回环闭　　在 DPV-SLAM 中重新使用 AnyLoc 视觉地点识别

张文正*、足立和树*、原义隆†、中村宗介‡

* 日本东京法政大学科学与工程研究生院　† 日本千叶千叶工业大学未来机器人技术中心（fuRo）　‡ 日本东京法政大学理工学部

arXiv:2601.02723v1 [cs.RO] 6 Jan 2026

摘要——回环闭合对于维持视觉SLAM的准确性和一致性至关重要。我们提出了一种提升 DPV-SLAM 回环闭合性能的方法。该方法将基于学习的视觉地点识别技术 AnyLoc 集成到系统中，以取代经典的视觉词袋（BoVW）回环检测方法。与依赖手工设计特征的 BoVW 不同，AnyLoc 利用深度特征表示，能够在多种视角和光照条件下实现更鲁棒的图像检索。此外，我们还提出了一种自适应机制，可根据环境条件动态调整相似度阈值，无需手动调参。在室内外数据集上的实验表明，所提方法在回环闭合准确率和鲁棒性方面均显著优于原始 DPV-SLAM。该方法为提升现代同步定位与建图系统的回环闭合性能提供了一种实用且可扩展的解决方案。

I. 引言

同步定位与建图（SLAM）是实现移动机器人系统自主导航的一项基础技术。在各种 SLAM 模态中，视觉 SLAM 仅依靠视觉输入（通常来自单目或双目相机），即可实现机器人定位和地图构建。

在单目视觉 SLAM 中，随着时间推移，累积漂移和尺度不一致不可避免地会出现，可能会损害所构建地图的全局一致性。回环闭合是缓解这些问题的关键过程，它通过识别先前访问过的位置，随后对地图进行优化。

在真实环境中，移动机器人常常面临光照变化、动态遮挡和显著视角变化等严峻条件。这些因素对视觉回环检测的鲁棒性提出了严格要求。在此类情况下，传统的视觉词袋（BoVW）方法尤其容易受到影响，常常导致错误匹配或漏检，进而降低同步定位与建图系统的精度和一致性。

近年来，基于学习的视觉特征表示取得了长足进展，由此催生了超越 BoVW 等经典方法的新型回环检测方法。具体而言，AnyLoc [1], 是一种视觉地点识别技术，通过大规模预训练实现了出色的泛化能力和语义区分能力；与传统方法相比，它对环境变化表现出更强的鲁棒性。

与依赖 ORB [2] 或 SIFT [3] 等手工设计的局部特征以及人工构建的词汇表的经典 BoVW 方法不同，AnyLoc 利用深度神经网络提取鲁棒且具有丰富语义信息的图像表示。其预训练策略有助于提升在不同环境和时间变化下的泛化能力，使其非常适合在真实环境中进行回环检测。

在我们此前的工作中 [4], ，我们评估了若干单目视觉 SLAM 系统，并发现 Deep Patch 视觉 SLAM（DPV-SLAM） [5]表现出色。DPV-SLAM 是一种基于 Deep Patch 视觉里程计（DPVO） [6],扩展而来的单目视觉 SLAM 方法，它高效利用深度特征表示，以低内存消耗实现实时运行。尽管 DPV-SLAM 在其视觉里程计模块中采用了深度学习，但在回环检测方面仍依赖经典的基于 BoVW 的方法。

本文提出以 AnyLoc 替换 DPV-SLAM 中基于 BoVW 的回环检测，从而增强其回环闭合模块。我们还提出一套完整的回环闭合流程，将自适应相似度阈值调整与几何验证相结合。我们使用真实世界相机数据对所提方法进行评估，以验证其有效性。

我们的贡献如下：

- 我们将基于学习的视觉地点识别方法 AnyLoc 集成到 DPV-SLAM 的回环闭合模块中。
- 我们提出一种自动阈值调整机制，可根据不同环境调整相似度阈值。
- 实验表明，在回环闭合准确率和鲁棒性方面，我们的方法优于原始 DPV-SLAM。

II. 相关工作

与所提方法相关的研究大致可分为三类：基于视觉词袋（BoVW）的回环检测方法、利用基于学习的视觉特征的回环检测方法，以及利用拓扑和语义信息的回环检测方法。

在许多传统视觉 SLAM 系统中，回环检测通常依赖 BoVW 方法，例如 DBoW2[7]。这些方法将 ORB 或 SIFT 等局部特征量化为视觉词，从而实现快速图像检索

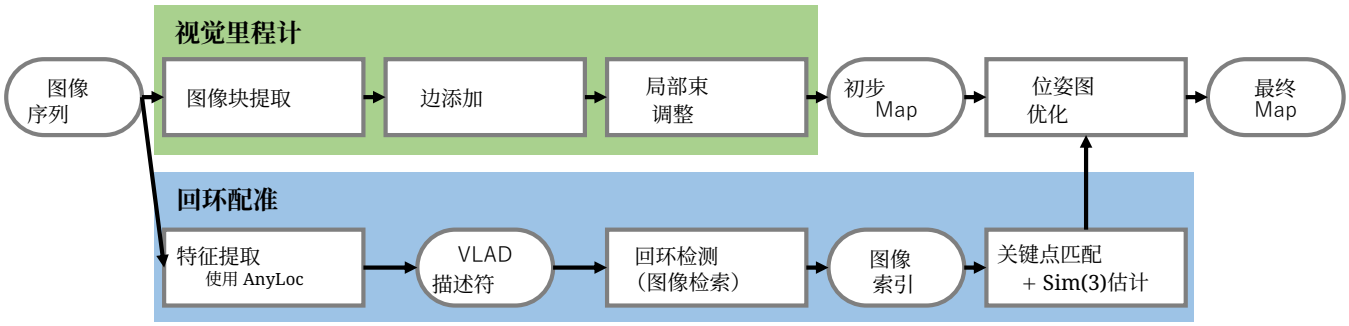

图 1. 采用基于 AnyLoc 的视觉地点识别的 DPV-SLAM 流程。

基于词频，具备较高的计算效率和处理速度。这类方法已应用于 ORB-SLAM [8] 和 DPV-SLAM 等系统。然而，基于 BoVW 的方法往往会因光照变化和视角变化而导致性能下降。

为应对这些挑战，基于深度学习的回环检测方法受到了关注。NetVLAD [9] 和 Patch-NetVLAD [10] 等技术将局部描述子聚合为全局图像描述子，从而实现稳健的视觉地点识别。其中，AnyLoc 在涵盖多个城市场景的大规模数据集上进行训练，具备较强的泛化能力，并能在结构化和非结构化环境中学习到具有语义意义的特征表示。

此外，也有利用空间结构和时间上下文的拓扑回环检测方法。例如，Topomap [11] 通过图表示捕捉环境的空间结构，有助于回环检测和地图一致性。同样，SemanticFusion [12] 所展示的语义信息整合方法，已证明能在复杂环境条件下有效提升识别的稳定性。

总而言之，现有的同步定位与建图系统在回环检测方面，已逐渐从经典的 BoVW 方法转向基于学习的方法。

受这一趋势启发，本文提出将深度视觉地点识别方法 AnyLoc 集成到原始 DPV-SLAM 框架中，以取代其 BoVW 模块，从而提高回环检测的鲁棒性和准确性。此外，本文还引入了自适应阈值机制，使系统能够根据环境变化动态调整图像匹配阈值。这一设计使系统能够在多样且具有挑战性的场景中保持稳定的性能。

III. 使用 AnyLoc 进行回环闭合

A. 概述

图 1 展示了所提方法的流程。在视觉里程计模块中，输入图像经过特征提取和位姿估计，随后通过光束法平差构建回环闭合前的地图先验。

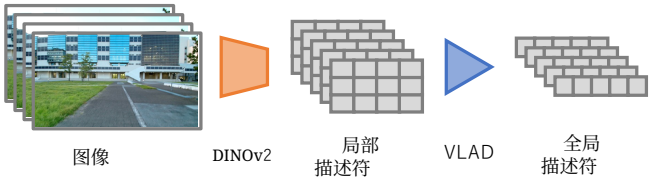

图2. 使用AnyLoc-VLAD-DINOv2提取全局描述子。

回环注册模块与这一流程并行运行，用于判断是否发生回环闭合。该模块使用 AnyLoc 提取全局描述子，并检索相似度分数较高的多个候选图像。随后，它根据阈值判断是否发生回环闭合，并进行几何验证。验证成功后，通过位姿图优化来优化地图，生成最终的一致地图。

B. 全局描述子提取

图2展示了全局描述子的提取过程。使用AnyLoc-VLAD-DINOv2从关键帧图像中提取全局描述子。在此过程中，DINOv2作为视觉特征提取的骨干网络，从输入图像中生成局部描述子。随后，使用VLAD特征聚合方法将这些局部描述子聚合为具有高度判别性的全局描述子。提取出的全局描述子被存储在数据库中，以供后续检索。

C. 自适应相似度阈值调整

在基于BoVW的AnyLoc等传统方法中，回环闭合判定所使用的相似度阈值必须手动设置，这会影响系统的泛化性能。为了解决这一问题，我们提出了一种相似度阈值自适应调整方法。

具体而言，在初始阶段，收集少量有效回环闭合候选的相似度分数（默认为5个），并计算其中位数，将其作为当前阈值（loop thresh）。该阈值反映了与环境相对应的分数分布。

然而，由于环境初始阶段不稳定，且后续环境变化也会产生影响，一旦检测到新的阈值，系统便会记录当时观察到的最大阈值

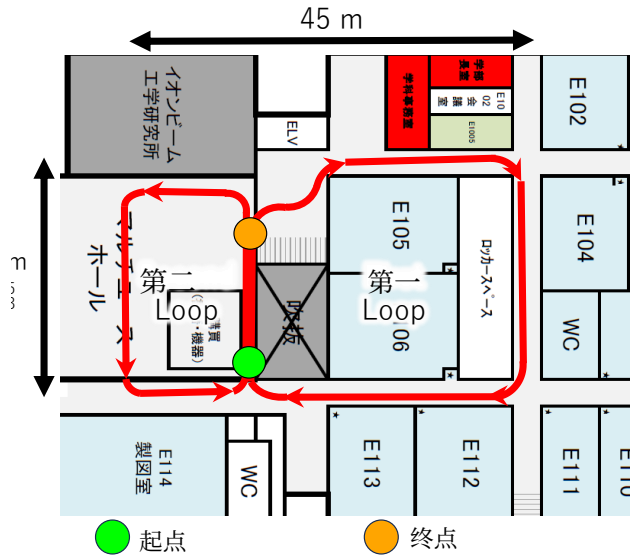

图3. 走廊环境。

并相应地更新先前的值。这种更新机制提升了回环检测过程的适应性和鲁棒性。

D. 回环检测与 *Sim(3)* 位姿配准

确定阈值后，计算每个关键帧的全局描述子与查询图像之间的相似度，并将超过回环阈值的关键帧视为回环候选。排除时间上相近的帧（例如，间隔在50帧以内的帧），然后选择相似度最高的候选帧。

回环检测后，使用 DISK [13] 提取关键点，并通过 LightGlue [14] 进行匹配。随后，结合 RANSAC [15] 与 Umeyama算法 [16]估计 Sim(3)变换。如果内点数量超过预定义阈值（30个点），则认为几何验证成功，并将估计的位姿作为回环约束加入位姿图。

E. 位姿图优化

在成功完成回环检测和几何验证后，估计得到的Sim(3)变换会作为回环闭合约束添加到位姿图中。我们采用 DPV-SLAM 中基于 CUDA 加速的块稀疏优化后端，执行非线性位姿图优化（PGO），从而高效、实时地完成全局位姿优化。优化过程在独立线程中异步运行，完成后由回调函数更新整个轨迹，实现持续的在线地图优化。

IV. 实验

A. 概述

我们在多个真实环境中进行了实验，并将我们的方法与以往的研究进行了比较。具体而言，本研究旨在回答以下问题：

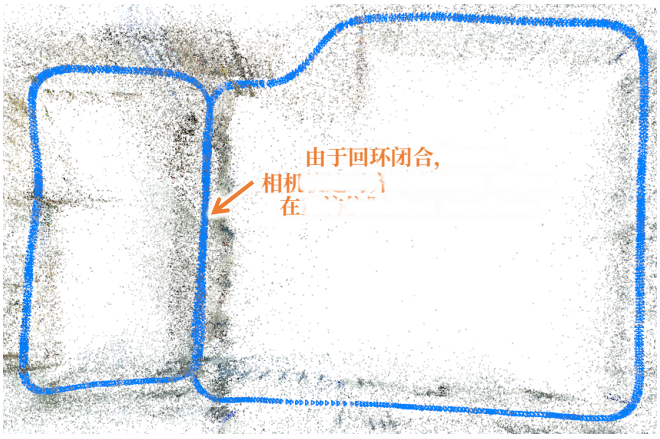

图 4. 走廊环境中的三维点云地图和相机轨迹（所提方法）。

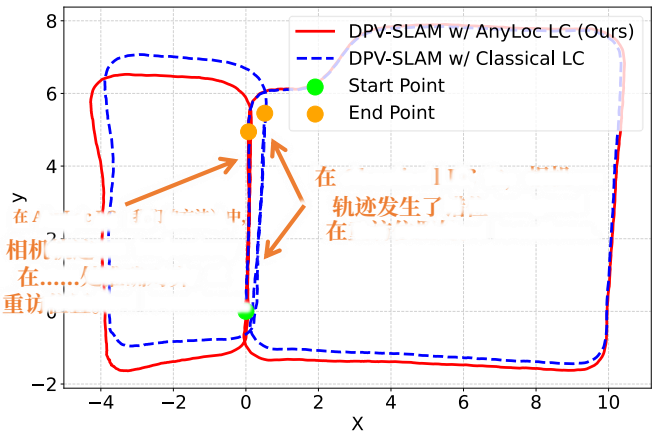

图 5. 走廊环境中的相机轨迹对比。

问题1： 与基线方法相比，所提方法能否提高回环检测的准确性，从而生成更准确的地图?

Q2: 与以往方法相比，所提方法在不同环境中是否展现出更强的泛化能力?

B. 环境

实验在法政大学小金井校区的室内和室外环境中进行。

图3展示了室内环境及其轨迹。图3中的室内环境较暗，并且多处出现人行横道。

图6展示了室外环境（庭院）及其轨迹。室外庭院环境的路径上有多个路段，其草地和混凝土等纹理相似。

此外，图9展示了室外环境（建筑物之间）及其轨迹。该环境中也有多处人行横道。

C. 数据采集与处理

使用 Logitech MX BRIO 700 (C1100) 摄像头以 30 Hz 的帧率采集图像。该摄像头

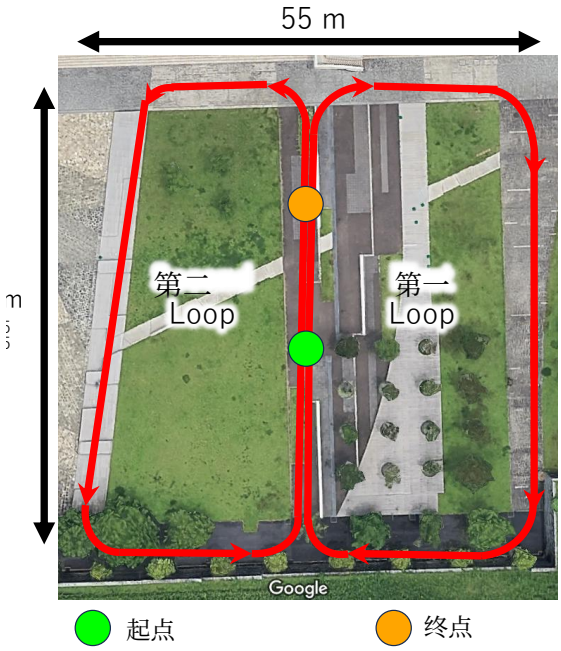

图6。室外环境（庭院）。

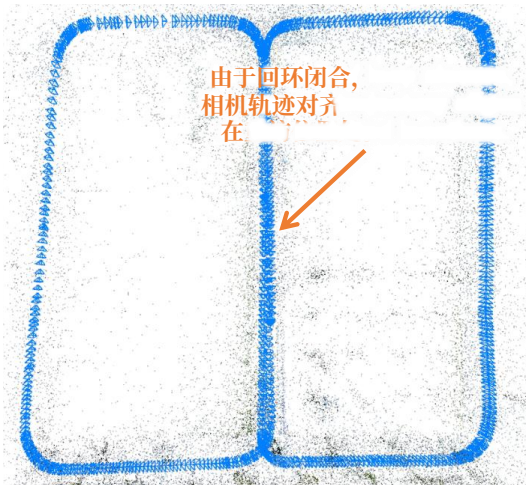

图7。室外庭院中的三维点云地图和相机轨迹（所提方法）。

在行走过程中由手持使用，并使用 rosbag 记录数据。

数据采集后，实验均离线进行。用于处理的计算机是一台运行 Ubuntu 22.04 的台式机，配备 Intel Core i7-13700 CPU、48 GB 内存，以及一块配有 16 GB 显存的 NVIDIA GeForce RTX 4070 Ti SUPER GPU。

D. 评估方法

为评估所提方法的有效性，我们进行了定性评估。评估的主要目标是从回环检测精度、地图重建质量和整体系统性能等方面，将我们的方法与 DPV-SLAM 进行比较。

在评估过程中，我们采用了离线实验

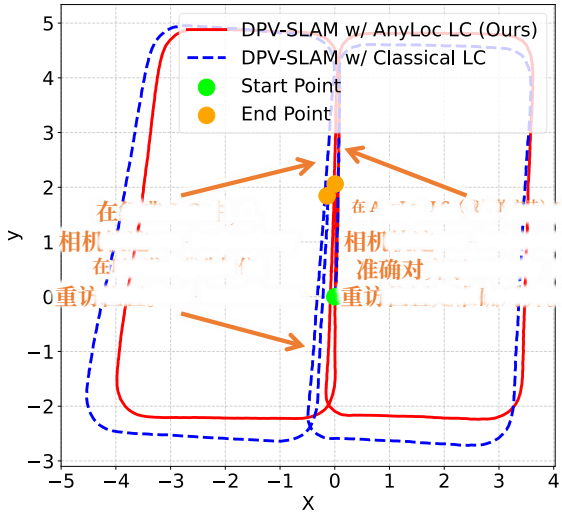

图 8. 室外环境（庭院）中的相机轨迹对比。

方法。数据使用统一的传感器配置采集，并分别通过所提方法和原始 DPV-SLAM 系统进行处理。通过比较两者在同一数据集上的回环检测性能和地图质量，我们旨在进一步验证所提方法在实际场景中的优势。根据与真实环境地图的对齐误差对建图结果进行评估，并通过视觉对比考察重建地图的完整性和一致性。

此外，我们对三种实验环境中由不同回环检测方法产生的图像检索结果进行了比较分析，从而揭示了各方法在回环检测精度方面的差异。

E. 实验结果

图 4 展示了所提方法在室内环境中的三维点云地图和相机轨迹。图 5 展示了同一环境中相机轨迹的对比。

接下来，图7展示了所提方法在室外环境（庭院）中的三维点云地图和相机轨迹，图8则对比了相机轨迹。

先前的方法未能检测到有效回环，导致沿回环的重访位置处相机轨迹明显错位。相比之下，所提方法成功检测到回环，同时保持了较高的位置精度，使回环闭合后重访位置处的相机轨迹彼此重合。这些结果表明，我们的方法有效提高了回环检测的准确性，从而实现了更精确的地图构建。这回答了**问题1**。

然后，图 10 展示了室外环境（建筑物之间）的三维点云地图和相机轨迹，而图 11 则对相机轨迹进行了比较。

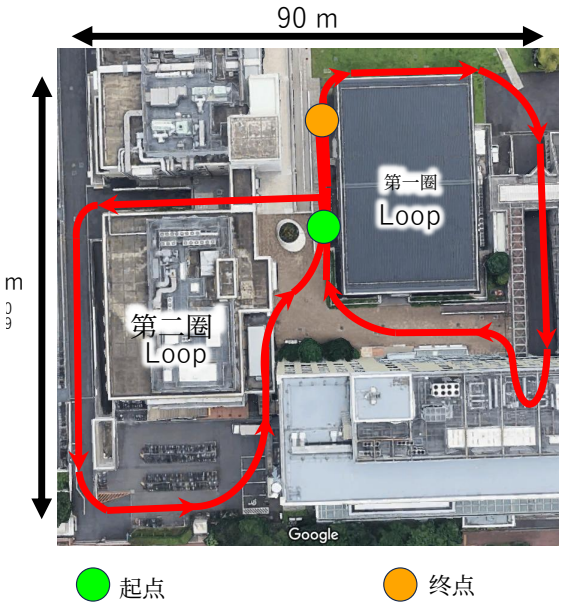

图 9. 室外环境（建筑物之间）。

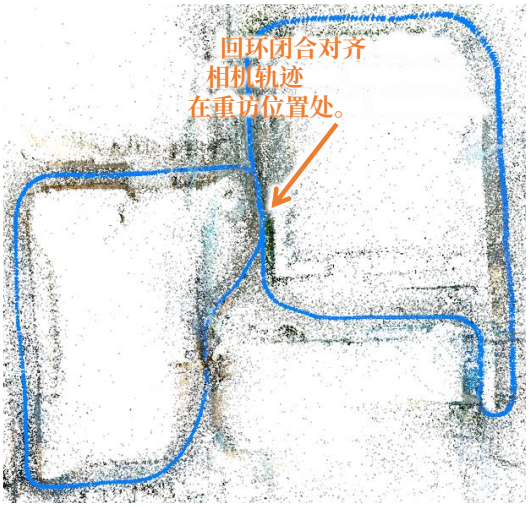

图 10. 建筑物之间的室外区域中的三维点云地图和相机轨迹（所提方法）。

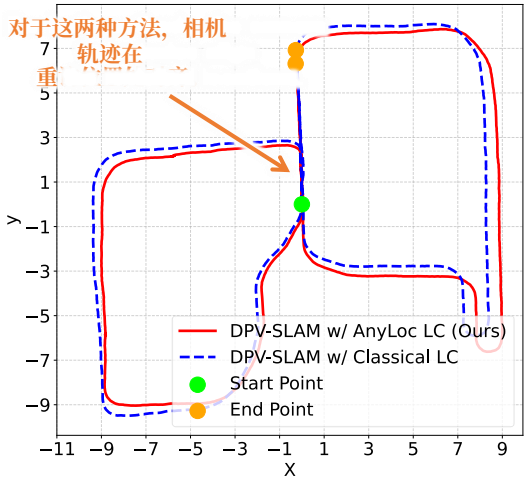

图 11. 室外环境（建筑物之间）中相机轨迹的比较。



图 12. 回环闭合检索性能（AnyLoc LC）。两幅图像之间检测到大量对应关系，其中一部分以高亮显示。这些匹配关键点随后用于配准和位姿图优化。



图 13. 回环闭合检索性能（Classical LC）。图中展示了查询图像及其匹配结果，并高亮显示了相似的特征区域。虽然通过调整阈值获得了最相似的图像，但匹配关键点数量不足，无法通过几何验证。

在此环境中，传统方法与所提方法之间未观察到显著差异。两种方法均正确检测到了回环，回环闭合使相机轨迹在重新访问的位置处发生重叠。

此外，如图 12 所示，所提方法能够持续检索到高度相似且成功通过几何验证的图像，从而实现可靠的回环闭合。相比之下，如图 13 所示，尽管经典方法检测到了回环闭合，但检索到的图像未能通过最终的几何验证，因此无法实现回环闭合。这些结果凸显了我们的方法在具有挑战性和复杂的环境中的鲁棒性和有效性。

如图14(a)–(c)所示，我们的方法始终在三个环境中识别出的闭环数量均多于传统方法。尽管调整经典方法的检测阈值可以适度增加其检测数量，但其性能仍不及我们的方法。这些观察结果表明，所提方法显著提升了回环检测的鲁棒性和整体能力，从而提高了生成地图的质量和准确性。

通过在三种不同环境中进行实验，所提方法在回环检测精度和轨迹精度方面均优于传统 DPV-SLAM。即使在低照度、重复纹理和动态干扰等具有挑战性的条件下，我们的方法仍表现出更高的鲁棒性和稳定性，显示出在实际应用中的巨大潜力。这回答了**Q2**。

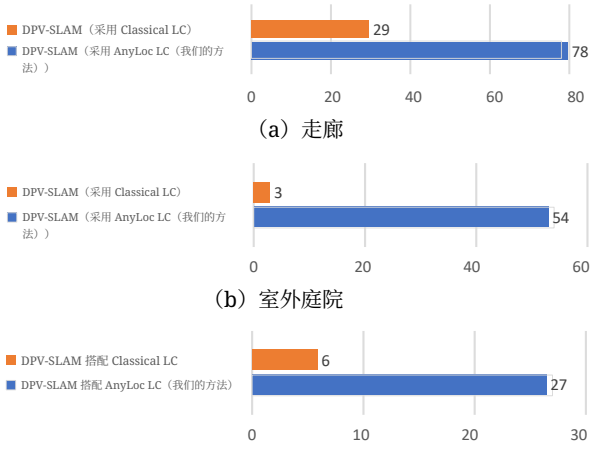

图14. 不同环境中检测到的回环数量对比。

V. 结论

本文提出将一种视觉地点识别方法 AnyLoc 集成到 DPV-SLAM 中，以替代传统的基于视觉词袋（BoVW）的回环检测。此外，我们设计了一套回环闭合流程，将自适应相似度阈值调整与几何验证相结合，并通过真实环境实验评估其有效性。真实环境中的实验结果表明，将 AnyLoc 与 DPV-SLAM 相结合可显著提高回环检测精度。

所提方法的一项局限是对图形处理器计算资源的需求较高。尽管系统可以在中央处理器上运行，但处理速度会显著降低。

未来的工作中，我们计划利用位置信息排除不必要的关键帧全局描述符，以提高回环检测速度。此外，我们还计划在标准基准数据集上评估所提方法。

参考文献

[1] N. Keetha, A. Mishra, J. Karhade, K. M. Jatavallabhula, S. Scherer, M. Krishna, and S. Garg, “AnyLoc：迈向通用视觉地点识别，” *IEEE* 机器人与自动化快报，第 9 卷，第 2 期，第 1286–1293 页，2023 年。[2] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, “ORB：SIFT 或 SURF 的高效替代方案，”发表于*IEEE* 国际计算机视觉会议（*ICCV*）论文集，2011 年。[3] D. G. Lowe, “基于尺度不变关键点的显著图像特征，” 国际计算机视觉期刊，第 60 卷，第 2 期，第 91–110 页，2004 年。[4] K. Adachi, Y. Hara, and S. Nakamura, “单目视觉 SLAM 的仿真评估：ORB-SLAM3、DROID-SLAM、DPVO 和 DPV-SLAM，”发表于*IEEE* 智能系统多传感器融合与集成国际会议（*MFI*）论文集，2025 年。[5] L. Lipson, Z. Teed, and J. Deng, “Deep Patch 视觉 SLAM，”发表于欧洲计算机视觉会议（*ECCV*）论文集，2024 年。[6] Z. Teed, L. Lipson, and J. Deng, “Deep Patch 视觉里程计，”发表于神经信息处理系统年会（*NeurIPS*）论文集，2023 年。[7] D. G´alvez-L´opez and J. D. Tard´os, “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。

[8] R. Mur-Artal, J. M. M. Montiel, and J. D. Tard´os, “ORB-SLAM：一种通用且精准的单目 SLAM 系统，” *IEEE* 机器人学汇刊，第 31 卷，第 5 期，第 1147–1163 页，2015 年。[9] R. Arandjelovi´c, P. Gronat, A. Torii, T. Pajdla, and J. Sivic, “NetVLAD：用于弱监督地点识别的卷积神经网络架构，”发表于*IEEE/CVF* 计算机视觉与模式识别会议（*CVPR*）论文集，2016 年。[10] S. Hausler, S. Garg, M. Xu, M. Milford, and T. Fischer, “Patch-NetVLAD：用于地点识别的局部-全局描述子多尺度融合，”发表于*IEEE/CVF* 计算机视觉与模式识别会议（*CVPR*）论文集，2021 年。[11] F. Bl¨ochliger, M. Fehr, M. Dymczyk, T. Schneider, and R. Siegwart, “Topomap：基于视觉 SLAM 地图的拓扑建图与导航，”发表于*IEEE* 国际机器人与自动化会议（*ICRA*）论文集，2018 年。[12] J. McCormac, A. Handa, A. Davison, and S. Leutenegger, “SemanticFusion：利用卷积神经网络进行稠密三维语义建图，”发表于*IEEE* 国际机器人与自动化会议（*ICRA*）论文集，2017 年。[13] M. Tyszkiewicz, P. Fua, and E. Trulls, “Disk：利用策略梯度学习局部特征，”发表于神经信息处理系统年会（*NeurIPS*）论文集，2020 年。[14] P. Lindenberger, P.-E. Sarlin, and M. Pollefeys, “Lightglue：极速局部特征匹配，”发表于*IEEE/CVF* 计算机视觉与模式识别会议（*CVPR*）论文集，2023 年。[15] M. A. Fischler and R. C. Bolles, “随机样本一致性：一种模型拟合方法及其在图像分析和自动制图中的应用，” *ACM* 通讯，第 24 卷，395，1981 6, 381no. pp. – 。[16] S. Umeyama, “两个点模式之间变换参数的最小二乘估计，” *IEEE* 模式分析与机器智能汇刊，第 13 卷，第 4 期，第 376–380 页，1991 年。