基于三维基础模型的回环检测：面向分布式协同同步定位与建图

Pierre-Yves Lajoie[1], Benjamin Ramtoula[2], Daniele De Martini[2*], Giovanni Beltrame[1*]

arXiv:2602.02430v1 [cs.RO] 2 Feb 2026

摘要——**分布式协同同步定位与建图（C-SLAM）技术常常难以识别地图重叠区域，这是因为不同机器人之间存在显著的视角变化。受近期三维基础模型的进展启发，这类模型能够在视角差异较大的情况下匹配图像。我们提出一种稳健的回环检测方法，利用这些模型建立机器人间测量值。与需要在集中式地图中进行完整三维重建、资源消耗巨大的方法不同，我们的方法将基础模型集成到现有的同步定位与建图流程中，从而实现可扩展且稳健的多机器人建图。我们的贡献包括：（1）将三维基础模型集成到去中心化协同SLAM中，以便从单目图像对中可靠地估计相对位姿；（2）提出对使用这些相对位姿至关重要的稳健离群值抑制技术；（3）开发专用的位姿图优化形式，以高效消除尺度歧义。我们将该方法与当前先进方法进行比较评估，结果表明，该方法提升了定位和建图精度，同时显著提高了计算效率和内存效率。这些结果凸显了我们的方法在大规模多机器人场景中部署的潜力。**

索引词—**多机器人同步定位与建图、定位**

I. 引言

DECENTRALIZED 协同同步定位与建图（C-SLAM）是多机器人系统在未知环境中运行的一项关键能力。在这类场景中，多台机器人必须独立探索环境，同时交换信息以构建共享的全局地图。当机器人的轨迹各异、传感器布置不同，导致其视角差异显著时，这项任务尤其具有挑战性，因为这会使识别地图重叠区域并生成可靠的机器人间回环闭合变得困难 [1]。此外，当网络问题或带宽限制导致无法进行集中式处理时，智能体之间必须以最少的数据传输完成回环检测，因为交换整张地图对带宽的要求过高。为应对这些挑战，我们提出了一种新颖的回环检测方法，利用三维基础模型 [2], [3] 的最新进展，在不共享地图的情况下处理极端视角变化并寻找重叠区域。具体而言，我们的方法利用三维基础模型的能力，从单目图像对中进行尺度不确定相对位姿估计；得益于在大规模数据集上的广泛预训练，即使面对相反视角和未知图像域也能实现这一点。我们将该方案命名为基于基础模型的回环检测协同同步定位与建图（FLOCC-SLAM）。

FLOCC-SLAM旨在与现有的单机器人里程计系统无缝集成，从而在各个机器人定位流水线之上实现稳健、高效且易于部署的协同同步定位与建图（C-SLAM），前提是这些系统能够提供局部一致的度量尺度里程计（例如VIO）。在位姿图优化过程中，我们的方法利用里程计提供的公制尺度，推断从三维基础模型获得的机器人间测量值的正确尺度。

本文介绍了三项主要贡献：

- 将三维基础模型（MASt3R [3]）集成到去中心化协同SLAM处理流程中，基于单目图像对估计机器人之间的相对位姿，从而能够在视角重叠有限的情况下检测机器人间回环闭合；
- 针对使用MASt3R进行三维相对位姿估计设计异常值检测和不确定性建模，以减少虚假回环闭合的发生；
- 一组专用的位姿图优化公式，用于融合各个机器人地图，解决生成的回环闭合中的三维尺度歧义，并提升整体定位精度。

总而言之，我们的贡献使我们能够利用三维基础模型的最新进展，增强协同SLAM的鲁棒性，并提升其在更广泛环境和多机器人任务中的表现。我们在多个多机器人数据集序列上，将所提方法的性能与最先进的去中心化协同SLAM算法进行了比较。结果表明，三维基础模型生成的强大表征能够显著提高定位精度，同时通过专门的优化和关键帧稀疏化，降低计算与内存开销。这使我们的方法成为一种颇具前景的解决方案，适用于未知环境中的大规模多机器人部署；在这些场景中，机器人间协作对于高效探索与建图至关重要。

本研究部分由 Vanier Canada Graduate Scholarships Award、MITACS Globalink Award、EPSRC 自主智能机器与系统博士培训中心 [EP/S024050/1], 和 Oxa 资助。*共同指导。

[1]P. Lajoie 和 G. Beltrame 任职于加拿大蒙特利尔的蒙特利尔理工学院计算机与软件工程系，´电子邮箱：{pierre-yves.lajoie,giovanni.beltrame}@polymtl.ca。2

B. Ramtoula 和 D. De Martini 任职于英国牛津大学工程科学系牛津机器人研究所，{benjamin,daniele}@robots.ox.ac.uk。

II. 背景与相关工作

在协同同步定位与建图（C-SLAM）中，多台机器人协同构建共享地图，并在其中进行定位。通过共享传感器数据并检测各自地图之间的重叠区域，机器人能够改善各自的定位效果，并创建一个全局一致的

环境视图。本节将介绍协同同步定位与建图（C-SLAM）以及基于图像的相对位姿估计的背景和相关工作。

1）协同同步定位与建图：在协同同步定位与建图中，机器人通常各自执行同步定位与建图，然后共享各自地图的信息，将其融合为对所遍历环境的全局一致估计。与单机器人同步定位与建图类似，协同同步定位与建图通常分为两个部分：负责特征提取和数据关联的前端，以及负责状态估计的后端 [4]。

前端最具挑战性的任务之一，是在重叠区域内高效检测并计算机器人间的回环闭合。这些回环闭合对应于独立机器人估计值之间的连接，可在不同机器人访问相同地点时发现。它们作为拼接点，将局部地图合并为环境的全局表示。为高效合并大型地图，回环检测通常分两个阶段进行：首先通过地点识别确定可能的地图重叠区域，然后进行配准，计算各重叠区域之间的三维相对位姿。

随后，后端根据所有机器人收集的测量值，估计最可能的位姿和地图。我们的工作聚焦于同步定位与建图的位姿图形式，其中将特征边缘化为位姿间测量值，因为这种方法对于较大的地图通常更高效 [4]。

然而，感知混淆现象会导致前端技术产生伪测量值，即不同地点被误认为同一地点。为解决这一在同步定位与建图中普遍存在的问题，已有多种方法被提出，例如成对一致性最大化 [5], 或渐进非凸性（GNC） [6]。

在本研究中，我们重点关注去中心化协同SLAM面临的关键挑战，具体而言，是如何生成成对的机器人间回环闭合并保持尺度一致性，同时无需交换完整地图。近年来，已有多种去中心化协同SLAM系统问世。DSLAM [7] 是早期的一种方法，它利用紧凑的学习型描述子，实现了前端高效的分布式地点识别。后续系统（如 Kimera-Multi [8], ）在此基础上进一步引入了稳健的分布式后端求解器。另一种系统 Swarm-SLAM [9] 则在这些进展之上引入了稀疏化策略，优先处理机器人间回环闭合，显著提升了前端效率。最新的 DVM [10], 与本文同期发布，是首个在每个智能体上完全独立运行前端和后端的去中心化单目协同同步定位与建图系统。

A. 相对位姿估计

在协同同步定位与建图中生成全局一致地图，很大程度上依赖于准确进行相对位姿估计的能力，这涉及在来自不同机器人的两个关键帧之间进行三维配准。

单目设置中通常采用的几何方法 [11] 只能估计具有相对尺度的相对位姿，这意味着变换的绝对公制尺度仍然未知。虽然可以将基于图像的估计值与惯性测量单元等运动传感器结合起来，恢复连续图像之间的尺度信息 [12], 但这些传感器无法恢复非连续图像之间相对位姿的尺度。此外，与连续图像相比，非连续图像通常具有更大的视角差异，传统的基于关键点的匹配方法——例如依赖手工设计特征的方法——往往难以应对。近年来，单目深度估计取得了进展，例如 DPT [13], 为尺度估计提供了很有前景的替代方案。然而，尽管这些方法具有潜力，却通常需要针对特定领域进行微调，这可能会限制其在未知环境中建图时的泛化能力。

为克服视角限制，SuperPoint [14] 和 SuperGlue [15] 等基于学习的方法采用深度学习技术来增强关键点匹配鲁棒性，并通过对整幅图像进行推理来提升性能。该领域还推出了新的数据集和基准，例如 Map-Free 挑战赛 [16], 它聚焦于视角和光照变化剧烈的场景；要在这些场景中取得成功，需要涌现出 MicKey [17]和 DUSt3R [2] 等新技术。前者直接在三维相机空间中预测度量对应关系，而非通常采用的二维像素空间；后者则将图像匹配问题重新表述为成对三维重建任务，通过预测并对齐三维点图来估计相对位姿。在 DUSt3R 的基础上，MASt3R [3] 回归局部特征，并针对成对匹配进行显式训练。

在这些最新进展的基础上，MAST3R-SfM [18] 与 MAST3R-SLAM [19] 将这些三维基础模型集成到完整的建图与定位框架中。与传统运动恢复结构及同步定位与建图系统类似，这些方法通过选取关键帧而非处理全部图像来降低计算复杂度，并构建一个图，将具有重叠视点的关键帧相互连接——类似于同步定位与建图中常用的位姿图。然而，尽管这些方法在单机器人情形下非常有效，但对于多机器人同步定位与建图却效果较差，因为它们依赖于整体式全局优化，联合精化位姿与场景几何。这涉及集中式处理，即必须收集所有关键帧图像并在中央服务器上处理，从而需要与所有机器人持续进行高带宽通信。此外，在多机器人场景中，关键帧数量以及来自里程计和机器人间回环闭合的约束数量比单机器人增长得更快，使得利用大型模型进行实时机载计算变得不切实际。

相比之下，我们的方法专门针对去中心化系统的约束而设计。通过利用三维基础模型，仅凭一对图像即可实现稳健的机器人间回环闭合，从而避免机器人之间大规模的地图交换。我们的方法并不依赖单一整体式的全局优化，而是将计算开销高昂的编码步骤分散执行，并将优化任务委托给动态选出的机器人，遵循 [9], 中的策略，从而缓解去中心化运行所固有的主要通信与计算瓶颈。

III. 方法

我们的位姿图视觉协同建图与定位方法假设每个机器人都能获取具有正确公制尺度的既有定位源，例如由视觉惯性 [12], 、双目立体视觉 [20], 或基于激光雷达的 [21] 系统所获得的定位源。我们将机器人 α 的输入位姿估计定义为$T_{\alpha,0}, \ldots, T_{\alpha,n} \in$SE(3),其中$T_{\alpha,i}$ 表示机器人 α 沿其轨迹的 $i \in n$ 个关键帧中每一帧的位姿估计。除每个关键帧外，我们的方法还将相应的图像帧 I_i^α作为输入，用于机器人间回环检测。

A. 地点识别

我们处理流程的第一步是在不同机器人的地图之间进行地点识别。通过识别两个或更多机器人曾到访的位置，我们可以创建机器人间回环闭合，将各个机器人的位姿图连接起来，形成一致的共享定位估计。在地点识别方面，虽然 [18], [19]采用了由 MASt3R 编码的特征与聚合选择性匹配核 [22], ，但我们选择了 CosPlace [23]，因为它开箱即用时性能出色。借助它，我们无需进行任何标定即可使用预训练模型展示最先进的协同同步定位与建图，从而适应新的领域；这对于在非受控环境中部署而言是一项关键优势。

当两个或更多机器人处于通信范围内时，它们会相互交换紧凑的 CosPlace 描述子。机器人 α 收到邻近机器人发送的描述子后 β，会使用余弦相似度将其与自身的描述子进行比较。对于每一对关键帧（I_i^α，I_j^β），我们都会计算余弦相似度分数 $s^{\alpha,\beta}_{i,j}$。通过最近邻搜索确定每个关键帧的最佳匹配；如果最佳匹配的相似度超过阈值，则将相应的关键帧对（I_i^α，I_j^β）视为机器人间回环候选。随后，这些候选会被送入后续的配准步骤，以计算两个关键帧之间的相对位姿 $T^{\alpha,\beta}_{i,j}$。

B. 配准

以往的方法 [8], [9]依赖使用手工设计的图像特征进行三维配准，而这些特征对视角变化高度敏感。我们利用近期的基础模型进行三维图像匹配 [2], [3]。这些模型能够推断来自差异显著的视角的图像之间的相对位姿，从而检测到更多机器人间回环闭合。

传统的协同同步定位与建图方法则使用双目立体视觉、RGB-D、为通过激光雷达数据进行具有准确尺度的相对位姿估计，我们仅使用单目图像，并在后续阶段解决尺度歧义（见第 III-C 节）。对于配准，机器人在本地使用 MASt3R 模型经过预训练且冻结的 ViT编码器，将回环候选图像中的每张图像（I_i^α，I_j^β）编码为压缩潜向量。该向量之后可解码，用于重建场景并估计相对位姿，如图 1 所示。I_j^β 的编码会共享给机器人 α，用于解码、特征匹配和机器人间相对位姿推断。该过程会生成相对位姿测量 $T^{\alpha,\beta}_{i,j}$，以及两帧之间的特征对应数量。

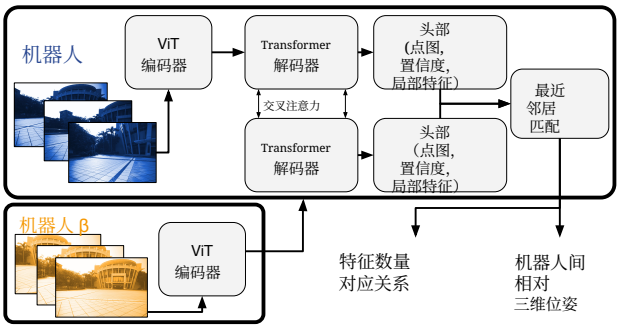


图 1：我们的推理流程示意图，突出展示了机器人间的回环闭合处理。

由于 MASt3R 偶尔会出现失败 [3],，我们提出了一种新颖的置信度估计指标。为计算该指标，机器人α 还会使用 MASt3R 对自身连续拍摄的一对图像（I^α_{i-1}，I^α_i）进行推理；这对图像具有可靠的里程计估计。随后，我们计算特征对应数量之比 $r^{\alpha,\beta}_{i,j}$，分子为机器人间回环图像对（I^α_i，I^β_j），分母为机器人内里程计图像对（I^α_{i-1}，I^α_i）：

$$r^{\alpha,i}_{\beta,j} = \frac{|(I^\alpha_i, I^\beta_j) \text{ feature correspondences}|}{|(I^\alpha_{i-i}, I^\alpha_i) \text{ feature correspondences}|} \tag{1}$$

由于里程计对通常具有较高的重叠度和大量对应关系，因此该比率 $r^{\alpha,\beta}_{i,j}$，能够有效地对回环闭合匹配的置信度进行归一化。

我们首先使用该比率筛除低于阈值 R_{thr} 的失败的注册；其次，在位姿图中用它为成功测量赋予权重。对于成功的回环闭合，我们使用逻辑函数将该比率映射为概率 $p^{\alpha,\beta}_{i,j}$：

$$p^{\alpha,i}_{\beta,j} = \left[1 + \exp(-k \cdot (r^{\alpha,i}_{\beta,j} - 1))\right]^{-1}. \tag{2}$$

概率及其参数 k 可以手动调整；如果事先有来自部署环境或类似环境的训练数据 [24]，也可以通过学习获得。

随后，将置信概率乘以与每个测量值关联的信息矩阵。信息矩阵在位姿图优化过程中充当权重，确保高置信度的回环闭合比低置信度的回环闭合对最终解有更大的影响。对于每个机器人间回环闭合，我们定义代价函数 $\phi^{\alpha,\beta}_{i,j}$：

$$\phi^{\alpha,i}_{\beta,j} = \left\| T^{-1}_{\alpha,i} \cdot T_{\beta,j} - \bar{T}^{\alpha,i}_{\beta,j} \right\|^2_{\Omega^{\alpha,i}_{\beta,j}(p^{\alpha,i}_{\beta,j})} \tag{3}$$

其中，$\bar{T}^{\alpha,\beta}_{i,j}$ 是相对位姿测量，$\Omega^{\alpha,\beta}_{i,j}$，是与置信度 $p^{\alpha,\beta}_{i,j}$ 成正比的测量信息矩阵。随后，可以将此代价函数与里程计约束一同纳入全局非线性最小二乘问题，以进行多机器人位姿图优化。

C. 多机器人位姿图与回环尺度优化

在我们的框架中，多机器人位姿图优化确保所有机器人的地图和估计轨迹在公共参考坐标系内对齐，从而提供环境的统一表示。该优化

问题被表述为因子图，其中变量表示未知量（例如，机器人位姿），而因子则定义这些变量子集上的函数。

a) 基础多机器人因子图：基础优化问题主要由两类因子组成：里程计因子，用于连接连续的关键帧位姿；以及回环因子，用于连接非连续的关键帧。机器人 α 的里程计代价函数定义如下：

$$\phi_{\alpha,i}^{\alpha,i-1} = \left\| T_{\alpha,i-1}^{-1} \cdot T_{\alpha,i} - \bar{T}_{\alpha,i}^{\alpha,i-1} \right\|_{\Omega_{\text{odom}}}^2 \tag{4}$$

其中，$\bar{T}^{\alpha,i-1}_{\alpha,i}$，表示相对位姿测量，$\Omega_{\text{odom}}$ 表示相应的信息矩阵。因此，对于多机器人优化问题，我们最小化所有代价函数之和：

$$\boldsymbol{T}^* = \underset{\boldsymbol{T}}{\operatorname{argmin}} \sum_{\alpha \in \mathcal{R}} \sum_{i \in (1:n_\alpha)} \phi_{\alpha,i}^{\alpha,i-1} + \sum_{(\alpha,i),(\beta,j) \in L_{\mathcal{R},\mathcal{R}'}} \phi_{\beta,j}^{\alpha,i} \tag{5}$$

其中，$\mathcal{R}$ 表示所有机器人的集合，n_α 表示机器人 α 的关键帧数量，$L_{\mathcal{R},\mathcal{R}'}$ 表示连接不同机器人的回环闭合集合，$\boldsymbol{T}$ 表示构成机器人轨迹的所有位姿集合。式（5）中的机器人集合 $\mathcal{R}$ 可以包含两个以上的机器人；事实上，我们使用所有相邻机器人进行优化，检测每一对机器人之间的回环闭合，并将其纳入优化问题。此外，我们不考虑机器人内回环 $\phi^{\alpha,i}_{\alpha,k}$，因为我们的目标是分离机器人间回环闭合的影响；不过，也可以将其作为额外的代价函数，直接纳入该公式。

MASt3R测得的相对位姿面临的挑战在于，其尺度往往不一致或非常不精确 [18], [19]。因此，为了将这些测量值正确地整合到优化问题中，我们必须预先借助其他传感器确定尺度，或者像本文所提出的那样，将回环尺度视为优化变量。一个简单的方法是利用两个里程计位姿之间经过正确尺度校准的里程计测量值来估计回环尺度。假设里程计和回环使用相同的缩放因子：

$$\bar{t}_{\beta,j}^{\alpha,i} = \frac{\left\| \bar{t}_{\alpha,i}^{\alpha,i-1} \right\|}{\left\| t_{\alpha,i}^{\alpha,i-1} \right\|} \cdot t_{\beta,j}^{\alpha,i} \tag{6}$$

其中，$t^{\alpha,i-1}_{\alpha,i}$，和 $t^{\alpha,i}_{\beta,j}$，是 MASt3R 输出的相对平移，而 $\bar{t}^{\alpha,i-1}_{\alpha,i}$ 是已知的里程计平移。虽然我们无法保证回环和里程计采用相同的缩放因子，但这通常能提供一个合理的初始猜测，详见 IV-C节。

b) 独立尺度多机器人因子图：本文进一步提出对回环闭合的尺度进行显式优化。为此，我们借鉴了 [25],的思路；该方法为通过 IMU航位推算估计、并利用零星的超宽带距离测量进行优化的行人轨迹引入了缩放因子。虽然这项工作侧重于重新缩放里程计估计，但我们将缩放应用于非连续的相对位姿测量。

首先，我们将 MASt3R 测得的、与回环闭合相关联的相对位姿$\bar{T}^{\alpha,i}_{\beta,j} \in \text{SE(3)}$分解为 $\bar{R}^{\alpha,i}_{\beta,j} \in \text{SO(3)}$和 $\bar{t}^{\alpha,i}_{\beta,j} \in \mathbb{R}^3$，$s^{\alpha,i}_{\beta,j}$。其中，$\bar{R}^{\alpha,i}_{\beta,j}$ 是测得的相对旋转矩阵，$\bar{t}^{\alpha,i}_{\beta,j}$ 是两个位姿之间测得的平移向量，而 $s^{\alpha,i}_{\beta,j}$ 用于将平移向量的大小调整到正确的尺度。经尺度调整后的相对位姿 $\hat{T}^{\alpha,i}_{\beta,j}$ 为：

$$\hat{T}_{\beta,j}^{\alpha,i} = \begin{bmatrix} \bar{R}_{\beta,j}^{\alpha,i} & s_{\beta,j}^{\alpha,i} \cdot \bar{t}_{\beta,j}^{\alpha,i} \\ 0 & 1 \end{bmatrix} \tag{7}$$

然后，我们定义了一个纳入尺度值的新回环代价函数：

$$\hat{\phi}_{\beta,j}^{\alpha,i} = \left\| T_{\alpha,i}^{-1} \cdot T_{\beta,j} - \hat{T}_{\beta,j}^{\alpha,i} \right\|_{\Omega_{\beta,j}^{\alpha,i}(p_{\beta,j}^{\alpha,i})}^2 \tag{8}$$

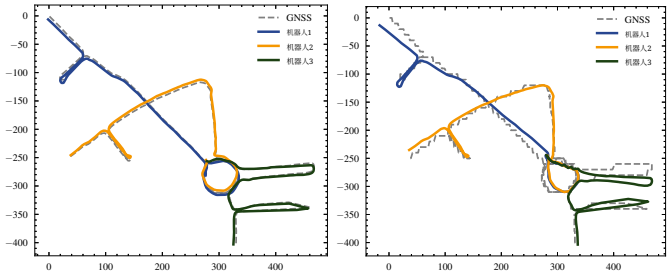
为了优化这一新型因子，我们需要提供相应的导数。具体而言，由于我们使用 GTSAM [26]作为因子图优化框架，因此提供了解析测量雅可比矩阵，以便高效计算；优化过程中，这些矩阵在当前估计值处的切空间中求得：$\boldsymbol{H}_{T_{\alpha,i}} = -\text{Adj}(\text{inv}(T_{\alpha,i}^{-1} \cdot T_{\beta,j}))$, $\boldsymbol{H}_{T_{\beta,j}} = \boldsymbol{I}$和 $\boldsymbol{H}_s = [\boldsymbol{0} - \bar{t}^{\alpha,i}_{\beta,j}]^\top$。在式（5）中用$\hat{\phi}^{\alpha,i}_{\beta,j}$ 代替 $\phi^{\alpha,i}_{\beta,j}$，即可推导出优化问题，其中每个回环的尺度都独立处理。

c) 平滑尺度多机器人因子图：我们还提出了第三种建模方式，其中加入了一个额外因子，用于关联相关回环闭合的尺度因子，确保它们保持相近。我们将关联相距不到十个关键帧的相对位姿的回环闭合归为同一簇。当机器人长时间重访某个特定区域时，通常会出现这类簇；此时地图之间存在一个较大的重叠区域，因而检测到多个回环闭合。先前的研究探讨了如何识别此类簇，以检测回环测量中的离群值 [27]。在我们的情形中，将它们关联起来的直觉依据是：由于 MASt3R 是数据驱动的，从相似图像域推断出的相对位姿可能具有相似的尺度因子。

用于关联尺度值的代价函数及其对应的雅可比矩阵定义为 $\phi_{s_i,s_j} = \|s_j - s_i\|^2_{\Omega_s}$、$\boldsymbol{H}_{s_i} = -1$ 和 $\boldsymbol{H}_{s_j} = 1$。其中，尺度残差 ϕ_{s_i,s_j}定义为两个不同回环闭合的尺度估计值 s_i 和s_j 之差，从而确保回环闭合簇中的尺度一致。对应的雅可比矩阵 $\boldsymbol{H}_{s_i}$ 和 $\boldsymbol{H}_{s_j}$ 分别是 ϕ_{s_i,s_j} 关于 s_i 和 s_j 的偏导数。为完整起见，我们在实验分析中也考虑了所有回环闭合共享同一尺度值的情形。我们的公式设计支持去中心化执行。在 [9], 优化之后，只要一组智能体处于通信范围内，便由动态选举的机器人执行优化，并将所得估计值广播给相邻智能体。针对我们的尺度感知公式设计完全分布式的求解器，是未来值得探索的方向。

IV. 实验

我们使用机器人机载计算机 NVIDIA Jetson AGX Xavier 在数据集序列上进行了实验



(a) FLOCC-SLAM (b) Swarm-SLAM 激光雷达

图 2：采用激光雷达配置时，S3E宿舍中 FLOCC-SLAM 与 Swarm-SLAM 的轨迹估计值。

（32GB）。我们在两种配置下将 FLOCC-SLAM 与开源的 Swarm-SLAM [9]进行基准测试：双目立体视觉和激光雷达。在双目配置下，使用 CosPlace [23],进行地点识别，并使用 PnP 和 RANSAC [20]进行位姿配准。在激光雷达配置下，使用 ScanContext [28],进行地点识别，并使用 TEASER++ [29]进行位姿配准。我们选择 Swarm-SLAM 作为强基线，因为它利用三维感知，使我们能够证明，我们仅依赖单目数据的闭环方法无需依赖三维输入数据，也能达到相当的精度。

为进行全面且公平的评估，我们还为 Swarm-SLAM 配置了几种后端优化器：快速的 Levenberg-Marquardt求解器（ LM） [26],、稳健的 GNC 优化器 [6], （可检测并拒绝回环闭合中的离群值），以及黎曼块坐标下降（RBCD） [8], ，这是一种集成 GNC 以排除离群值的分布式求解器。值得注意的是，采用双目立体视觉模式和 RBCD 的 Swarm-SLAM 与 Kimera-Multi [8], 十分相似，区别仅在于地点识别——Kimera-Multi 使用 DBoW2 [30], ，而 Swarm-SLAM 采用 CosPlace [23]。

对于 FLOCC-SLAM，我们测试了三种不同的因子图形式：基础多机器人位姿图、具有独立尺度值的形式（**IS**），以及具有平滑尺度值的形式（**SS**）。除非另有说明，否则均采用基础形式。由于基础形式不优化尺度，因此仅依赖式（6）中提出的尺度初始化技术。对于地点识别，我们将相似度阈值 $s^{\alpha,\ \beta}_{i\ j}$，设置为一个宽松的 0。1，因为我们的处理流程能够在后续配准阶段稳健地处理误报。

我们首先在 S3E 数据集 [31],上评估这些技术，该数据集包含三台在大型、动态的室内和室外环境中导航的机器人。我们选取了最具挑战性的完整序列，这些序列的重叠最少且视角各异。此外，我们还在 GrAco 数据集 [32],上进行了测试，该数据集涉及六台在大型室外环境中运行的机器人。选择这些数据集，是因为它们的轨迹具有挑战性，且与我们的处理流程兼容。我们使用 GrAco 数据集进行了一项消融研究。评估时，我们使用 *Evo* 软件包 [33], 将结果与提供的真值进行比较：高精度全球导航卫星系统数据（用于室外路段，报告的精度为±1厘米 [31], [32]）和动作捕捉系统（用于室内路段，精确率达到毫米级 [31]）。在我们的实验中，所有技术均使用相同的里程计估计，以便隔离地图合并的影响。

A. 完整系统性能

在图 2 中，我们展示了 FLOCC-SLAM 在具有挑战性的 S3E 宿舍序列上的表现。与 Swarm-SLAM 基线相比，FLOCC-SLAM 的结果与真值更加贴合。表 I 中的详细结果为这些可视化提供了佐证，其中报告了平均平移误差 (ATE)、回环闭合数量N，以及各方法在四个序列上所需的因子图求解器计算时间。

FLOCC-SLAM 使用由**激光测距** 求解器优化的平滑尺度构型（**SS**）时，取得了最佳性能。这种方法生成的回环闭合数量比其他方法高出数个数量级，显著提升了最终解的精度。为识别成功匹配，我们采用了对应关系比率阈值 R_{thr} 0.3。尽管部分离群点可能通过筛选，但基于比率的置信度机制（见式 (2)）能够减轻其影响，无需使用 GNC 等计算开销更大的鲁棒求解器。虽然只需少量准确的回环闭合即可完成地图合并，但我们的方法能够生成更多回环闭合，这是一项显著优势，尤其是在潜在回环闭合稀少的场景中。此外，直接处理尺度歧义，而不依赖近似尺度估计值或代价高昂的离群点剔除机制，也显著提升了性能。

采用平滑尺度（SS）的 FLOCC-SLAM 精度提升表明，簇内的回环尺度可能确实存在相关性。这使得相对于 IS 所增加的计算时间——如在宿舍序列等情形中所观察到的——在资源允许时成为一种值得的权衡。值得注意的是，在 S3E 宿舍序列上，双目立体视觉 Swarm-SLAM 配置无法合并全部三条轨迹，原因是使用传统立体匹配技术难以从对向视角检测回环闭合。同样，Swarm-SLAM 的激光雷达配置在该序列上表现不佳，原因是结构重复的宿舍走廊中存在显著的感知混淆。

为进一步分析我们所提出的方法，以下各小节给出了我们开展的一项消融研究，用以评估各个新颖组件的有效性与影响。

B. 相对位姿估计准确性与鲁棒性

在我们的第一项消融实验中，我们使用了涉及六台机器人的 GrAco 数据集 [32], ，并根据 GNSS 数据提取了相距不足 10 米的 88,096 个图像对。随后，我们对每一对应用配准流程，以获得相对位姿。

我们首先研究的问题是如何确定匹配的有效性。由于感知混淆，地点识别有时会失败，错误地将不同地点识别为同一地点。DUSt3R [2],以及随后出现的MASt3R [3],进一步加剧了这一挑战，因为它们如今能够以较高置信度匹配几乎任何视角的图像，因此更难判断哪些图像匹配对应有效的相对位姿，并应作为机器人间回环

表 I：平均平　　移误差（绝对轨迹误差）、回环闭合数量（N）以及 S3E序列上的优化时间　　s.

		校园			教学			广场			宿舍		
		绝对轨迹误差（米）	N	时间（秒）	绝对轨迹误差（米）	N	时间（秒）	绝对轨迹误差（米）	N	时间（秒）	绝对轨迹误差（米）	N	时间（秒）
FLOCC-SLAM	GNC	19.25 ± 6.81	1122	27.45	2.40 ± 0.76	2274	41.58	9.53 ± 3.87	226	9.25	46.00 ± 30.36	175	34.79
	IS-LM	5.50 ± 2.16	1122	2.46	1.87 ± 0.65	2273	3.46	5.64 ± 2.46	226	1.41	3.84 ± 1.35	175	1.50
单目闭环 SS-LM		**5.41 ± 2.13**	1122	1.82	**1.84 ± 0.67**	2273	4.74	**5.07 ± 2.56**	226	1.63	**2.41 ± 0.99**	175	11.05
Swarm-SLAM	RBCD	9.41 ± 7.59	29	286.63	3.59 ± 1.56	110	143.31	113.58 ± 80.99	6	120.33	7.41 ± 5.32	226	147.80
	GNC	10.50 ± 9.47	29	21.38	3.69 ± 1.52	110	13.97	164.83 ± 99.39	6	36.45	10.18 ± 5.47	226	26.14
激光雷达回环闭合	LM	10.21 ± 8.97	29	1.23	3.85 ± 1.63	110	0.99	165.94 ± 98.98	6	1.73	8.34 ± 4.82	226	4.02
Swarm-SLAM	RBCD	6.79 ± 7.08	16	744.65	2.23 ± 0.94	27	840.53	42.97 ± 25.27	7	2.46		✗	
	GNC	6.56 ± 7.07	16	48.64	2.25 ± 1.15	27	23.85	9.73 ± 6.67	7	15.34		✗	
立体视觉闭环	LM	6.56 ± 7.07	16	2.90	2.20 ± 1.05	27	3.91	33.37 ± 22.98	7	1.27		✗	

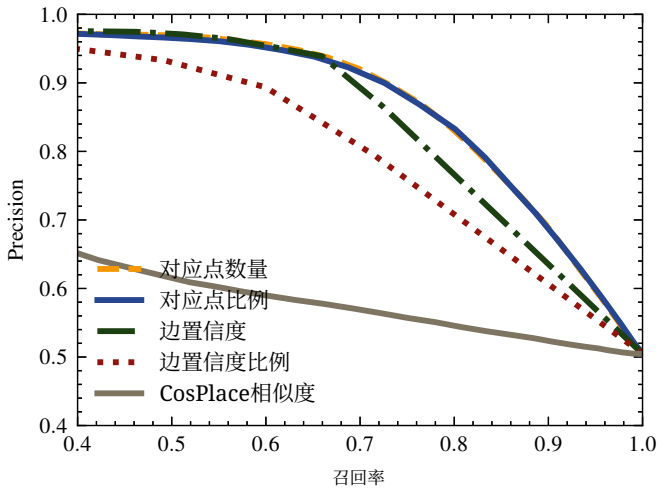


图 3：不同内点/外点检测技术的精确率-召回率曲线。基于 MASt3R 的对应关系结合回环与里程计比率，可获得最佳性能。

闭合纳入位姿图中。在图 3 中，我们评估了五种不同指标的精确率和召回率。如果某匹配相对于真值的平移误差超过 2 米，则将其视为离群值；测量值根据真值数据进行尺度缩放。我们发现，仅依赖 CosPlace 相似度不足以区分配准内点和离群值。作为替代方案，我们考虑了基于 MASt3R 输出的指标，包括 MASt3R 生成的平均置信度，以及回环匹配与里程计匹配之间的置信度比率。我们还考察了特征对应的数量及对应点比例。结果表明，对应点数量和对应点比例都表现出色；由于比例不带单位，相较于受图像尺寸和视场影响的对应点数量，它更便于调节。

C. 多机器人尺度与位姿图优化

在以下实验中，我们评估了涉及完整 6 机器人 GrAco 序列的完整位姿图解。我们将特征对应点比例映射为置信度指标（见式（2））， 以便在位姿图优化期间对机器人间回环闭合进行加权。

在表 II 和表 III 中，我们分析了不同因子图形式的影响，以及里程计骨干的作用。我们测试了两种里程计骨干：VINS-Mono [12], , 使系统完全采用单目方案；以及精度更高的基于激光雷达的 LIO [21], , 以展示当里程计的尺度近乎完美时我们方法的有效性。尽管单目里程计取得了合理的结果，但轨迹不同部分的尺度误差——尤其是起始阶段在线尺度估计尚未收敛时——导致性能下降。由于 GrAco 数据集缺少视觉惯性里程计标定序列，所有机器人轨迹都受到这些初始尺度误差的影响。在实际场景中，适当的标定对于有效部署视觉惯性里程计仍至关重要。我们的方法与基于 LIO 的方案相结合，取得了出色的精度：在 3.5 公里的轨迹中，误差低于 4 米。

对于每种骨干网络，我们比较了先前介绍的各种形式。为完整起见，我们还针对每种因子图形式报告了非鲁棒 LM 求解器与鲁棒 GNC求解器的结果。此外，我们同时报告了在机器人机载计算机以及配备 AMD Ryzen 7 3700X CPU 和 NVIDIA RTX 3070 GPU 的台式服务器上的计算时间。鉴于一些其他现有的协同同步定位与建图方法将状态估计卸载到服务器 [34], 这一比较颇具意义。需要注意的是，位姿图优化在 CPU 上运行，而地点识别与配准均在图形处理器上执行。我们为每项结果提供了不同的尺度初始化技术，包括使用真值、直接使用 MASt3R 的原始输出，或采用我们基于里程计的尺度估计（见式 (6)）。表 II 和表 III 中的结果证实，采用独立尺度或平滑尺度的形式取得了最佳或接近最佳的性能，且无需 GNC 代价高昂的鲁棒优化。这表明使用鲁棒求解器带来的收益有限。事实上，当与平滑尺度形式结合时，它甚至可能降低性能，这可能是因为 GNC 的外点剔除机制与我们基于置信度的加权相冲突。与 GNC 所需的数分钟相比，我们的优化在服务器上不到一秒即可完成，在机载计算机上仅需数秒。

表 II：比较　　采用 VINS-Mono 里程计骨干的不同位姿图优化方法　　.

	真值尺度			MASt3R尺度			里程计尺度		
	绝对轨迹误差（米）	在板时间（秒）	服务（秒）	绝对轨迹误差（米）	在板时间（秒）	服务（秒）	绝对轨迹误差（米）	在板时间（秒）	服务（秒）
LM	4.76 ± 4.50	1.60	0.33	39.93 ± 44.43	3.63	0.91	39.59 ± 45.68	3.27	0.79
GNC	4.89 ± 4.51	31.09	10.28	11.98 ± 5.21	32.93	10.49	5.56 ± 4.07	35.86	11.78
IS-LM	6.99 ± 6.47	3.27	0.79	7.11 ± 6.81	3.24	0.78	7.89 ± 7.01	3.49	0.87
IS-GNC	6.34 ± 6.51	104.21	29.65	6.40 ± 6.82	96.54	27.38	7.86 ± 7.01	112.94	31.98
SS-LM	7.43 ± 6.93	3.89	0.94	8.23 ± 7.34	3.50	0.82	7.42 ± 7.06	3.95	0.94
SS-GNC	7.45 ± 7.06	166.24	45.99	91.50 ± 59.82	296.27	85.31	58.94 ± 62.61	305.20	85.38

表 III：Co　　采用 LIO 里程计骨干的不同位姿图优化形式比较　　.

	真值尺度			MASt3R尺度			里程计尺度		
	绝对轨迹误差（米）	在板时间（秒）	服务（秒）	绝对轨迹误差（米）	在板时间（秒）	服务（秒）	绝对轨迹误差（米）	在板时间（秒）	服务（秒）
LM	2.75 ± 1.36	1.44	0.29	10.36 ± 3.65	1.44	0.30	6.90 ± 3.95	1.51	0.30
GNC	2.73 ± 1.36	29.61	9.58	9.25 ± 2.98	29.12	9.56	2.95 ± 1.32	30.71	9.89
IS-LM	3.45 ± 2.03	3.16	0.74	3.14 ± 1.80	2.83	0.66	3.41 ± 1.96	3.01	0.70
IS-GNC	3.45 ± 2.00	93.70	26.45	3.12 ± 1.77	87.59	24.63	3.40 ± 1.94	94.17	26.31
SS-LM	2.92 ± 1.93	4.06	0.95	3.29 ± 1.94	3.42	0.80	3.11 ± 1.81	3.85	0.84
SS-GNC	2.93 ± 1.90	156.52	41.42	3.23 ± 1.92	139.69	37.32	13.89 ± 11.23	342.36	89.20

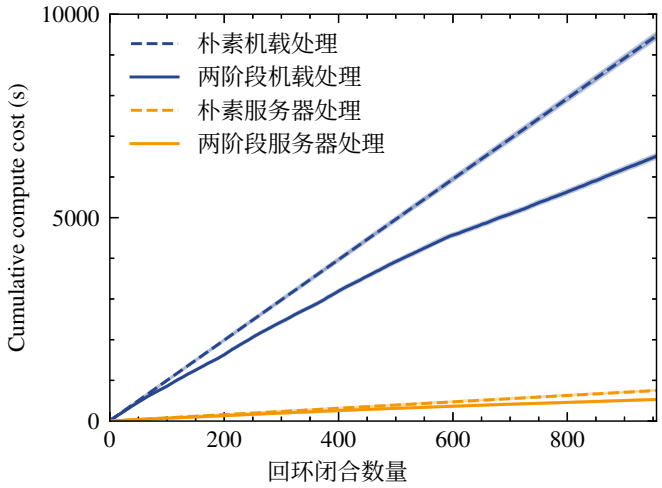


图 4：在协同同步定位与建图中，采用 MASt3R 两阶段处理所节省的计算量。将编码器和解码器处理分开，可避免对已建立回环闭合的图像重复编码，从而减少计算量。这对机载计算机尤其有益，因为 MASt3R 处理对其计算资源的要求远高于桌面服务器。

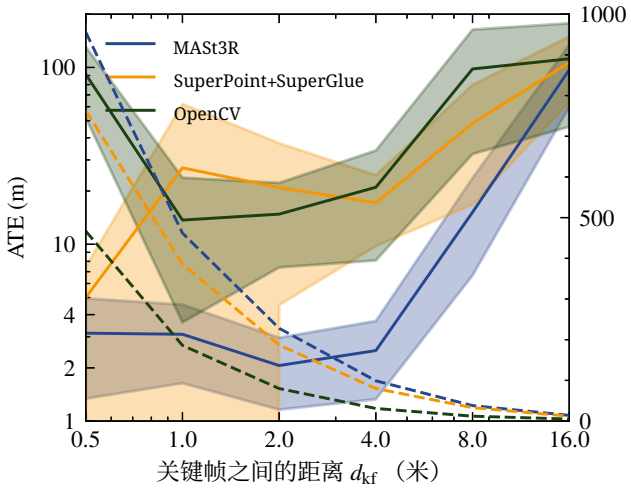


图 5：平均轨迹误差（绝对轨迹误差）随关键帧距离的变化，每种配准技术以不同颜色表示。实线表示绝对轨迹误差，阴影区域表示标准差；虚线表示检测到的回环闭合数量。增大关键帧距离会降低内存使用量，但通常会降低精度。即使关键帧更为稀疏，我们的方法仍能保持更低的误差。

D. 资源效率

除了与鲁棒方法相比我们的尺度感知优化所带来的计算效率之外，我们的方法还具有进一步的效率优势。具体而言，我们将配准分为两个阶段：参与该回环的两台机器人各自在本地对关键帧进行编码，而解码仅在其中一台机器人上完成，如图 1 所示。这样，属于多个回环闭合的关键帧只需编码一次——在管理大量回环闭合候选时尤为有用。在图 4 中，我们展示了在 GrAco 数据集上相对于一种对所有候选都对两幅图像进行编码的 na¨ıve 基线所获得的计算增益。可以看到，与.....相比，MASt3R 推理在机器人机载计算机上（蓝色）代价相当高

使用配备更强大图形处理器的服务器（橙色）。这与我们在表 II 中对位姿图优化的观察一致

以及 III，其中机载计算速度约比服务器慢 3 到 4 倍。因此，通过采用更先进的机器人与服务器之间的负载分担策略，还能进一步提升性能。减少需要处理并存入内存的关键帧数量，也能提高资源效率。不过，减少关键帧通常会降低精度，因为建图会变得更加粗略，能够检测到的回环闭合也会减少。在图 5 中，我们评估了增大关键帧之间的距离（从而按比例减少关键帧数量）对解的绝对轨迹误差的影响。我们还将其与使用 SuperPoint+SuperGlue 和 OpenCV 计算的回环进行了比较。随着关键帧距离增大，其他方法的性能会迅速下降，而我们的方法即使在距离达到 4 米时，仍能保持较低的绝对轨迹误差。这表明它

能够支持更稀疏、内存效率更高的协同同步定位与建图方案，同时不牺牲精度。采用更稀疏关键帧所带来的效率提升不仅体现在节省内存上：还可以通过地点识别和配准，以更低的通信带宽、更短的时间来比较稀疏地图。高精度稀疏地图适用于部署在小型机器人或消费电子产品上。

V. 结论

本文介绍了 FLOCC-SLAM，这是一种新颖的去中心化协同同步定位与建图方法，利用三维基础模型解决多机器人回环检测所面临的挑战。通过结合单目位姿估计与尺度优化，FLOCC-SLAM 有效提升了机器人间回环的计算效果，尤其适用于当前方法因机器人视角差异较大而难以应对的场景。实验结果和消融研究表明，FLOCC-SLAM 的表现优于最先进方法，尤其在准确性方面。我们还证明，当具备尺度正确的高精度里程计时，可以轻松、高效地将尺度恢复回环纳入多机器人位姿图优化。然而，我们的方法依赖单机器人系统提供的度量尺度里程计，以锚定全局地图的尺度。此外，该方法假设里程计足够稳定，能够为置信度指标提供可靠的基准。尽管预训练 MASt3R 模型展现出令人瞩目的泛化能力，但在与训练数据差异极大的领域（例如水下图像）中，性能可能会下降，因此需要进行微调。

展望未来，三维基础模型与协同同步定位与建图系统有望实现更紧密的集成，从而提高测量精度，并改善对已探索环境的整体表示。创建更多样化的大规模协同同步定位与建图数据集，也将极大推动该领域的发展；这些数据集应包含不同类型的传感器，以及具有挑战性且贴近现实的多机器人轨迹。

参考文献

[1] S. Lowry 等，“视觉地点识别：综述”，*IEEE*机器人学汇刊，第 32 卷，第 1 期，第 1–19 页，2016 年。[2] S. Wang、V. Leroy、Y. Cabon、B. Chidlovskii 和 J. Revaud，“DUSt3R：让几何三维视觉变得简单”，载于 *CVPR*，2024 年，第 20 697–20 709 页。[3] V. Leroy、Y. Cabon 和 J. Revaud，“利用 MASt3R 将图像匹配建立在三维基础上”，载于*ECCV*，2024 年，第 71–91 页。[4] C. Cadena 等，“同步定位与建图的过去、现在与未来：迈向鲁棒感知时代”，*IEEE*机器人学汇刊，第 32 卷，第 6 期，第 1309–1332 页，2016 年。[5] J. G. Mangelson、D. Dominic、R. M. Eustice 和 R. Vasudevan，“用于鲁棒多机器人地图合并的成对一致测量集最大化”，载于 *ICRA*，2018 年，第 2916–2923 页。[6] H. Yang、P. Antonante、V. Tzoumas 和 L. Carlone，“鲁棒空间感知的渐进非凸优化：从非最小求解器到全局离群点剔除”，*IEEE*机器人与自动化快报，第 5 卷，第 2 期，第 1127–1134 页，2020 年。[7] T. Cieslewski、S. Choudhary 和 D. Scaramuzza，“数据高效的去中心化视觉SLAM”，载于*ICRA*，2018 年，第 2466–2473 页。[8] Y. Tian、Y. Chang、F. Herrera Arias、C. Nieto-Granda、J. P. How 和 L. Carlone，“Kimera-Multi：面向多机器人系统的鲁棒分布式稠密度量语义 SLAM”，*IEEE*机器人学汇刊，第 38 卷，第 4 期，第 2022–2038 页，2022 年。[9] P.-Y. Lajoie 和 G. Beltrame，“Swarm-SLAM：面向多机器人系统的稀疏去中心化协同同时定位与建图框架”，IEEE机器人与自动化快报，第 9 卷，第 1 期，第 475–482 页，2024 年。[10] J. Bird、J. Blumenkamp 和 A. Prorok，“DVM-SLAM：面向多智能体系统的去中心化视觉单目同时定位与建图”，arXiv 预印本 arXiv:2503.04126，2025 年。[11] R. Hartley 和 A. Zisserman，《计算机视觉中的多视图几何》，第 2 版。美国：剑桥大学出版社，2003 年。[12] T. Qin、P. Li 和 S. Shen，“VINS-Mono：一种鲁棒且通用的单目视觉惯性状态估计器”，IEEE机器人学汇刊，第 34 卷，第 4 期，第 1004–1020 页，2018 年。[13] R. Ranftl、A. Bochkovskiy 和 V. Koltun，“用于密集预测的视觉变换器”，载于 ICCV，2021 年，第 12 159–12 168 页。[14] D. DeTone、T. Malisiewicz 和 A. Rabinovich，“SuperPoint：自监督兴趣点检测与描述”，载于 CVPRW，2018 年，第 337–33 712 页。[15] P.-E. Sarlin、D. DeTone、T. Malisiewicz 和 A. Rabinovich，“SuperGlue：利用图神经网络学习特征匹配”，载于 CVPR，2020 年，第 4937–4946 页。[16] E. Arnold 等，“无地图视觉重定位：相对于单幅图像的度量位姿”，载于 ECCV，2022 年，第 690–708 页。[17] A. Barroso-Laguna、S. Munukutla、V. A. Prisacariu 和 E. Brachmann，“在三维空间中匹配二维图像：从度量对应关系估计度量相对位姿”，载于 CVPR，2024 年，第 4852–4863 页。[18] B. Duisterhof、L. Zust、P. Weinzaepfel、V. Leroy、Y. Cabon 和 J. Revaud，“MASt3R-SfM：面向无约束运动恢复结构的完整集成解决方案”，arXiv 预印本 arXiv:2409.19152，2024 年。[19] R. Murai、E. Dexheimer 和 A. J. Davison，“MASt3R-SLAM：基于三维重建先验的实时稠密SLAM”，arXiv 预印本 arXiv:2412.12392，2024 年。[20] M. Labb´e 和 F. Michaud，“RTAB-Map：面向大规模和长期在线运行的开源激光雷达与视觉同时定位与建图库”，《现场机器人学》期刊，第 36 卷，446，2019 年 2，416第。页–。[21] T. Shan、B. Englot、D. Meyers、W. Wang、C. Ratti 和 D. Rus，“LIO-SAM：通过平滑与建图实现紧耦合激光雷达惯性里程计”，载于 IROS，2020 年，第 5135–5142 页。[22] G. Tolias、T. Jenicek 和 O. Chum，“用于实例级识别的深度局部描述子的学习与聚合”，载于 ECCV，2020 年，第 460–477 页。[23] G. Berton、C. Masone 和 B. Caputo，“重新思考面向大规模应用的视觉地理定位”，载于 CVPR，2022 年，第 4878–4888 页。[24] Y. Tian、K. Khosoussi 和 J. P. How，“一种具有可证明性能保证的资源感知协同闭环检测方法”，《国际机器人研究杂志》，第 40 卷，第 10-11 期，第 1161–1180 页，2020 年。[25] P.-Y. Lajoie、B. H. Baghi、S. Herath、F. Hogan、X. Liu 和 G. Dudek，“PEOPLEx：利用惯性测量单元、超宽带、蓝牙低功耗和 Wi-Fi 的行人机会式定位”，载于国际通信会议，2024 年，第 3518–3523 页。[26] F. Dellaert 等，“佐治亚理工学院平滑与建图（GTSAM）”，http://gtsam.org/，2020 年。[27] F. Wu 和 G. Beltrame，“用于鲁棒位姿图优化的基于聚类的惩罚缩放”，IEEE机器人与自动化快报，第 5 卷，第 4 期，第 6193–6200 页，2020 年。[28] G. Kim 和 A. Kim，“扫描上下文：用于三维点云地图内地点识别的自我中心空间描述子”，载于 IROS，2018 年，第 4802–4809 页。[29] H. Yang、J. Shi 和 L. Carlone，“TEASER：快速且可验证的点云配准”，IEEE机器人学汇刊，第 37 卷，333，2021 年 2，314第。页–。[30] D. Galvez-L´opez 和 J. D. Tardos，“用于图像序列快速地点识别的二进制词袋”，IEEE机器人学汇刊，第 28 卷，第 5 期，第 1188–1197 页，2012 年。[31] D. Feng，“S3E：面向协同SLAM的多机器人多模态数据集”，2024 年。[32] Y. Zhu、Y. Kong、Y. Jie、S. Xu 和 H. Cheng，“GRACO：面向地面与空中协同定位与建图的多模态数据集”，IEEE机器人与自动化快报，第 8 卷，第 2 期，第 966–973 页，2023 年。[33] M. Grupp，“Evo：用于评估里程计和SLAM的Python软件包”，2017 年。[34] P. Schmuck、T. Ziegler、M. Karrer、J. Perraudin 和 M. Chli，“COVINS：用于集中式协作的视觉惯性SLAM”，载于 ISMAR-Adjunct，2021 年，第 171–176 页。