了大量实验，证明了所提出的边缘特征增强模块的有效性，并在定位精度方面取得了最先进的性能。

II. 相关工作

基于事件的视觉定位已从多个角度展开研究。本节回顾该领域主流研究方向的关键进展。

检索方法通过将事件流转换为可查询表征来避免跨模态匹配，例如图像 [1], [3], [4]或边缘图 [20],，从而能够针对参考数据库进行基于相似度的匹配。其他方法则构建手工设计的 [5] 或学习型描述子 [6], [8], ，或将隐式表征编码到网络权重中 [7]。然而，这些方法通常需要代价高昂的数据库构建和维护，并且只能实现粗略定位，可作为我们方法的初始化步骤。

另一类研究侧重于训练神经网络，直接根据事件回归相机位姿 [15]–[18],，并隐式编码环境信息。例如，Nguyen等人 [15]使用卷积神经网络和堆叠空间长短期记忆网络，根据短暂的事件突发估计位姿。尽管这类方法无需显式地图，但其可扩展性有限，且对未见环境的泛化能力较差。

位姿细化方法假设已知三维场景表示以及事件相机状态的初始估计。根据场景表示的类型，这些方法大致可分为几何方法和光度方法。几何方法通常构建包含边缘特征的三维地图，并通过将事件与这些特征对齐来优化相机位姿。例如，Yuan等人 [9] 构建了仅由垂直线组成的场景地图，并通过将二维事件线与其三维对应线进行匹配，实现三自由度定位。较新的方法[10]–[12] 利用深度传感器或运动恢复结构（SfM）技术构建半稠密地图，并通过事件与半稠密地图之间的极性感知几何对齐来估计六自由度位姿。然而，由于这类方法严重依赖显式边缘特征，因此在纹理含噪或较弱的场景中性能会下降。光度方法则从包含颜色信息的场景地图中推断强度变化 [13], [14], [21]。例如，Bryner等人 [14] 提出了一种利用生成式事件模型和光度三维地图，通过对齐观测到的强度变化图像与预测的强度变化图像来估计位姿的方法。然而，传统光度地图通常缺乏视角相关外观建模，因此在视点变化时表现能力有限。为此，Liu等人 [21] 引入了基于高斯泼溅的地图，能够编码更复杂的环境光照效果。不过，高斯泼溅中的渲染伪影或不一致性会直接影响定位精度。

与上述方法相比，激光雷达地图具有稠密、精确且不受光照影响的几何信息，因此非常适用于基于事件的定位。此外，激光雷达地图通常可直接从第三方提供商处获取，无需进行任务特定场景构建。尽管潜力巨大，鲜有方法解决事件与激光雷达点云之间的跨模态匹配难题。 [22], [23] 是最早尝试将这一方法用于事件相机与激光雷达外参标定的工作之一。然而，这些方法通常依赖任务特定标记或受控环境，限制了其可扩展性。E2PNet [24] 提出了一种基于学习的类图像事件特征表示，并将其用于现有的基于RGB的方法 [25], [26], 以实现事件到点云定位。不过，该方法仅报告了粗定位性能。与我们的方法最密切相关的是 EVLoc[19], ，它通过直接学习事件与激光雷达地图之间的稠密流场来解决对齐问题。然而，它并未明确利用模态不变特征来增强跨模态对应关系。相比之下，我们

提出了一种双任务框架，通过以相互促进的方式联合学习边缘检测器和光流估计器，明确针对这一表征瓶颈，从而实现特征空间中的一致性，并显著提升事件-LiDAR 定位的鲁棒性和精度。

III. 预备知识

在本节中，我们描述从基线EVLoc [19], 继承的通用模块，包括数据预处理和光流估计网络。

*1)*数据预处理：给定初始相机位姿$\boldsymbol{T}_{\text{init}}$，我们将原始激光雷达点 P_w 变换到相机坐标系，并使用针孔相机模型将其投影到图像平面，以获得深度图I_D。同时，在给定时间窗口内生成结构清晰的事件图像。 [19] 中的事件图像生成包含两个步骤：去模糊和去噪。在我们的实现中，我们保留去模糊步骤，但由于其计算成本较高而省略去噪。取而代之的是，我们采用Prophesee SDK [27]中的时空对比度滤波器和拖尾滤波器对原始事件流进行预处理。这些滤波器能够有效减少噪声和事件数量。滤波之后，我们应用 [19] 中的事件图像生成算法生成最终的事件图像 I_{EV}。

2）光流估计网络：与 [19],相同，我们使用经过修改的 RAFT [28] 网络，将跨模态事件 – 深度匹配问题表述为稠密事件–深度光流预测任务，并采用掩码 L2 损失进行训练：

$$\mathcal{L}_{\text{flow}} = \frac{\sum_{u,v} \mathbf{1}\left[\boldsymbol{f}_{\text{gt}}(u,v) \neq 0\right] \left\|\boldsymbol{f}(u,v) - \boldsymbol{f}_{\text{gt}}(u,v)\right\|_2}{\sum_{u,v} \mathbf{1}\left[\boldsymbol{f}_{\text{gt}}(u,v) \neq 0\right] + \epsilon}, \quad (1)$$

其中， $\boldsymbol{f}$ 表示预测的事件—深度光流。 $\boldsymbol{f}_{\text{gt}}$ 表示真实事件–深度光流，该光流通过使用真值位姿 $\boldsymbol{T}_{\text{gt}}$和初始位姿 $\boldsymbol{T}_{\text{init}}$ 将三维点投影到图像平面，并计算对应二维投影之间的差值得到。 $\mathbf{1}\left[\boldsymbol{f}_{\text{gt}}\ (u,\ v)\ \neq 0\right]$在真实光流有效（非零）时返回 1，否则返回 0。 ϵ 是一个较小的常数，用于确保数值稳定性。

备注：EVLoc [19] 展示了使用光流估计网络预测事件与激光雷达地图之间稠密对应关系的可行性。然而，该方法对事件与激光雷达点云之间跨模态一致性的分析有限。这项任务的主要挑战源于深度图与事件图像之间的模态与外观差异，这使得跨模态匹配比模态内对齐困难得多。这自然表明，让深度图更具“事件特征”以缩小这种差异，或许能够简化匹配过程。受此启发，我们重新设计了框架，以显式增强跨模态一致性，从而进一步提高定位精度。我们将在下一节介绍所提方法的细节。

IV. 方法

在本节中，我们介绍一种新颖的框架，以相互促进的方式整合光流估计器和边缘检测器（第 IV-A 节），并通过两个新颖的跨任务特征融合模块（第 IV-B 节和第 IV-C 节）实现。我们的框架概览如图2所示。

A. 边缘检测网络

众所周知，事件通常在梯度丰富的区域被触发，即沿着事件图像中的边缘。这一认识促使我们提出从深度图中提取边缘以促进跨模态对齐的思路。然而，深度图仅编码几何边缘结构，缺少事件图像中存在的纹理边缘信息，因此这种对齐并不总是可靠，尤其是在复杂或噪声较大的环境中。为此，我们提出学习一个编码器-解码器边缘检测器，从深度中提取边缘特征，并隐式增强各模态之间的特征级一致性，详见第IV-B节和第IV-C节。

由于缺乏专门为深度图设计的基于学习的边缘检测器，我们采用了广泛使用的HED网络 [29], ，该网络最初是为RGB图像开发的。HED是一种卷积网络，从中间层提取分辨率为$(\mathbf{1}, 1/2, 1/4, 1/8, 1/16)$的五个特征图。每个特征图经过 1×1卷积并上采样以匹配输入分辨率，从而产生五个侧输出。随后，这些输出通过 1×1 卷积融合，生成最终的边缘图。

我们使用输入事件图像生成真实边缘图。具体而言，由于由真值位姿生成的深度图与事件图像对齐，我们首先对深度图进行膨胀，以确保覆盖稀疏事件像素。然后，针对每个事件像素，我们获取其对应的深度值 d。使用针孔相机模型，我们在相机坐标系中重建对应的三维点 P_{edge}。接下来，根据真值位姿与初始位姿之间的相对变换，使用投影函数 π对三维边缘点进行重投影：

$$I'_{\text{EV}} \triangleq \pi(P_{\text{edge}}, \boldsymbol{T}_{\text{init}}^{-1}\boldsymbol{T}_{\text{gt}}) \quad (2)$$

最后，我们通过对重投影事件图像 I'_{EV}进行二值化，生成真实边缘图 I_{ED} 。

网络使用类别平衡交叉熵损失函数 [29], 进行训练，其定义如下：

$$\begin{aligned}\mathcal{L}_{\text{edge}} = -\sum_{u,v} \Big[& wI_{\text{ED}}(u,v)\log\left(p(u,v)\right) \\ & + (1-w)(1-I_{\text{ED}}(u,v))\log\left(1-p(u,v)\right)\Big]\end{aligned} \quad (3)$$

这里， $p(u,v)$表示像素 (u,v) 处的预测边缘概率， w 表示根据真实边缘图 I_{ED}中非边缘像素的比例计算得到的类别平衡权重。该损失通过为少数类（边缘像素）赋予较高权重，解决边缘像素与非边缘像素之间的不平衡问题。

B. 跨任务特征融合

如上所述，生成的深度图和事件图像被输入光流估计器和边缘检测器，它们首先分别生成事件特征 F_{EV}、

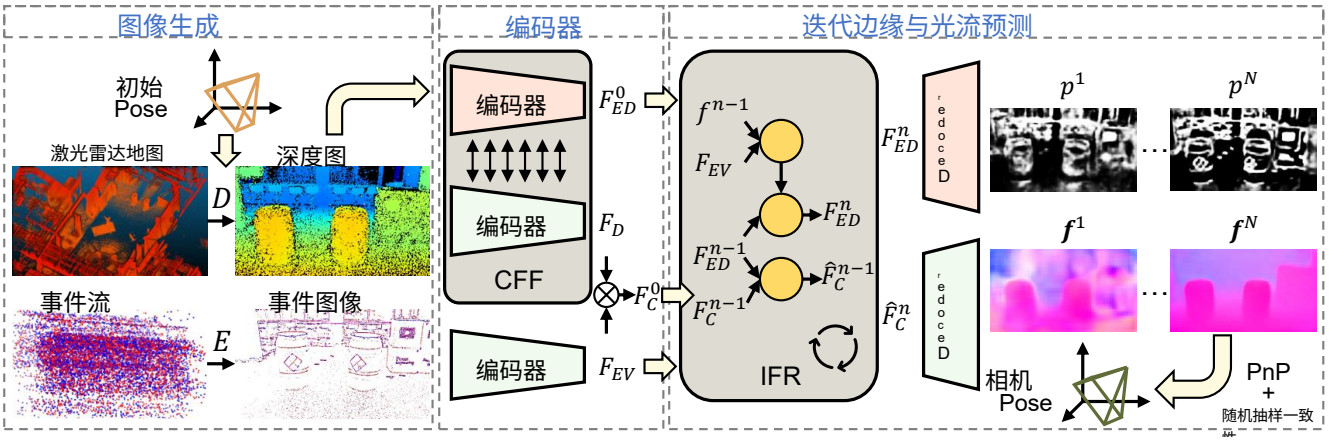


图2. 所提出方法的三个阶段：1）图像生成：基于给定的初始位姿估计，使用针孔相机模型生成深度图 D，并使用事件图像生成器从原始事件流构建事件图像 E；2）编码器：将深度图和事件图像输入光流估计分支（**绿色模块**），同时将深度图作为边缘检测分支（**红色模块**）的输入。编码后的深度特征 F_D 和边缘特征 F_{ED} 通过 CFF 模块进行跨尺度融合，以生成边缘感知表征；3）迭代边缘与光流预测：将这些特征与事件特征 F_{EV} 及相关体 F_C（通过使用当前光流对 F_D 和 F_{EV} 的逐元素乘积进行索引构建；参见图5）一同输入 IFR模块，在其中对光流 $\boldsymbol{f}$ 和边缘表示 F_{ED} 进行迭代式联合细化，最终由相应的解码器生成光流估计 $\boldsymbol{f}^{1\to N}$ 和边缘估计 $p^{1\to N}$。最终的光流估计 $\boldsymbol{f}^N$ 用于在随机抽样一致性循环中通过 PnP求解器恢复相机位姿。IFR模块中的黄色圆圈表示特定操作。详情请参见第IV-C节和图5。

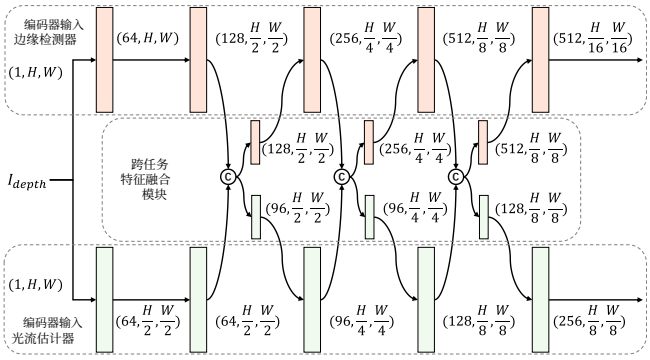


图3. *CFF* 模块的架构。边缘检测分支和光流估计分支均采用五层编码器，且每层的空间分辨率不同。如图所示，特征融合在中间三层进行。

深度特征 F_D 和边缘特征 F_{ED}。两个分支中的深度编码器通过我们提出的 CFF 模块进行整合，如图3所示。边缘检测器和光流估计器的编码器所生成的多尺度特征图逐步融合并传递。光流估计器编码器从输入深度图中提取五个分辨率为 $(1/2,\ 1/2,\ 1/4,\ 1/8,\ 1/8)$ 的特征图，而边缘检测器编码器则生成分辨率为 $(1,\ 1/2,\ 1/4,\ 1/8,\ 1/16)$ 的特征图。由于分辨率不匹配，融合仅在中间三层进行。第 i 层的特征图经过拼接后，通过两个独立的1×1 卷积恢复各任务分支的原始通道维度，具体表述如下：

$$\hat{F}_s^i = \mathrm{Conv}_{1\times1}^s\big(\mathrm{Concat}(F_{\mathrm{ED}}^i,\ F_{\mathrm{D}}^i)\big), \tag{4}$$

其中 $s \in \{\mathrm{ED},\ \mathrm{D}\}$ 表示任务分支（边缘检测或深度），$\hat{}\,F_{\mathrm{ED}}^i$ 和$\hat{}\,F_{\mathrm{D}}^i$ 为增强后的特征。尽管两个编码器都处理相同的输入深度图，但它们关注的侧重点不同：边缘检测编码器强调细粒度低层边缘线索，而光流估计编码器主要提取运动相关特征，既关注局部对应关系，也关注全局运动上下文。通过跨任务特征融合，深度特征对局部纹理细节的敏感度得到增强，而边缘特征则变得更具几何结构，从而提升了跨模态一致性。如图4所示，应用 CFF 模块后获得的深度特征图包含更丰富的纹理细节，这些细节与真实边缘图中的结构相吻合。

C. 迭代特征细化

如图5所示，提取的特征被输入 IFR模块。该模块由两个部分组成：**(1)** 边缘更新分支（**上方**），通过整合来自对齐的事件特征的信息来细化边缘特征；**(2)** 光流更新分支（**下方**），通过将边缘特征融入相关体来细化光流估计。

首先，通过计算所有深度特征 F_D 与事件特征 F_{EV} 之间的点积，构建相关性金字塔。初始光流 $\boldsymbol{f}^0$ 设为零。在迭代 n（$n \in [1, N]$）时，使用当前光流估计值 $\boldsymbol{f}^{n-1}$ 在相关性金字塔中进行索引，并提取相应的相关性体 F_C^n。在光流更新分支中，首先沿通道维度将相关性体与当前边缘特征 F_{ED}^n 拼接，然后通过一个 1×1 卷积恢复原始通道数。GRU单元使用 $\hat{}\,F_C^n$、当前光流和上下文特征 F_{ctx}（其中 F_{ctx}和 h^0 来自 EVLoc风格上下文编码器 [19]）更新隐藏状态，再由光流头 $\boldsymbol{H}$ 将所得状态映射为增量光流更新。接着，在边缘更新分支中，利用更新后的光流扭曲事件特征 F_{EV}，使其在空间上与边缘特征 F_{ED}对齐。当前边缘特征 F_{ED}^n取自第四编码器层，以匹配事件特征的分辨率，并拆分为上下文特征 c_{ED} 和隐藏状态 h_{ED}^n。一个GRU单元

以上下文特征、隐藏状态和对齐的事件特征为输入，并输出更新后的隐藏状态。将更新后的隐藏状态与原始上下文特征拼接，形成细化后的边缘特征 F_{ED}^{n+1} 。

$$\hat{F}_C^n = \text{Conv}_{1\times1}\big(\text{Concat}(F_C^n, F_{ED}^n)\big), \tag{5}$$

$$\boldsymbol{f}^n = \boldsymbol{f}^{n-1} + H(\text{GRU}\big(h^{n-1}, \hat{F}_C^n, \boldsymbol{f}^{n-1}, F_{\text{ctx}}\big)), \tag{6}$$

$$F_{ED}^{n+1} = \text{Concat}(c_{ED}, \text{GRU}\big(h_{ED}^n, W(F_{EV}, \boldsymbol{f}^n), c_{ED}\big)). \tag{7}$$

通过这种方式整合边缘特征，相关性体融入了局部结构细节，补充了全局运动上下文，使模型能够预测更精细的光流。值得注意的是，这两个分支相互促进：光流的改善有助于获得对齐得更好的事件特征，而更准确的边缘特征则增强了融合后的相关性表示，进而进一步引导光流细化过程。请参见图6底部几行。

D. 其他细节

损失函数： 如上所述，边缘检测器和光流估计器以相互促进的方式进行联合训练。最终损失定义为以下两项的加权组合：

$$\mathcal{L} = \alpha \times \mathcal{L}_{flow} + \beta \times \mathcal{L}_{edge}, \tag{8}$$

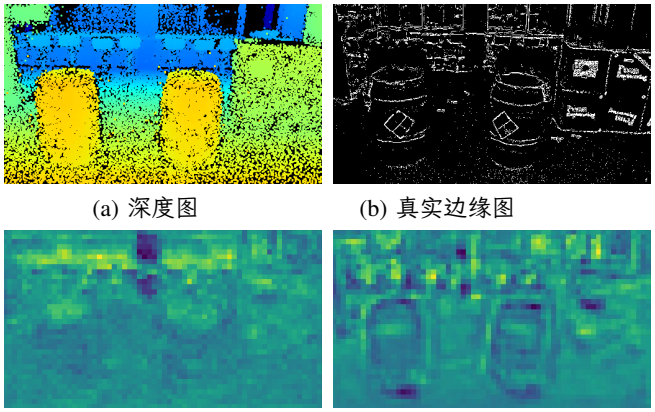
其中 α 在我们的实验中为 1， β 为 100。

位姿估计： 我们将深度值非零的像素视为有效像素，并仅在这些位置提取光流预测（即 2D-3D对应关系）。我们使用公开可用的 PoseLib 库 [30], ，将最大重投影误差设为 12.0，并使用 Huber损失进行优化，同时将其他所有参数保持为默认值。

五、实验

A. 数据集与实现细节

1）数据集：M3ED数据集 [31] 是一个典型的数据集，提供高分辨率激光雷达地图和真实轨迹，因此我们主要在该数据集上评估我们的方法。为进行额外评估，我们还纳入了 DSEC数据集 [32],



(a) 深度图 (b) 真实边缘图

(c)基线特征图 (d)添加 *CFF* 后的特征图

图4. 编码后的深度特征图可视化。可视化特征图中每个像素的值表示该位置对应特征向量的均值。应用 CFF 模块后的深度特征图包含更丰富且与结构相吻合的纹理细节。

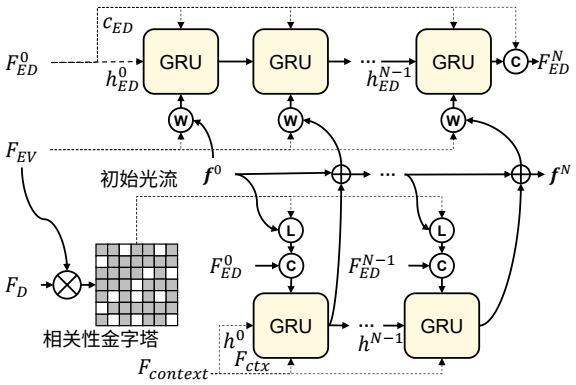


图5. *IFR* 模块的架构。边缘特征（$F_{ED}^0 \rightarrow F_{ED}^N$）和光流估计（$\boldsymbol{f}^0 \rightarrow \boldsymbol{f}^N$）通过相互增强逐步细化。W、L 和 C 分别表示变形、索引和拼接。

表 I 消融研究结果。除上界 Oracle 外，粗体数字表示最佳性能。*Oracle*。

	CFF	IFR	EPE	平均误差 平移[厘米]↓	旋转[°]↓	中位误差 平移[厘米]↓	旋转[°]↓
Oracle			1.66	4.14	0.39	3.31	0.22
基线 [19]			6.10	9.01	1.10	8.11	0.97
直接掩码			6.11	8.99	1.02	7.80	0.85
(a)	✓		5.75	**8.06**	0.94	7.20	0.83
(b)	✓	✓	**5.52**	8.07	**0.93**	**6.85**	**0.74**

该数据集包含 M3ED 中没有的更复杂驾驶场景。然而，DSEC 仅提供视差图像。为便于评估，我们使用针孔相机模型将视差推导的深度反投影至三维空间，为每个样本重建三维点云。此外，由于我们的方法依赖高质量的预构建激光雷达地图来生成稠密且准确的深度估计，因此我们排除了受动态物体影响严重的序列，因为这些物体的运动会在重建地图中引入显著的噪声伪影。

2) 数据处理：为模拟粗定位中的不确定性，我们通过随机均匀采样扰动真实值（DSEC 中为单位矩阵），生成初始位姿估计，其中平移偏移量不超过 ± 50 厘米，旋转偏移量不超过 $\pm 5°$。为确保一致性和效率，所有输入图像均按各数据集特定的固定分辨率进行裁剪和缩放。对于 M3ED 数据集，图像缩放并裁剪至 512×288；对于 DSEC，图像裁剪至 480×360。

3) 训练细节：对于 M3ED，由于该数据集涵盖多种不同环境（室内飞行、室外停车场、森林等），每个场景分别训练一个模型；例如，在 $falcon\ indoor\ flight$ 场景中，前两个序列用于训练，最后一个序列用于评估。对于只有一个序列的场景（例如，$spot\text{-}indoor\text{-}building\text{-}loop$），我们采用时间划分，将前 70% 的数据用于训练，剩余 30% 用于评估，以确保训练集和测试集没有重叠。相比之下，一个模型在所有

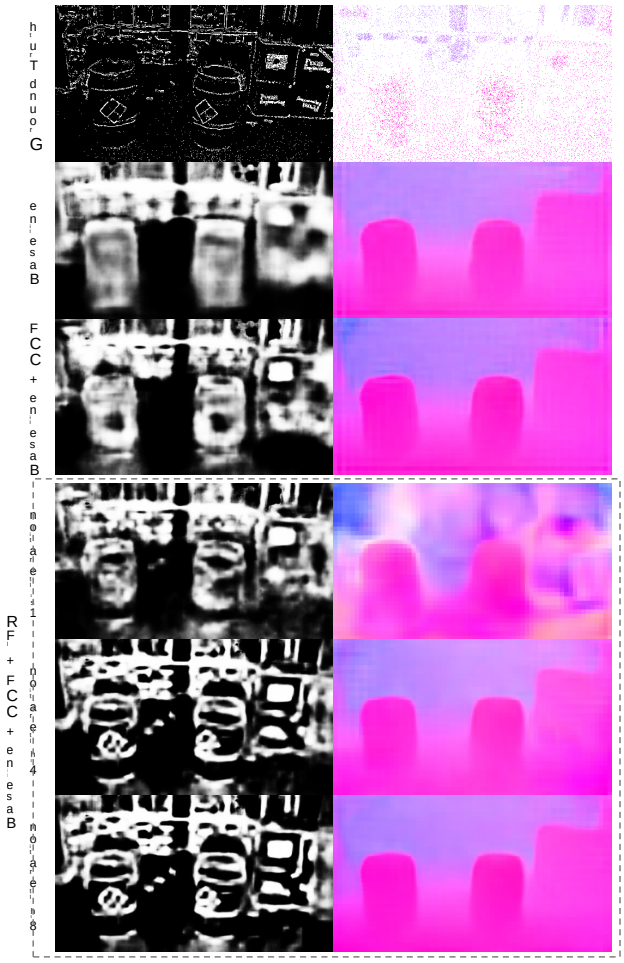


图6。 预测边缘和光流的可视化。 **左列：** 边缘预测； **右列：** 光流估计。

DSEC 场景上共同训练——因为这些场景均呈现具有相似传感器配置和运动模式的驾驶环境——并将每个场景的最后一个序列用于评估。在每次实验中，模型训练 100 个 epoch，批量大小为 2，权重衰减为 1×10^{-4}，初始学习率为4×10^{-5}，并使用 *OneCycleLR* 学习率调度器和 Adam 优化器 [33]。IFR模块在训练期间设置为运行 12 次迭代，在测试期间设置为运行 24 次迭代。所有实验均在单张 NVIDIA RTX 4090 GPU 上进行。

4）评估指标： 评估时，使用预测位姿的平均误差和中位误差来衡量定位精度，而使用平均端点误差（EPE）来评估估计光流的精度。

B. 消融研究

为评估所提出模块的有效性，我们进行了一系列消融实验，如表 I 和图6所示。基线为 EVLoc [19]，，我们使用与本方法相同的图像分辨率和实验设置对其重新训练。在*Oracle* 设置中，我们在训练和测试之前使用真实边缘图对深度输入进行掩码处理，从而生成掩码深度图，使其纹理与输入事件图像十分相似。

这最大限度地缩小了不同模态之间的外观差异。正如预期，该模型的表现显著优于基线，展现了在具备完美边缘信息时基于光流的匹配所能达到的理论上界。接下来，我们探索一种渐进式流程直接掩码，将预测的边缘图直接用于掩码处理深度输入。边缘检测器独立训练，图6第2行展示了预测边缘图的示例。此配置的表现略优于基线，验证了即使边缘预测不够完美，也能改善深度–事件匹配。

在此基础上，我们逐步引入所提出的模块。通过 CFF 模块将边缘特征融合到光流估计器中，模型 *(a)* 在简单掩码的基础上取得了进一步提升。这表明，隐式的特征级融合比直接掩码更有效。此外，如图6第3行所示，与独立边缘检测器生成的边缘相比，预测边缘呈现出更精细的细节，这说明 CFF 模块通过整合互补信息，也增强了边缘检测分支。加入 IFR模块后，定位精度进一步提升（模型 *(b)*），证实了该模块有助于细化光流估计。与独立边缘检测器 HED（图6第2行）相比，我们的方法能够生成更清晰、细节更丰富的边缘预测，表明 IFR模块能够有效地将事件特征中的局部纹理线索融合到边缘表示中。此外，随着迭代的推进，预测边缘和细化后的光流估计都变得愈加完整、准确，如图6最后三行所示，突显了迭代细化过程的相互促进作用。

C. 结果分析

本节在相同的实验设置下，将我们的方法与最先进的事件-LiDAR 定位方法 EVLoc [19]进行比较，并与基于 RGB 图像的图像到 LiDAR 方法 I2D-Loc [34]进行对比。结果见表 II。M3ED 未提供标准相机的真实位姿。虽然可以通过插值获得近似位姿，但由此生成的深度图与相机帧存在明显错位。相比之下，DSEC 为每张 RGB 图像提供真实视差图，因此我们可以在标准相机实验中采用上述相同的真实值生成方法。

如表 II 所示，在不同光照条件下，我们提出的方法在所有序列上的表现始终优于 EVLoc，展现了该方法的整体优势。在 M3ED 的室内环境中（$falcon\ \underline{i}ndoor\ \underline{f}light$ 和 $spot\ \underline{i}ndoor\ \underline{b}uilding\ \underline{l}oop$），与 EVLoc 相比，LEAR 将平移误差降低了约 15–20%，将旋转误差降低了约 $20-25\%$。在更具挑战性的室外场景中，两种方法的整体表现均不如室内环境，这表明有限或无规律的场景纹理会增加跨模态匹配的难度。尽管如此，与基线相比，我们的方法并未出现性能下降

表 II M3ED 和 DSEC 上 I2D-Loc、EVLoc 与所提方法的定位性能比较。采用平移和旋转的中位误差作为评估指标。粗体数字表示最佳性能。

数据集	测试场景	Time	I2D-Loc [34]		EVLoc [19]		Ours	
			平移[厘米]↓	旋转[°]↓	平移[cm]↓	旋转[°]↓	平移[cm]↓	旋转[°]↓
M3ED [31]	*Falcon* 室内飞行	-	-	-	8.11	0.97	**6.85**	**0.74**
	Spot 室内建筑环线-	-	-	-	12.88	2.07	**10.94**	**1.63**
	室外日间 *Penno* 停车场的猎鹰	day	-	-	25.11	1.32	**22.40**	**1.18**
	猎鹰从森林进入森林	day	-	-	20.46	1.87	**19.35**	**1.84**
	Spot 户外地点白天 *penno* 短环路	day	-	-	17.89	0.88	**14.37**	**0.74**
	Spot 林间道路	day	-	-	13.55	0.60	**12.96**	**0.57**
	汽车 城市 白天 *penno*	day	-	-	14.15	0.60	**11.88**	**0.56**
	汽车 森林 进入池塘	day	-	-	19.76	0.86	**16.23**	**0.75**
	户外地点 夜晚 *penno* 短环路 -	夜晚	-	-	34.85	1.71	**24.97**	**1.04**
	汽车 城市 夜晚 *penno*	夜晚	-	-	20.84	**0.72**	**19.15**	0.87
DSEC [32]	因特拉肯 *00*	黎明/黄昏	10.45	0.42	7.11	0.30	**5.56**	**0.25**
	图恩 *00*	黎明/黄昏	11.15	0.55	8.26	0.42	**8.08**	**0.41**
	苏黎世市 *00*	黎明/黄昏	11.30	0.52	7.94	0.37	**6.78**	**0.34**
	苏黎世市 *01*	黎明/黄昏	9.88	0.36	7.63	0.30	**6.15**	**0.28**
	苏黎世市 *02*	黎明/黄昏	8.68	0.33	7.93	0.31	**6.68**	**0.27**
	苏黎世市 *03*	夜晚	14.01	0.74	8.79	0.61	**8.14**	**0.51**
	苏黎世市 *04*	黎明/黄昏	11.82	0.51	7.49	0.30	**6.16**	**0.26**
	苏黎世市 *05*	day	9.27	0.37	6.42	0.28	**5.36**	**0.24**
	苏黎世市 *06*	day	14.26	0.62	8.12	0.39	**6.98**	**0.37**
	苏黎世市 *07*	day	11.33	0.43	7.27	0.34	**6.25**	**0.32**
	苏黎世市 *08*	黎明/黄昏	8.72	0.36	6.62	0.27	**5.81**	**0.26**
	苏黎世市 *09*	夜晚	8.05	0.32	6.38	0.29	**5.92**	**0.28**
	苏黎世市 *10*	夜晚	11.35	0.49	8.37	0.35	**7.35**	**0.34**
	苏黎世市 *11*	黎明/黄昏	9.56	0.41	6.14	0.26	**5.21**	**0.23**

，仍取得了小幅提升，说明模型能够继续从复杂环境中学习有用特征。夜晚序列尤其具有挑战性：即使经过去噪和去模糊，生成的事件帧仍几乎不包含所拍摄场景的有用信息，反而含有大量噪声；此外，事件相机的高分辨率进一步加剧了噪声影响。尽管缺乏有用的结构信息，LEAR 的表现仍与 EVLoc 相当。

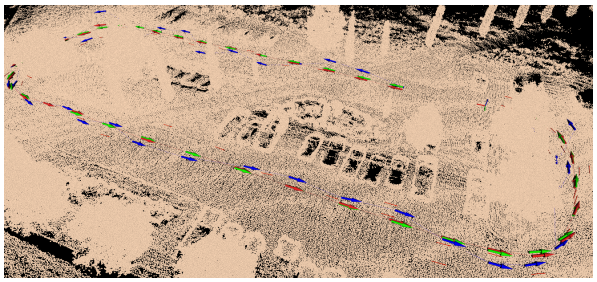


图 7. *M3ED* 上真实位姿（绿色）、初始位姿（蓝色）和我们的方法预测的位姿（红色）对比。我们的预测与真实值高度吻合，展现出很高的准确性。

DSEC仅包含驾驶场景，但光照条件多样。LEAR始终优于EVLoc，平移误差最多降低36%，旋转误差最多降低29%，同时也超越了I2D-Loc，平移误差和旋转误差分别最多降低61%和44%。相较于I2D-Loc的显著提升凸显了事件相机在跨模态数据关联方面的优势。我们认为，这可能归因于白天与夜晚RGB图像之间大得多的域差距，导致模型收敛到局部最小值。相比之下，事件图像保持一致，跨越白天与夜晚两种条件。

图 7 展示了 M3ED 中一个场景的定位结果。从粗略初始位姿（蓝色）出发，LEAR 对其进行优化，得到一个精确的预测结果（红色），与真实值（绿色）高度吻合。

表 III 运行时间和内存消耗对比。

方法	阶段	运行时间 [毫秒]	内存 [MB]	平移[厘米]/旋转[°]
EVLoc [19]	特征编码	4.33	1202	8.11/0.97
	光流估计（24 迭代次数）	51.80		
Ours	特征编码	7.71	1232	6.85/0.74
	光流估计（24 迭代次数）	100.43		
	特征编码	7.71	1232	7.01/0.75
	光流估计（12 次迭代）	51.30		

表 III 总结了基线方法和我们的方法在 M3ED 上的运行时间与内存消耗。在迭代次数相同的情况下，我们的方法几乎不增加额外内存开销，但运行时间比 EVLoc 更长（主要耗费在光流估计上）。尽管如此，这一开销仍在可接受范围内，并带来了定位精度的显著提升。此外，我们的方法在 12 次迭代时仍优于 EVLoc 在 24 次迭代时的表现，而两者在这种情况下的运行时间相近。这些结果证明了我们框架中引入的增强特征表示和迭代细化的有效性。

VI. 结论与未来工作

本文提出了一种用于激光雷达地图中基于事件的视觉定位的双分支框架。我们的方法将光流估计网络与边缘检测网络相结合，并引入了跨任务特征融合模块和迭代特征细化模块。这些组件使两个分支能够在特征层面进行隐式融

合，从而增强跨模态特征的一致性，以便为定位找到准确的对应关系。在多个公开数据集上进行的大量实验表明，所提出方法是有效的。

局限性：尽管我们的方法具有诸多优势，但它依赖于二维空间中的局部光流估计，这要求输入深度图与事件图像之间有足够的重叠。此外，成功匹配在很大程度上取决于是否存在足够的深度变化。在深度信息退化的场景中，例如平坦地面或光滑墙面，由于缺乏可区分的几何特征，我们的方法可能会失效。未来，我们计划研究不经投影的直接事件到点云配准，这或许能在初始位姿精度有限或几何变化较差的情况下提供更强的鲁棒性。

VII. 致谢

本研究项目获得了奥地利科学基金（FWF）通过项目协议（I 6747-N）EVELOC 的资助。

参考文献

[1] T. Fischer 和 M. Milford，《利用时间窗口集合进行基于事件的视觉地点识别》，*IEEE*机器人与自动化快报，第 5 卷，第 4 期，第 6924–6931 页，2020 年。[2] A. J. Lee 和 A. Kim，《Eventvlad：利用事件相机重建的边缘进行视觉地点识别》，载于 *IEEE/RSJ*智能机器人与系统国际会议，2021 年，第 2247–2252 页。[3] G. B. Nair、M. Milford 和 T. Fischer，《通过事件相机中的快速和慢速自适应偏置增强视觉地点识别》，载于 *IEEE/RSJ*智能机器人与系统国际会议。IEEE，2024 年，第 3356–3363 页。[4] M. Milford、H. Kim、S. Leutenegger 和 A. Davison，《迈向基于事件相机的视觉同步定位与建图》，载于 机器人学：科学与系统研讨会，2015 年。[5] T. Fischer 和 M. Milford，《需要多少事件？利用稀疏但不断变化的像素进行基于事件的视觉地点识别》，*IEEE*机器人与自动化快报，第 7 卷，12 282，2022 4，12 275 无页码。[6] D. Kong、Z. Fang、K. Hou、H. Li、J. Jiang、S. Coleman 和 D. Kerr，《Event-vpr：利用基于事件的视觉传感器进行视觉地点识别的端到端弱监督深度网络架构》，*IEEE*仪器与测量汇刊，第 71 卷，18，2022 1 页。[7] S. Hussaini、M. Milford 和 T. Fischer，《脉冲神经网络在视觉地点识别中的应用》，*IEEE*机器人学汇刊，2024 年。[8] O. Claxton、C. Malone、H. Carson、J. J. Ford、G. Bolton、I. Shames 和 M. Milford，《通过验证定位估计改进基于视觉地点识别的机器人导航》，*IEEE*机器人与自动化快报，2024 年。[9] W. Yuan 和 S. Ramalingam，《利用事件传感器进行快速定位与跟踪》，载于 *IEEE*机器人与自动化国际会议。IEEE，2016 年，第 4564–4571 页。[10] R. Yuan、T. Liu、Z. Dai、Y.-F. Zuo 和 L. Kneip，《Evit：利用窗口化非线性优化在半稠密地图中进行基于事件的视觉惯性跟踪》，载于 *IEEE/RSJ*智能机器人与系统国际会议。IEEE，2024 年，第 10 656–10 663 页。[11] Y.-F. Zuo、W. Xu、X. Wang、Y. Wang 和 L. Kneip，《在复杂条件下对事件相机进行跨模态半稠密六自由度跟踪》，*IEEE*机器人学汇刊，第 40 卷，第 1600–1616 页，2024 年。[12] N. Xu、L. Wang、Z. Yao 和 T. Okatani，《Mets：用于基于事件的高速位姿跟踪的运动编码时间表面》，国际计算机视觉期刊，第 1–19 页，2025 年。[13] G. Gallego、J. E. Lund、E. Mueggler、H. Rebecq、T. Delbruck 和 D. Scaramuzza，《基于事件的相机跟踪：利用光度深度图进行六自由度跟踪》，*IEEE*模式分析与机器智能汇刊，第 40 卷，第 10 期，第 2402–2412 页，2017 年。[14] S. Bryner、G. Gallego、H. Rebecq 和 D. Scaramuzza，《利用非线性优化基于光度三维地图进行直接事件相机跟踪》，载于 IEEE机器人与自动化国际会议。IEEE，2019 年，第 325–331 页。[15] A. Nguyen、T.-T. Do、D. G. Caldwell 和 N. G. Tsagarakis，《利用堆叠空间LSTM网络对事件相机进行实时六自由度位姿重定位》，载于 IEEE/CVF计算机视觉与模式识别会议研讨会论文集，2019 年，第 0–0 页。[16] Y. Jin、L. Yu、G. Li 和 S. Fei，《基于CNN-LSTM和图像去噪的六自由度事件相机重定位系统》，《Expert Systems with Applications》，第 170 卷，114535 页，2021 年。[17] H. Lin、M. Li、Q. Xia、Y. Fei、B. Yin 和 X. Yang，《利用熵帧和注意力网络对事件相机进行六自由度位姿重定位》，载于第 18 届 ACM SIGGRAPH 虚拟现实连续体及其工业应用国际会议论文集，2022 年，第 1–8 页。[18] H. Ren、J. Zhu、Y. Zhou、H. Fu、Y. Huang 和 B. Cheng，《一种简单有效的基于点的网络，用于事件相机六自由度位姿重定位》，载于 IEEE/CVF计算机视觉与模式识别会议论文集，2024 年，第 18 112–18 121 页。[19] K. Chen、J. Zhang 和 F. Fraundorfer，《Evloc：通过事件-深度配准在激光雷达地图中进行基于事件的视觉定位》，arXiv 预印本 arXiv:2503.00167，2025 年。[20] A. J. Lee 和 A. Kim，《Eventvlad：利用事件相机重建的边缘进行视觉地点识别》，载于 IEEE/RSJ智能机器人与系统国际会议。IEEE，2021 年，第 2247–2252 页。[21] T. Liu、R. Yuan、Y. Ju、X. Xu、J. Yang、X. Meng、X. Lagorce 和 L. Kneip，《Gs-evt：基于高斯泼溅的跨模态事件相机跟踪》，arXiv 预印本 arXiv:2409.19228，2024 年。[22] K. Ta、D. Bruggemann、T. Br¨odermann、C. Sakaridis 和 L. Van Gool，《L2e：利用激光将信息转化为事件，对激光雷达和事件相机进行六自由度外参标定》，载于 IEEE机器人与自动化国际会议。IEEE，2023 年，第 11 425–11 431 页。[23] W. Xing、S. Lin、L. Yang 和 J. Pan，《利用边缘对应关系对事件-激光雷达组合进行无靶标外参标定》，IEEE机器人与自动化快报，第 8 卷，第 7 期，第 4020–4027 页，2023 年。[24] X. Lin、C. Qiu、S. Shen、Y. Zang、W. Liu、X. Bian、M. M¨uller、C. Wang 等，《E2pnet：通过时空表示学习进行事件到点云配准》，《Advances in Neural Information Processing Systems》，第 36 卷，第 18 076–18 089 页，2023 年。[25] J. Li 和 G. H. Lee，《Deepi2p：通过深度分类进行图像到点云配准》，载于 IEEE/CVF计算机视觉与模式识别会议论文集，2021 年，第 15 960–15 969 页。[26] Q.-H. Pham、M. A. Uy、B.-S. Hua、D. T. Nguyen、G. Roig 和 S.-K. Yeung，《Lcd：用于二维-三维匹配的学习得到的跨域描述子》，载于 AAAI 人工智能会议论文集，第 34 卷，第 07 期，2020 年，第 11 856–11 864 页。[27] Prophesee，《Prophesee Metavision SDK》，网址：https://docs. prophesee.ai/stable/index.html。[28] Z. Teed 和 J. Deng，《Raft：用于光流的循环全配对场变换》，载于欧洲计算机视觉会议。Springer，2020 年，第 402–419 页。[29] S. Xie 和 Z. Tu，《整体嵌套边缘检测》，载于 IEEE/CVF计算机视觉国际会议论文集，2015 年，第 1395–1403 页。[30] V. Larsson 等贡献者，《Poselib——相机位姿估计的最小求解器》，2020 年。 [在线]。网址：https://github.com/vlarsson/PoseLib[31] K. Chaney、F. Cladera、Z. Wang、A. Bisulco、M. A. Hsieh、C. Korpela、V. Kumar、C. J. Taylor 和 K. Daniilidis，《M3ed：多机器人、多传感器、多环境事件数据集》，载于 IEEE/CVF计算机视觉与模式识别会议研讨会论文集，2023 年 6 月，第 4015–4022 页。[32] M. Gehrig、W. Aarents、D. Gehrig 和 D. Scaramuzza，《Dsec：用于驾驶场景的立体事件相机数据集》，IEEE机器人与自动化快报，第 6 卷，第 3 期，第 4947–4954 页，2021 年。[33] D. P. Kingma 和 J. Ba，《Adam：一种随机优化方法》，arXiv 预印本 arXiv:1412.6980，2014 年。[34] K. Chen、H. Yu、W. Yang、L. Yu、S. Scherer 和 G.-S. Xia，《I2d-loc：通过图像到激光雷达深度流进行相机定位》，《ISPRS Journal of Photogrammetry and Remote Sensing》，第 194 卷，第 209–221 页，2022 年。