

分类号：TP242.6

密 级：公开

学校代码：10363

学 号：2220320174



安徽工程大学

Anhui Polytechnic University

硕士学位论文

题 目 动态遮挡场景下基于改进YOLOv6s
网络的SLAM算法

论 文 作 者	<u>许瑞珩</u>
校 内 导 师	<u>陈孟元 教授</u>
校 外 导 师	<u>陈青</u>
专业学位类别	<u>电子信息</u>
研究方向（领域）	<u>模式识别与智能系统</u>

论文提交日期：2025 年 5 月 31 日

安徽工程大学学位论文原创性声明

本人郑重声明：我恪守学术道德，崇尚严谨学风。所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已明确注明和引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品及成果的内容。论文为本人亲自撰写，我对所写的内容负责，并完全意识到本声明的法律结果由本人承担。

学位论文作者签名：许瑞珩

日期：2025 年 5 月 12 日

安徽工程大学学位论文版权使用授权书

学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅或借阅。本人授权安徽工程大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

保密 ☐，在 ____ 年解密后适用本版权书。

本学位论文属于

不保密 ☒。

学位论文作者签名：许瑞珂

指导教师签名：陈吉

日期：2025 年 5 月 12 日

日期：2025 年 5 月 12 日

动态遮挡场景下基于改进 YOLOv6s 网络的 SLAM 算法

摘 要

同步定位与地图构建（Simultaneous Localization And Mapping, SLAM）作为移动机器人实现自主导航的核心技术，在智能仓储、自动驾驶、增强现实等众多领域发挥着关键作用。视觉 SLAM 因相机成本低廉、能获取丰富的环境纹理和语义信息，成为当下研究的重点方向。传统算法在动态场景下，难以精确区分动态目标和被遮挡物体的边缘，在判断动态特征点时容易误判，剔除动态物体后背景修复效果不佳，导致定位精度下降、构图不准确以及跟踪不稳定等问题，限制了视觉 SLAM 技术在复杂动态环境中的实际应用。

针对动态遮挡环境下物体之间存在相互遮挡关系时易出现目标边缘模糊分割效果不佳的问题，采用基于改进 YOLOv6s 网络的 YDO-SLAM 算法(improved SLAM algorithm of YOLOv6s network based on dynamic occlusion scenario, YDO-SLAM)。在物体识别与分割阶段，对 YOLOv6s 网络进行优化。在 Backbone 主干网络的 SimSPPF 模块中结合两种池化方法的优势，并在双分支 FPNB 特征融合模块上设计了双分支架构，使 SLAM 可以高效整合跨层语义信息与同层多尺度特征，增强网络对边缘信息的捕获能力，从而提升动态场景下物体分割的精度，为后续背景修复奠定基础。

针对移除动态物体时导致大量静态特征误剔除的问题，YDO-

SLAM 算法采用双重约束特征策略。基于改进的 Z-score 评分，通过计算有效特征点深度值的均值、标准差以及结合四分位数差自适应调整阈值，精准识别异常深度值的特征点，提高动态特征点剔除的准确性，确保在复杂环境下能稳定保留静态特征点，维持系统定位精度与跟踪稳定性。当图像中动态物体过多时，剔除动态物体将导致背景信息缺失，从而造成特征点过少。通过基于双方向映射的背景修复模型挖掘前后帧中的静态信息，恢复图像的完整性，从而补充高质量的特征点，为 SLAM 系统的位姿估计和地图构建提供更完整、准确的信息，增强系统在动态遮挡场景下的性能。

为验证 YDO-SLAM 算法的有效性，本文在公开数据集 TUM 上进行多次实验，并在真实场景中进行实地测试。实验结果表明，在复杂动态环境下，该算法定位精度和构图能力显著提升，相较于 DS-SLAM、DM-SLAM 和 Dyna-SLAM 算法，绝对轨迹误差分别降低 86.73%、20.75%及 17.21%，与传统算法相比，能够更精准地估计相机位姿，大幅减少轨迹误差，构建的地图更精确清晰，有效排除动态物体干扰，真实反映环境结构。同时，算法展现出良好的稳定性和鲁棒性，最后搭建硬件实验平台验证所提算法在真实场景中的可行性。

关键词:同步定位与地图构建，动态场景，语义分割，运动判断，背景修复

IMPROVED SLAM ALGORITHM OF YOLOV6S NETWORK BASED ON DYNAMIC OCCLUSION SCENARIO

ABSTRACT

Simultaneous Localization and Mapping (SLAM) serves as a core technology for autonomous navigation in mobile robots, playing a crucial role in various fields such as intelligent warehousing, autonomous driving, and augmented reality. Visual SLAM has become a major research focus due to the low cost of cameras and their ability to capture rich environmental texture and semantic information. However, traditional algorithms struggle in dynamic environments, failing to accurately distinguish between dynamic objects and occluded static structures. Misjudgment of dynamic feature points and ineffective background restoration after removing dynamic objects lead to decreased localization accuracy, inaccurate mapping, and unstable tracking, limiting the practical application of visual SLAM in complex dynamic environments.

To address the challenge of blurred object boundaries and poor segmentation performance caused by mutual occlusions in dynamic environments, this paper proposes the YDO-SLAM algorithm based on an improved YOLOv6s network. In the object recognition and segmentation

phase, YOLOv6s is optimized by enhancing the Backbone network's SimSPPF module with a combination of two pooling strategies and introducing a novel dual-branch architecture in the FPNB feature fusion module. These improvements enable SLAM to efficiently integrate cross-layer semantic information and multi-scale features within the same layer, enhancing edge information capture and improving object segmentation accuracy in dynamic scenarios, thereby laying a solid foundation for subsequent background restoration.

To mitigate the issue of excessive removal of static features when eliminating dynamic objects, the YDO-SLAM algorithm incorporates a dual-constraint feature strategy. By leveraging an improved Z-score evaluation, it precisely identifies outlier depth values of feature points by calculating the mean, standard deviation, and interquartile range-based adaptive threshold adjustment. This approach significantly improves the accuracy of dynamic feature point removal, ensuring the stable retention of static feature points in complex environments while maintaining system localization accuracy and tracking stability. When an image contains an excessive number of dynamic objects, their removal leads to background information loss and a subsequent reduction in feature points. To address this, a bi-directional mapping-based background restoration model is introduced, which exploits static elements from consecutive frames to reconstruct occluded regions. This strategy restores image completeness,

supplements high-quality feature points, and provides SLAM with more comprehensive and accurate information for pose estimation and map construction, thereby enhancing system performance in dynamic occlusion scenarios.

To validate the effectiveness of the YDO-SLAM algorithm, extensive experiments were conducted on the public TUM dataset and in real-world environments. Experimental results demonstrate that in complex dynamic environments, the proposed algorithm significantly improves localization accuracy and mapping quality. Compared to DS-SLAM, DM-SLAM, and Dyna-SLAM, the absolute trajectory error is reduced by 86.73%, 20.75%, and 17.21%, respectively. Compared with traditional algorithms, YDO-SLAM achieves more precise camera pose estimation, significantly reduces trajectory errors, and constructs clearer and more accurate maps while effectively eliminating dynamic object interference to reflect the true environmental structure. Furthermore, the algorithm exhibits excellent stability and robustness. Finally, a hardware experiment platform was built to verify the feasibility of the proposed algorithm in real-world scenarios.

Xu Ruiheng(Electronic Information)

Supervised by Professor Chen Mengyuan

KEY WORDS: simultaneous localization and mapping, dynamic scene, semantic segmentation, motion judgment, background restoration

目 录

第 1 章 绪论.....	1
1.1 本文研究背景及意义.....	1
1.2 国内外研究现状.....	3
1.2.1 视觉 SLAM 研究现状	3
1.2.2 动态遮挡场景下视觉 SLAM 研究现状	6
1.3 研究内容与章节安排.....	8
第 2 章 视觉 SLAM 与 YOLOv6s 网络基础.....	11
2.1 视觉 SLAM 系统框架	11
2.1.1 视觉里程计.....	11
2.1.2 后端优化.....	12
2.1.3 闭环检测.....	13
2.1.4 系统建图.....	14
2.2 YOLOv6s 网络动态物体分割	15
2.2.1 整体架构概述.....	15
2.2.2 头部网络预测机制.....	16
2.2.3 物体的动态特征剔除.....	17
2.3 本章小结.....	18
第 3 章 基于双重特征约束的动态判断方法.....	19
3.1 局部自适应的去模糊方法.....	19
3.1.1 传统去模糊方法.....	19
3.1.2 基于注意力引导的局部自适应去模糊方法.....	21
3.2 基于双重约束筛选动态特征.....	23
3.2.1 结合改进 Z-score 评分判断动态特征	24
3.2.2 基于点线特征的相机位姿优化.....	26
3.3 实验结果与分析.....	27
3.3.1 图像去模糊实验.....	28
3.3.2 动态特征剔除实验.....	29
3.3.3 轨迹对比与建图实验.....	30

3.3.4 轨迹精度对比试验.....	33
3.4 本章小结.....	34
第 4 章 基于网络实例分割与双向映射的背景修复方法.....	35
4.1 基于改进 YOLOv6s 的物体分割.....	36
4.1.1 改进 simspfp 模块.....	36
4.1.2 双分支 FPNB 融合特征.....	37
4.2 基于改进双向映射的背景修复.....	39
4.2.1 背景修复原理.....	40
4.2.2 约束条件与优化策略.....	41
4.3 实验结果与分析.....	41
4.3.1 动态遮挡物体分割算法实验.....	41
4.3.2 系统定位精度实验.....	43
4.3.3 背景修复实验.....	44
4.4 硬件平台设计及真实场景实验.....	46
4.4.1 机器人硬件平台.....	46
4.4.2 真实场景实验.....	46
4.5 本章小结.....	48
第 5 章 全文总结展望.....	49
5.1 全文总结.....	49
5.2 展望.....	50
参考文献.....	51
攻读学位期间发表的学术论文的科研成果.....	55
致谢.....	56

第 1 章 绪论

1.1 本文研究背景及意义

在当今科技快速发展的背景下，机器人技术的应用范围不断扩大，其中同步定位与地图构建（Simultaneous Localization And Mapping, SLAM）为机器人智能自主运行的核心技术，SLAM 能在没有预设环境信息的情况下，通过内置传感器来收集环境信息，实现定位与跟踪，这一过程为机器人的导航、路径规划以及执行定位跟踪任务提供了有力支持。视觉 SLAM 基于传感器的不同，主要被分为激光 SLAM 及视觉 SLAM 两大类。激光 SLAM 借助激光雷达的高精度测距优势，技术相对成熟，在工业场景中的应用较为广泛，但激光雷达存在成本高昂、设备体积较大以及获取环境信息不够丰富等缺点，限制其在大规模应用及复杂场景下适应的可能。相比之下，视觉 SLAM 以相机作为传感器，则有着信息获取全面、成本低廉、设备轻便等优势，逐步成为 SLAM 领域的研究热点^{[1][2][3]}。

在实际应用中 SLAM 技术遭遇诸多挑战，特别是在复杂动态场景中，定位与地图构建的效果欠佳。根据公开数据集 TUM 的统计，动态场景中遮挡率普遍超过 30%，动态物体密度可达 2-5 个/平方米，且动态目标与背景的深度值差异范围通常在 0.5-3 米之间。传统视觉 SLAM 算法通常假设环境是静态的，因此在应对此类场景时，难以准确区分动态物体与静态背景，导致定位误差显著增加（平均绝对轨迹误差可达 0.5 米以上）。当动态物体进入相机视野，会导致特征点匹配错误，使算法对相机运动状态产生误判，进而影响定位精度和建图的可靠性。在机器人导航过程中，若无法精准识别和处理动态物体，可能导致机器人碰撞障碍物或迷失方向，无法顺利完成定位跟踪任务^{[4][5]}。物联网迅速演进与 5G 技术广泛应用并行发展以及智能设备间的信息交互需求显著增强，使移动机器人如何在复杂环境中实现高效、精确的感知与定位成为当前亟需攻克的关键技术难题。例如，在智能工厂中，大量的移动机器人需要在动态变化的生产线上高效协作，精准地完成物料搬运、设备巡检等任务；在智慧城市的建设中，服务型机器人要在人流如织的街道和建筑物内为人们提供导航、配送等服务，必须能够快速且准

确地应对动态场景^{[6][7]}。在这些场景中，传统视觉 SLAM 算法的局限性被进一步放大，迫切需要新的技术和算法来提升其在动态环境下的性能。移动机器人的应用如图 1-1 所示。



图 1-1 移动机器人的技术应用

众多学者正不断开发 SLAM 技术与深度学习相结合的策略应对现有挑战。深度学习在目标检测、语义分割等领域取得显著成果，能够从图像中提取丰富的语义信息，为视觉 SLAM 处理动态场景提供了新思路。例如，DynaSLAM 利用 MaskR-CNN 实现实例分割，旨在识别动态物体，并集成多视图几何技术进行运动评估与目标区域排除，以此增强在动态环境中的 SLAM 效能；而 DS-SLAM 则采取融合 SegNet 语义分割模型与运动检测机制的方式，对可能包含动态物体的区域进行标记区分，显著降低了动态元素对系统性能的干扰。

现有的基于深度学习的视觉 SLAM 算法仍存在缺陷^{[8][9][10]}：（1）在复杂动态遮挡场景下，对被遮挡物体和动态目标的边缘识别精度不足，导致分割模糊，影响系统的后续定位和建图；（2）在判断动态特征点时，容易误剔除静态特征点，使静态特征点数量匮乏，造成跟踪丢失；（3）在剔除动态物体后，未能充分利用前后帧信息修复背景，导致信息缺失，影响系统的位姿估计和建图效果。

并且深度学习算法通常对硬件要求较高,计算复杂度大,限制了其在资源受限的设备上应用。

综上所述,在动态遮挡场景下传统 SLAM 算法的运行所面临的挑战在于精确识别被遮挡物与动态对象的界限。并且在动态特征点筛选过程中,部分静态特征点可能被误剔除,导致有效静态特征点数量不足,从而影响系统的跟踪稳定性,甚至引发跟踪丢失。本文构建基于改进 YOLOv6s 网络的 YDO-SLAM 算法,采用多尺度特征融合、动态特征剔除和背景修复等方法,着重解决动态遮挡场景下的特征运动状态判断,精确地识别被遮挡物体与动态目标的边缘,从而精确地分割动态目标并提升统定位跟踪的鲁棒性。在跟踪的过程中同时提取点特征和线特征,根据获取的语义信息并利用改进 Z-score 评分和对极几何算法剔除其中的动态特征,随后通过双向映射背景修复模型对被剔除动态物体的区域补全静态信息,并融入高质量特征点,旨在提升 SLAM 系统在动态遮挡环境下的稳定性和精确度。SLAM 系统运行效果如图 1-2 所示。

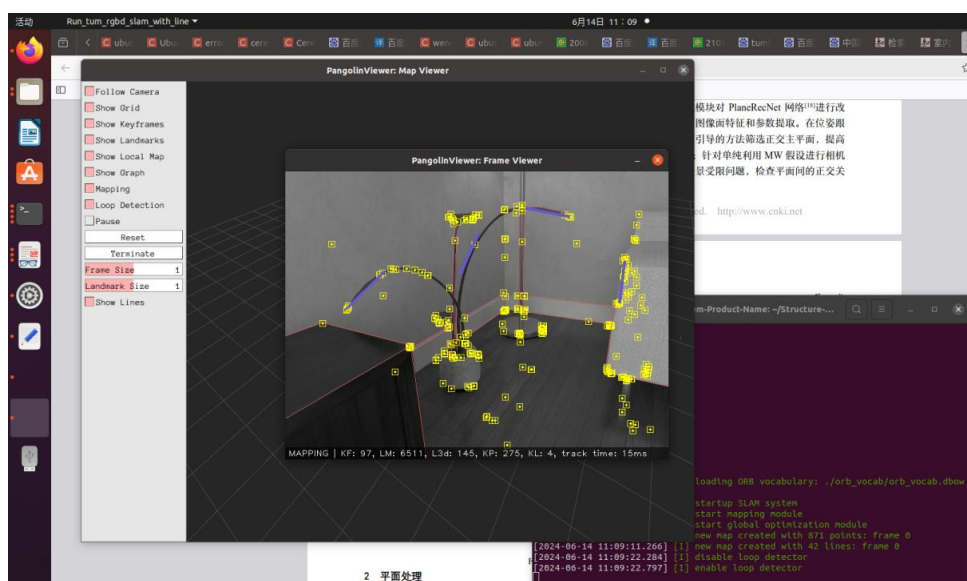


图 1-2 SLAM 系统运行示意图

1.2 国内外研究现状

1.2.1 视觉 SLAM 研究现状

视觉 SLAM 作为机器人领域的重要技术,近年来取得了显著进展,并广泛应用于各种实际场景。视觉 SLAM 根据图像处理策略主要被分为直接法与特征

点法两大类别^{[11][12][13]}。

基于直接法的视觉 SLAM 系统利用像素灰度信息实现目标追踪，通过量化相邻图像帧之间的光度差异，从而估算相机的位置和姿态变化，使其在低纹理环境下依然能够有效运行，并支持稠密或半稠密地图的构建。例如，Newcombe 等人于 2011 年提出的 DTAM 系统^[14]，首次将直接法应用于视觉里程计，能够实现相机的实时跟踪与建图，该方法通过深度估计、匹配及关键帧选择构建场景的稠密地图，一定程度上缓解了低纹理场景下的跟踪问题，但其在实时建图时依赖 GPU 加速，计算成本较高。Engel 等人于 2014 年提出 LSD-SLAM 算法^[15]，基于光照不变性假设，利用所有像素的强度信息估算帧间位姿，进而构建全局稳定的半稠密环境地图，在低纹理场景下依然具备较高的定位精度与鲁棒性。然而，该方法对光照变化较为敏感，在光照条件变化剧烈或不均匀的场景中容易导致跟踪失败，影响系统稳定性。Engel 等人于 2016 年进一步推出 DSO 算法^[16]，优化了位姿估计误差模型，在精度和鲁棒性方面表现优越，但缺少回环检测模块。

特征点法是当前视觉 SLAM 的主流方法，其基本原理在于从图像帧中提取特征点并进行匹配进而优化相机位姿估计。相比直接法，特征点法具有尺度和旋转不变性，并在模糊或含噪声的图像中展现出更强的鲁棒性，因此在复杂环境下具有更稳定的性能。Davison 等人于 2007 年引入了 MonoSLAM 系统^[17]，首次利用单目相机技术实现了移动机器人的实时定位与地图构建，该方法通过稀疏特征点提取结合卡尔曼滤波进行位姿优化，但容易受到传感器噪声影响，导致特征点跟踪丢失。同年，Klein 和 Murray 提出 PTAM 算法^[18]，采用非线性优化替代后端卡尔曼滤波，从而降低传感器噪声的干扰，且该方法引入关键帧筛选机制提升位姿优化和建图效率，有效减少数据冗余，但在场景变化明显的环境下依然出现跟踪失败的问题。

随着研究的深入，多传感器融合的视觉 SLAM 技术逐渐成为研究的重点。例如，VI-SLAM^[19]结合惯性测量单元 IMU 与视觉传感器，采用松耦合或紧耦合的方式融合数据以发挥两者优势，提升系统的抗干扰能力和定位精度。香港科技大学团队开源的 VINS-Mono 通过 IMU 预积分与特征观测数据融合优化，增强了定位精度和初始化鲁棒性。然而，该系统缺少重定位模块，在长时间运行或丢失跟踪时，恢复能力相对受限^[20]。Mur-Artal 等人于 2015 年推出 ORB-SLAM 系列

算法^{[21][22]}，该算法作为完整的视觉 SLAM 系统，集成了跟踪、建图和闭环检测等模块。ORB-SLAM 利用 ORB 特征提取与匹配提高跟踪准确率和定位精度，并引入视觉词袋模型（BoW）检测图像帧间相似性来判断是否发生回环，从而增强系统的全局一致性和长期稳定性。ORB-SLAM2 优化了特征点提取与匹配的精度和效率，成为目前应用最广泛的视觉 SLAM 算法，其系统架构如图 1-3 所示。

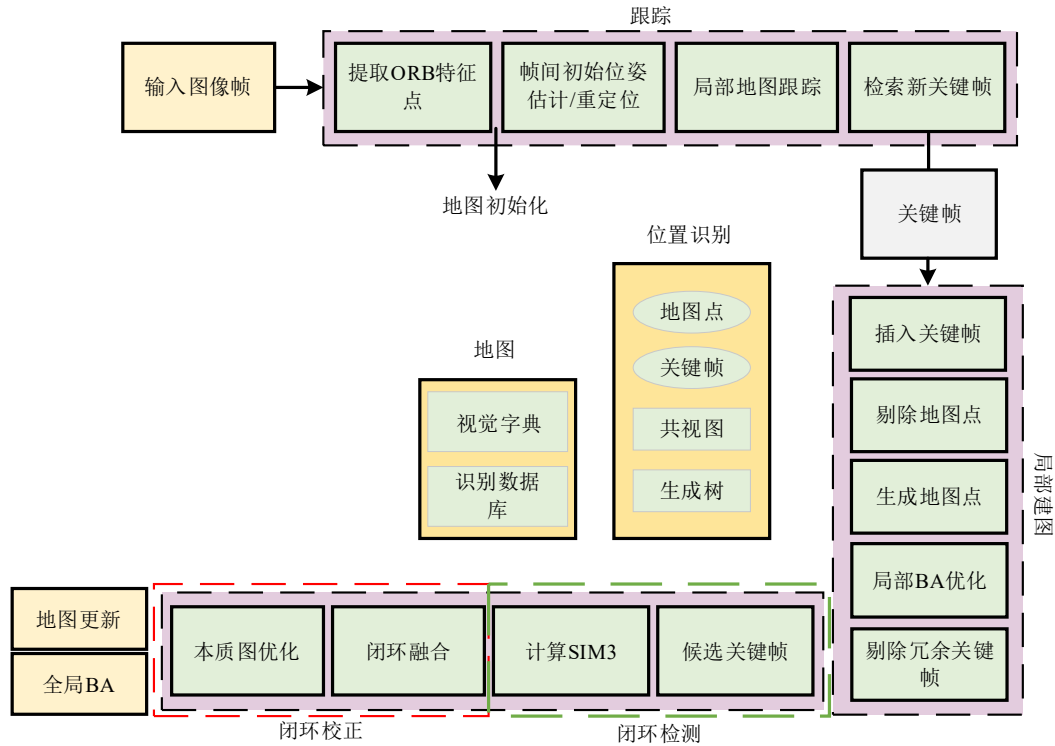


图 1-3 ORB-SLAM 系统框架

传统的视觉 SLAM 算法往往在假设环境为静态的基础上运行，但在实际应用时的动态场景下难以区分动态目标与静止背景，从而引发定位精度下降以及地图构建失真等关键问题。此问题在行人密集、车辆流动性较强的城市环境等动态物体较多的场景中尤为突出，使得传统 SLAM 系统难以保持稳定的轨迹估计和高精度的地图构建。且在光照变化剧烈或存在动态阴影的情况下，系统往往难以有效区分真实的动态物体和光照引起的伪动态信息，从而进一步影响系统的鲁棒性。为克服这些挑战，研究团队在视觉 SLAM 领域融合了深度学习方法。例如，DS-SLAM 算法^[23]采用 SegNet 语义分割网络对潜在动态物体区域进行分割标记，并结合运动约束条件来判断其动态特征，以此构建静态语义八叉树地图。SegNet 基于全卷积网络架构能够获取图像的高层语义信息，从而提高对动态目标的检测

能力。然而，该方法在动态区域占比较大的场景下鲁棒性较弱，对动态区域的分割精度仍有待提升。尤其是在遮挡严重或动态物体与背景颜色相近的情况下，SegNet 无法准确分割出完整的动态目标，导致部分动态信息残留在 SLAM 系统中，影响最终的定位与建图效果。Bescos 等人于 2017 年提出 DynaSLAM 算法^[24]利用 Mask R-CNN 网络与多视图几何结合，检测并剔除场景中的动态物体，在动态环境下具备一定的鲁棒性，但 Mask R-CNN 实例分割网络对于被遮挡潜在动态物体的分割效果不佳，其系统框架如图 1-4 所示。

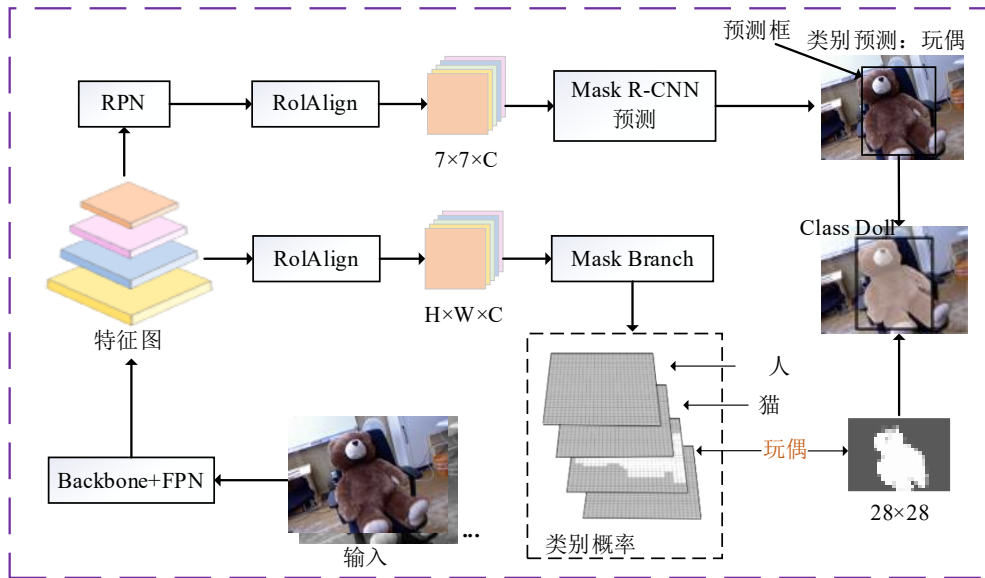


图 1-4 Mask R-CNN 系统框架

在复杂度更高的动态遮挡环境下视觉 SLAM 适用性依然较差^[25]，难以精确识别被遮挡物体与动态目标的边缘，导致分割模糊。在识别动态特征点时，可能误剔除静态特征点而引发跟踪丢失或其他定位误差。针对上述问题，构建基于改进 YOLOv6s 网络的 YDO-SLAM 算法，旨在提升视觉 SLAM 在动态遮挡场景下的分割性能和动态特征判断。

1.2.2 动态遮挡场景下视觉 SLAM 研究现状

基于几何检测的视觉 SLAM 算法通常假设静态特征点占主要比例，但在复杂动态环境下难以成立，特征点动静态分类不准确导致算法的鲁棒性较弱。为解决这一问题，国内外研究者积极探索深度学习与 SLAM 技术的融合，以提升系统在动态场景下运行的鲁棒性。Yu 等人于 2018 年提出 DS-SLAM 算法引入语义分割网络 SegNet^[26]和移动一致性检查系统在动态场景下的性能，并生成密集的

语义八叉树图^[27]；Toni Vidal-Garcia 等人将 MaskR-CNN 网络与多视图几何约束方法相结合，开发出 DynaSLAM 算法，该方法在动态 SLAM 领域具有重要影响。张驰等人于 2019 年采用了单阶段目标检测器（SSD）识别动态物体，并引入一种基于相邻帧中特征点速度一致性原理的静态特征补偿策略，进而构建 Dynamic-SLAM^[28]。王天驰等人于 2020 年整合了光流信息于 MaskR-CNN 网络中，并通过引入重投影偏差矢量来识别包含深度信息的特征点，进而提出了 DM-SLAM^[29]算法。Ji T 等人^[30]于 2021 年通过深度图聚类 and 重投影误差方法完成了非先验动态运动物体的检测，利用 SegNet 网络对关键帧中的运动目标进行检测。同年，Huang T 等人采用了两种一致性检查策略，用于检测并过滤动态特征^[31]。张磊等人^[32]于 2022 年采用了点与线特征以估算相机的位置与姿态，并利用 K 均值聚类技术融合了深度信息的以及深度学习驱动的语义分割方法，旨在识别潜在的动态特征。同时，由于深度信息在许多场合并不能很好的剔除动态点，Wang Y 等人提出了 DRG-SLAM^[33]利用 Segnet 神经网络分割动态对象，结合线特征和平面特征提高了系统的鲁棒性，该系统在低动态场景下展现出良好的性能。

综上所述，本文构建了一种基于动态遮挡场景改进 YOLOv6s 网络的 SLAM 算法。不同于传统的高低层特征融合，在融合后的图层上进行了一次 FPNB 处理，以获取更多的尺度特征。这一系列改进旨在提升最终的前景和背景图像分割效果，既保留了 YOLOv6s 本身的分割性能，又解决了目标分割中背景特征不足、分割效果不明显的问题。经过这些改进后，背景的特征语义信息更为丰富，不仅保留了动态特征点的排除效果，还在处理被遮挡部分的背景时表现出色。这一优化方案有效弥补了原始 YOLOv6s 在目标分割任务中的不足，为网络在复杂场景下的性能提升提供了有力支持，通过更精细的特征提取和融合，成功实现了在保持分割效果的同时，使背景信息更加显著，为目标分割任务的实际应用提供了更可靠的解决方案。在排除动态特征的基础上引入了双向映射模型，通过利用前后帧中的静态信息复原被移除部分的 RGB 图像背景。随后在修复的图像中进行特征点再提取，旨在保证全面的背景信息，进而增强 SLAM 在位姿估计与地图构建任务时的精确度与稳定性。

1.3 研究内容与章节安排

在动态遮挡场景下传统 SLAM 算法难以区分被遮挡物体和动态目标的边缘，致使分割模糊，严重影响系统区分场景中动态物体和静态背景的准确度^[34]。另一方面，在筛选动态特征点时静态特征点容易被错误地误剔除，进而造成特征点数量过少导致跟踪丢失，使 SLAM 无法实现稳定可靠的定位与地图构建。当在动态物体较多的场景下剔除动态物体后，图像的特征信息过少，进而影响回环检测以及全局地图的构建^{[35][36]}。

针对上述问题，本文构建了动态遮挡场景下基于改进 YOLOv6s 网络的 SLAM 算法。在网络架构方面，对 YOLOv6s 网络进行了针对性改进。首先通过在 simspfpf 模块上结合不同的池化方式使得网络在复杂动态场景下对边缘信息的识别能力显著增强，提升区分被遮挡物体与动态目标边缘的能力。并将 RepBlock 融合特征方式替换为 FPNB。实现了高低层语义特征信息的融合以及同层内部的多尺度融合，提升了动态场景中人物和背景的分割效果，增强了被遮挡部分的修复能力，为后续精确剔除动态物体和背景修复提供了可靠的基础^[37]。

在特征融合与提取上，利用卷积网络对输入图像执行上采样和下采样操作，从而生成具备不同尺度与语义通道的图像。这样做既减少了传统下采样过程中背景信息的损失，增加了网络感受野，又能将不同尺度的图像表征特征更好地融合，有效防止被遮挡物体和动态物体边缘模糊。

在动态特征的剔除方面，采用基于深度信息的 Z-score 评分方法和对极几何约束的双重约束特征点筛选策略。第一阶段，基于改进的 Z-score 评分来剔除动态异常点，并根据四分位数差的大小来调整 Z-score 的阈值。对于数据分布偏离正态分布的情况，引入基于分位数设定的自适应阈值，有效降低错检与漏检的风险。在第二阶段，采用了对极几何约束进行二轮筛选，计算特征点与极线之间的距离，当距离超越预设的阈值时，则断定该特征点为动态特征点，后对追踪线程中获取的特征线进行筛选，当特征线的端点坐标位于检测框内部，则认定该线为动态线并予以排除，由此显著减少了动态物体对 SLAM 系统的干扰，从而提升了系统的稳定性和准确性。

针对背景修复部分，采用基于双方向映射的背景修复模型，该方法利用前、

后两帧中的静态信息恢复动态目标遮挡的静态背景部分。在背景修复后通过对修复区域的特征点再提取，补充全局特征信息，旨在提升 SLAM 系统在动态遮挡环境下的定位精确度和鲁棒性，论文总体技术路线如图 1-5 所示。

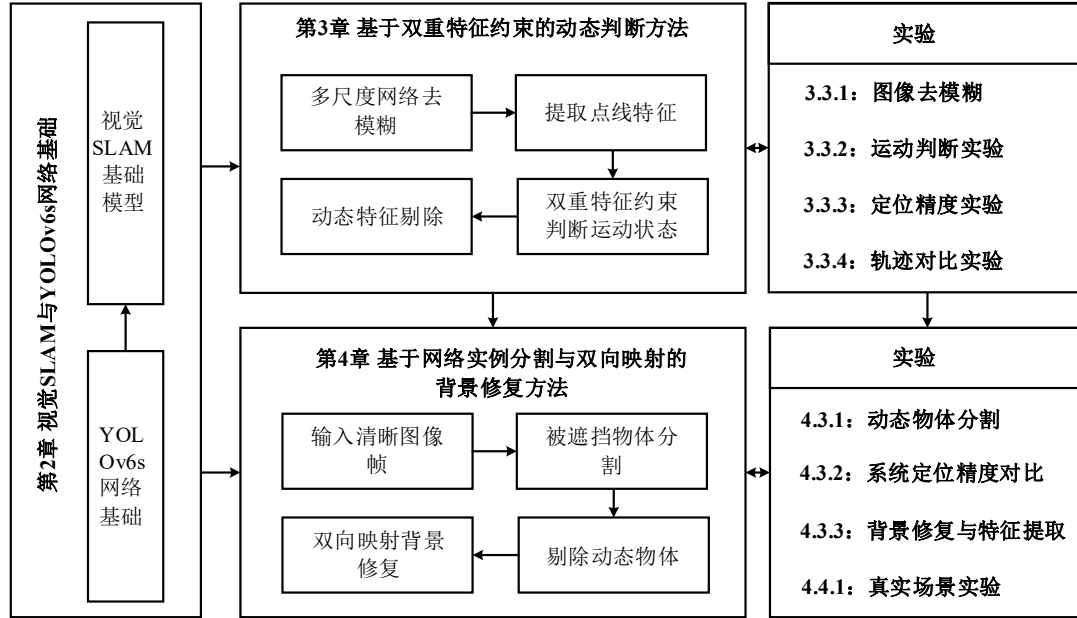


图 1-5 论文总体技术路线

各章节的主要内容如下：

第 1 章：绪论首先阐述研究背景及意义，强调在机器人技术广泛应用的背景下，视觉 SLAM 作为核心技术的重要性，特别是在动态遮挡场景下传统算法所面临的挑战。随后，对视觉 SLAM 技术按传感器类型进行分类，包括单目 SLAM、双目 SLAM 和 RGB-D SLAM，分析各自的优势与局限性。聚焦视觉 SLAM 研究热点，重点探讨动态遮挡场景下的算法现状，并剖析当前方法在物体分割、动态特征点识别及背景修复方面的不足。

第 2 章：基于 SLAM 与 YOLOv6s 网络的基础原理，本文深入分析其在视觉里程计、后端优化、回环检测及地图构建中的应用。随后，介绍 YOLOv6s 网络在图像处理中的工作原理，包括特征提取机制、多层次特征融合及动态物体识别。

第 3 章：针对动态场景中图像模糊对特征提取的干扰以及传统动态特征点识别方法存在的误判风险，本文提出了一种集成多尺度信息的自适应去模糊策略，以提升特征提取的准确性。并采用多层次神经网络架构，实现输入图像在不同尺度上的特征提取与融合，减少传统下采样过程中背景信息的损失，拓展网络的感受野，从而提高关键特征点的识别精度。在此基础上利用基于双约束的动态特征

筛选策略，对传统 Z-score 评分方法进行优化，通过调整基于四分位数间距的阈值，提高异常值检测的准确性，并增强其在不同数据分布场景下的适用性。进一步结合相机位姿估计与对极几何约束，精准判断特征点的运动状态，有效剔除动态特征点，降低动态目标对系统稳定性的影响。通过轨迹对比、地图构建及轨迹精度评估等实验，验证了所提算法的有效性以及在动态环境下的适应性与鲁棒性。

第 4 章：对动态遮挡环境下物体边界模糊及背景修复效果不佳的问题，对 YOLOv6s 模型进行了针对性改进，通过结合实例分割与双向映射机制提升目标检测和背景修复的准确性。在 SimSPPF 模块优化方面结合两种池化方法的优势，增强网络对边界信息的捕获能力，从而提升动态场景下物体分割的精度。在双分支 FPNB 特征融合模块上设计了双分支架构，高效整合跨层语义信息与同层多尺度特征，提升物体分割精度并为后续背景修复奠定基础。随后采用基于双向映射的背景修复模型挖掘前后帧中的静态信息，恢复图像的完整性。通过动态遮挡物体分割实验和背景修复实验对优化算法进行性能评估，并与现有算法进行对比来验证其在复杂动态遮挡场景下的分割与修复能力。

第 5 章：概述了基于改进 YOLOv6s 网络的 YDO-SLAM 算法在动态遮挡场景下视觉 SLAM 任务中的主要成果，包括在物体分割、动态特征点剔除和背景修复等方面的改进与优势，以及在实验中的优越性能表现。同时分析了研究中仍然存在的局限性，例如算法在极端复杂场景下的适应性仍有待进一步提升。最后，展望未来研究方向，提出可能的优化措施和改进策略为动态遮挡场景下视觉 SLAM 技术的深入研究提供参考。

第 2 章 视觉 SLAM 与 YOLOv6s 网络基础

本章重点探讨基于实例分割的视觉 SLAM 的基础理论。并分别对视觉里程计、后端优化、回环检测及地图构建等关键模块进行初步分析和介绍。随后阐述 YOLOv6s 网络模型语义分割从动态遮挡环境中提取语义信息的原理，为后续在动态遮挡场景下展开视觉 SLAM 研究提供了理论支撑^{[38][39][40]}。

2.1 视觉 SLAM 系统框架

视觉 SLAM 系统的基本结构由视觉里程计、后端优化、回环检测和地图构建四个模块组成。其中前端视觉里程计负责接收由传感器获取的图像与数据并为后续的任务提供必要的输入信息。闭环检测模块负责识别摄像头是否重新定位到同一位置，再通过前端视觉里程计和闭环检测的数据对相机位姿估计进行优化。后端优化模块根据位姿优化来调整相机姿态，避免系统定位跟踪的误差累积。地图构建模块基于相机位置姿态的估算结果，生成与实际环境相匹配的地图。视觉 SLAM 系统运行的流程如图 2-1 所示。

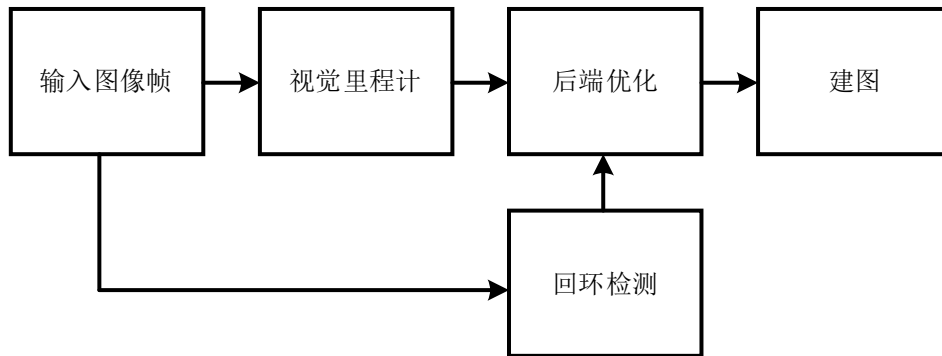


图 2-1 视觉 SLAM 基础框架图

2.1.1 视觉里程计

视觉里程计通过分析连续图像帧间的差异来估算相机的运动状态，并为后续的优化过程提供图像信息。从图像中提取特征点并通过匹配计算相机在两帧之间的运动和场景结构来实现运动估计。

视觉里程计的主要分为特征点法与直接法两大类。特征点法具有较强的稳定

性，并且对光照变化和动态物体的干扰不敏感，特征是图像信息的表达形式，常见的特征包括角点、边缘和区块，角点和边缘相较于区块更具代表性。在不同图像之间常见的特征提取方法是识别角点并建立其对应关系，从而实现匹配。角点被视为核心特征，常用的角点检测技术包括 Harris 角点、FAST 角点以及 GFTT 角点等。传统的角点检测方法在光照变化、尺度变化及旋转不变性方面存在一定的局限性，面对复杂场景时可能会丢失或匹配不准确，结合角点检测与特征描述的 ORB 算法能够有效应对光照变化、尺度变化和旋转变换带来的影响。ORB 是公认的高效且稳健的特征检测与提取算法，其架构包括特征点识别与特征描述两个核心步骤。角点的检测如图 2-2 所示。

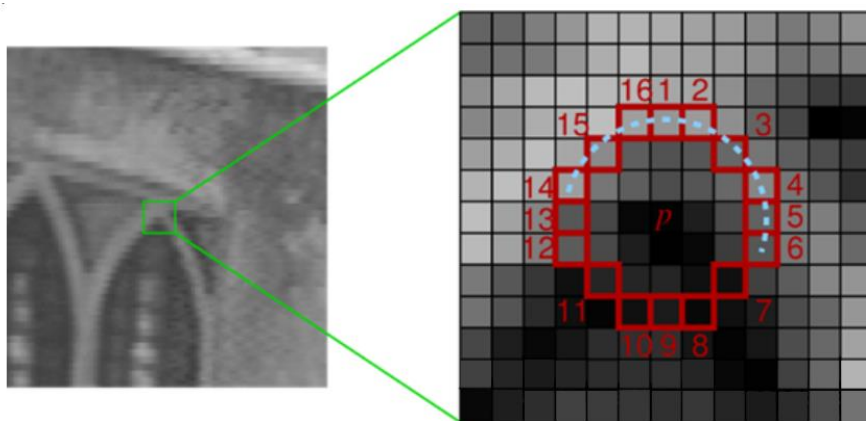


图 2-2 FAST 角点检测

2.1.2 后端优化

前端视觉里程计通过相邻帧图像间的特征匹配与运动估计逐帧推算相机位姿和构建局部地图。但实际运行中多种因素会导致前端估计误差的产生。相机传感器的噪声导致获取的图像存在不确定性，直接影响了基于图像的特征提取与匹配的准确性，使原本不对应的特征点错误匹配将会使相机位姿的估计产生偏差。随着相机不断采集新的图像帧并进行位姿推算，误差也会持续积累。

后端优化通过带有偏差的信息来评估系统状态，并度量评估的变异性，此过程与执行最大后验概率推断相一致。系统状态涉及了机器人运动轨迹和环境地图的构建。因此后端优化需要尽可能降低噪声的影响，并通过引入状态估计理论来量化定位与建图的不确定性提升 SLAM 系统性能。结合滤波器或非线性优化方法估计系统状态的均值与方差，从而提高 SLAM 结果的精度^[41]。

2.1.3 闭环检测

闭环检测的目标是消除随时间积累的定位误差。当机器人行驶一段距离后物理位置在理论上到达起始点，但累积误差导致估算位置偏离，此时闭环检测模块能够辨识重复访问的区域，并实施全局优化策略，校正由时间推移产生的漂移误差，增强 SLAM 系统的长期定位精度与鲁棒性。闭环检测效果图如图 2-3 所示。

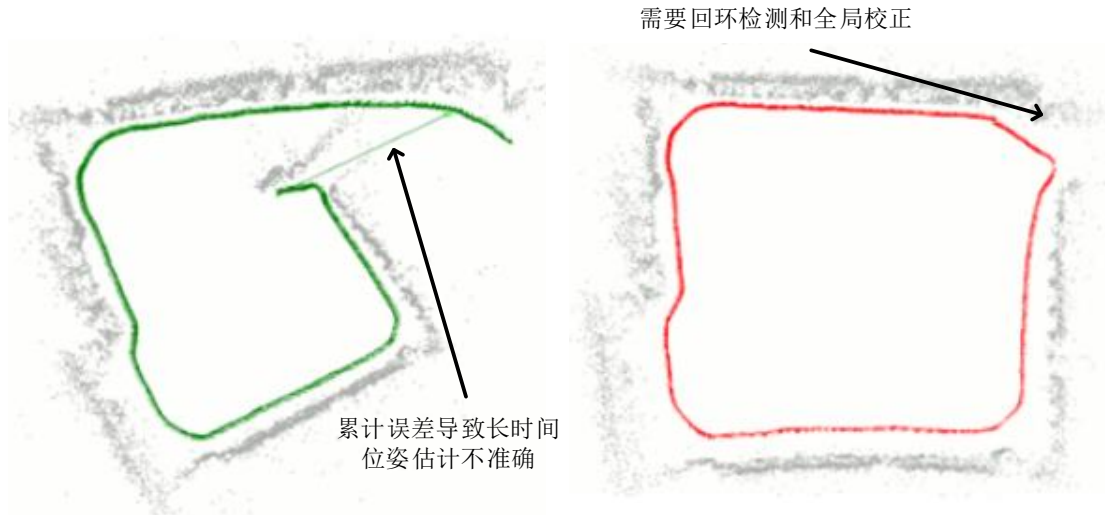


图 2-3 闭环检测效果图

闭环检测可通过多种策略实现。例如，对任意两幅图像执行特征匹配并评估准确匹配点的数量来判断两者间是否存在关联性。这种方法虽然简单有效，但假设任意两张图像都可能形成闭环的计算量较大，在 SLAM 系统长时间运行时容易导致冗余匹配。例如，当可能的回环数为 N 时，计算复杂度呈指数级增长，因此在大多数实时系统中难以应用。另一种方法是从过去 n 帧数据中随机选取 5 帧与当前帧匹配。该方法可保持恒定的计算时间，但由于采用随机抽样策略，随着帧数 N 的增加检测效率下降甚至遗漏闭环。

由于对不同图像匹配的方法在闭环检测中难以适用，通过引入词袋模型（Bag of Words, BoW）来优化匹配效率。词袋模型通过聚合大量特征描述子构建出视觉词典，用于表示图像中的局部特征。该模型用来统计图像中各视觉单词的出现频率，并形成向量化表示，从而将图像转换为特征向量，便于后续的相似性计算和匹配。

视觉词典由大量视觉词汇组成，每个视觉词汇代表一种特定的特征。例如，在图像分类任务中，人、车、狗等可作为视觉词典中的基本单元。在闭环检测中，

词袋模型利用图像的局部特征（如 SIFT、ORB 等）进行编码，并基于词典计算图像之间的相似度。假设任意图像 A 的特征分布可以表示为向量，则计算其在词典中的匹配关系，如公式 2-1 所示。

$$A = 1, w_1 + 1, w_2 + 0, w_3 \quad (1-1)$$

假设图像 A 的特征向量为 $V_A = [1, 1, 0]^T$ ，该向量可表示图像中特定类别特征的存在，无需依赖像素级灰度信息，与直接使用灰度值相比，词袋模型提供了更具鲁棒性的表征方式，使其在光照变化和局部遮挡等情况下依然具备稳定的匹配能力。由于用向量表示类别特征不受物体空间位置及排列序列的影响，只要目标在图像中存在，其特征向量 V_A 就不会变化。

词袋模型强调视觉单词的是否存在，而不考虑其具体排列顺序，因此视觉词典可视为类别特征的集合。假设图像 B 的特征向量表示为 V_B ，若不考虑视觉单词的数量分布，则 V_B 可被简化为二值特征向量，两幅图像之间的匹配程度可通过基于 V_A 和 V_B 的相似性评估方法进行计算，如公式 2-2 所示。

$$s(A, B) = 1 - \frac{1}{W} \|A - B\|_1, A, B \in R^W \quad (1-2)$$

范数采用 L_1 范数，即向量各元素的绝对值之和。当两个向量完全一致时，相似性评估结果为 1；若完全相反值为 0。这种评估方法用于反映图像之间的相似程度。

2.1.4 系统建图

视觉 SLAM 主要用于在无先验信息的环境中通过传感器实现位姿估计与环境建图。地图通常被定义为地标点的集合，目的是围为了精确定位这些地标点的位置并完成地图构建，根据地图的表达方式，SLAM 建图主要分为稀疏地图、稠密地图等多种形式。从定位的角度来看，稀疏地图虽然可用于相机位姿估计，但由于缺乏连续的表面信息，仅依赖有限的特征点进行空间推断，导致环境重建存在较大不确定性，从而限制了其在路径规划和避障等任务中的适用性，而稠密地图凭借其丰富的环境信息，在高精度环境重建、自主导航和机器人交互等 SLAM 应用中具有不可替代的优势^{[42][43]}。

2.2 YOLOv6s 网络动态物体分割

YOLOv6s 是一种高效的目标检测网络，在兼顾检测速度与精度的同时，采用轻量化设计并优化了整体网络架构。其网络由骨干网络（Backbone）、颈部网络（Neck）和头部网络（Head）三部分组成。主干网络负责提取低级与高级特征，为目标检测提供深层次的语义信息。颈部网络通过特征增强与多尺度特征融合，强化特征表达能力，提高检测的鲁棒性。检测头基于颈部网络的特征输出，执行目标分类与边界框回归，确保检测结果的准确性。该分层架构使 YOLOv6s 在保持高效推理能力的同时，进一步优化检测精度，并提升模型在复杂场景下的适应性。

2.2.1 整体架构概述

YOLO 在目标检测领域表现出卓越的性能优势，兼具检测速度快、精度高，并具备较强的实时性。YOLO 通过将目标检测任务转化为回归问题，实现端到端的检测流程，与传统的候选区域生成方法相比大幅提升了检测效率。通过将输入图像分解为若干网格单元，在每一单元内预测预定数量的边界框及其对应的类别概率。每个边界框由中心坐标、宽度、高度、置信度和类别概率构成，确保 YOLO 在单一正向传播过程中能够同步预测所有目标的边界框与类别标识。

在 YOLO 系列模型的发展过程中，YOLOv1 至 YOLOv3 被视为经典的单阶段（one-stage）目标检测框架，其优化重点在于改进特征提取网络，以提升检测精度和计算效率。例如，YOLOv4 引入 CSPNet 作为 Backbone，减少梯度信息冗余，提高计算效率与模型收敛速度。基于先前优化策略的改进，YOLOv5 进一步提升了检测精度与推理速度，在实际应用中展现出更优性能。借助其卓越的端到端检测性能及兼备高精度与高速度的特点，YOLO 系列已确立为目标检测领域的主导架构，普遍应用于自动驾驶、安全监控以及工业质量检查等多个领域。此外，YOLO 的网络设计思路为后续的一次性检测模型的创新与发展，提供了关键的理论依据与改进路径。

相较于先前的版本，YOLOv6s 在多个方面进行了优化，最显著的提升体现在网络架构的改进。YOLOv6s 借鉴了 RepVGG 的设计理念，强化了重参数化能

力。通过引入深度可分离卷积和跨阶段连接优化内存带宽的利用，使其在嵌入式设备或计算资源受限的环境中具备更强的适应性，有效提升了特征提取的准确性与计算效率。YOLOv6s 的网络模块设计如图 2-4 所示，可以看出多个计算块采用了 Rep 标记，表明该模型以 RepVGG 结构为核心进行优化。相较于 YOLOv5，YOLOv6s 的 GPU 利用率相对较低，提升计算性能并增强推理速度。

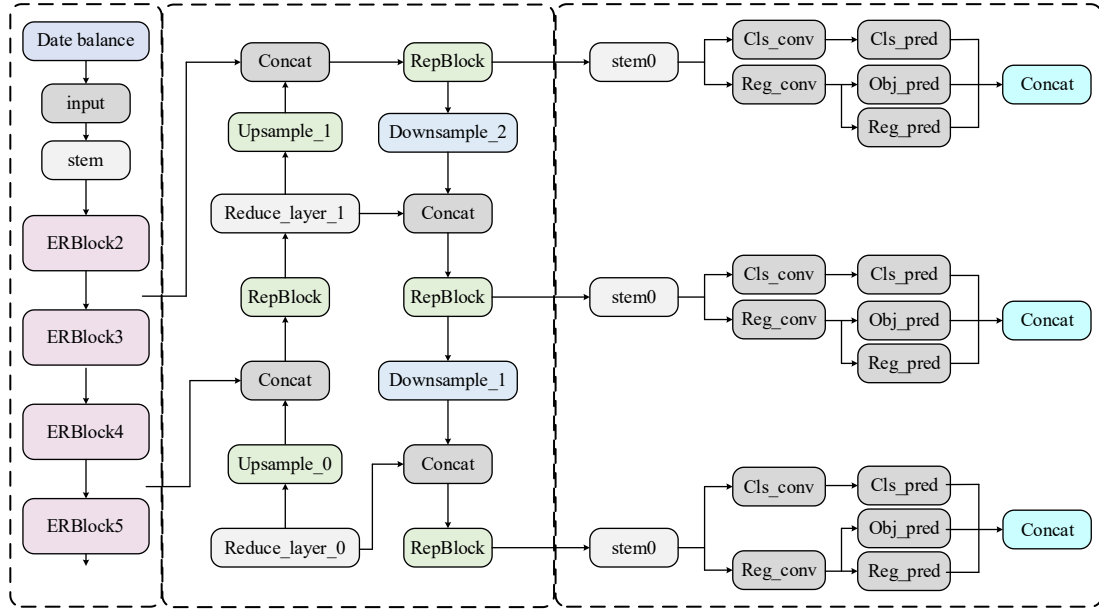


图 2-4 YOLOv6s 模型架构

2.2.2 头部网络预测机制

在动态遮挡场景下，目标检测网络的性能对比主要体现在能否准确的识别并剔除动态物体。然而，传统目标检测网络在该环境下，容易出现边界识别模糊、目标错检或漏检的问题，进而导致 SLAM 系统特征点提取和位姿估计的效果不佳。

头部网络是目标检测网络架构中的关键部分，根据特征提取网络和颈部网络的输出，进行目标类别预测和边界框回归。并采用多尺度预测机制，在不同尺度特征图上进行目标检测，以增强对不同层次目标的识别能力。通过对各尺度特征图的信息融合与分析，该网络能够更精准地检测大小不一的目标。使 SLAM 系统在动态场景下能更精确地区分动态与静态目标，提高特征点提取的精确度，减少误剔除的情况。此外，优化的目标边界预测能更好地应对物体遮挡问题，使得系统在复杂环境中保持稳定的跟踪效果，提升全局地图的构建质量。

2.2.3 物体的动态特征剔除

YOLOv6s 在物体分割时采用了特征提取与分析、多尺度特征融合、基于检测结果的分割转换以及训练优化等方法。骨干网络通过层层卷积不断提取图像特征，将颜色、纹理、形状等低级视觉信息逐步转换为深层次的特征表示。例如，浅层卷积主要关注边缘、角点等基础特征，而随着网络的加深，其提取的特征逐渐向更具语义信息的特征演化，为后续的物体分割提供坚实基础。颈部网络使用 Rep-PAN Neck 结构，以步长为 2 的重参数化卷积层减少内存需求，同时通过自上而下和自下而上的通道融合不同层次特征，高层语义特征和底层细节特征相互补充，让网络可以精准地定位物体边界。

在动态遮挡环境下处理图像中因物体移动造成的模糊问题，网络在分割过程中通常采用双线性内插法和上下采样等图像处理方法。其中双线性内插法通过计算相邻像素的加权平均值，填补由运动模糊或遮挡造成的像素信息缺失，从而恢复图像的清晰度。双线性内插法基于局部区域内的像素值进行插值，从而平滑图像中的不连续变化，减少了动态物体快速移动时引起的局部模糊现象。其次，物体移动时在图像中的尺度会不断发生变化，为使网络能够处理不同尺度的目标，YOLOv6s 网络中的多个模块和通道组成了颈部网络。颈部网络在整合语义信息和融合多尺度特征的过程中多次使用上下采样的操作。上采样通常在网络的深层使用，在网络检测较大物体时通过反卷积或转置卷积的方式将低分辨率特征图尺寸放大，将每个像素点扩展到更大的尺寸，填充缺失的空间信息，从而保留更多的空间细节。下采样则通常在网络的较浅层使用，通过应用步长大于 1 的卷积操作将特征图的尺寸缩小，并压缩图像中冗余的空间信息，从而减少计算量并聚合局部信息。

针对动态遮挡环境下动态物体移动时造成边缘模糊和尺度变化的问题。通过双线性内插法和上下采样方法使网络能够填补由运动模糊和遮挡造成的信息缺失，并从不同尺度的特征中分离出静态背景与动态物体，平衡计算量的同时压缩冗余信息，从而降低动态物体对特征提取与匹配过程的干扰，提升系统的整体精确性和鲁棒性，使 YOLOv6s 网络在动态环境下能够提供更加精准的目标检测与定位。双线性内插法和卷积如图 2-5 所示。

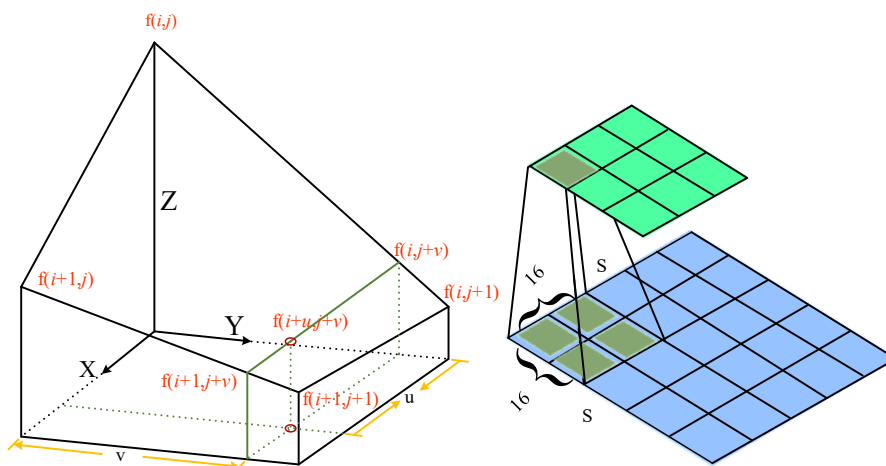


图 2-5 双线性内插法和卷积原理示意图

视觉 SLAM 算法通常假设场景为静态环境，而场景中的动态物体是导致系统偏差的主要因素。为此，本章采用 YOLOv6s 实例分割网络来识别动态目标区域，并在特征提取与匹配过程中剔除动态物体的影响。YOLOv6s 实例分割网络剔除动态特征如图 2-6 所示。

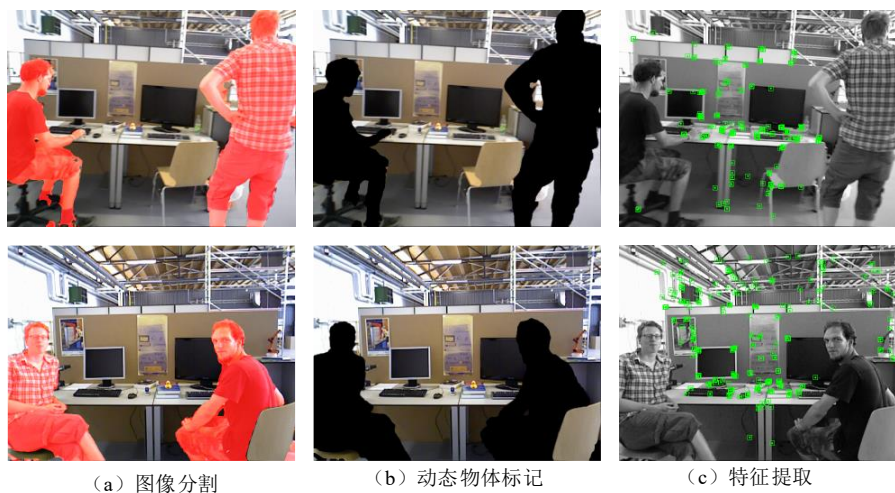


图 2-6 YOLOv6 网络分割剔除动态特征

2.3 本章小结

本章通过对视觉 SLAM 和 YOLOv6s 网络的介绍，叙述了 YOLOv6s 网络在处理动态遮挡物体方面的优势，明确了视觉 SLAM 算法结合 YOLOv6s 网络对移动机器人在动态遮挡环境下定位与环境感知中的提升，为后续改进 YOLOv6s 网络以提升视觉 SLAM 在复杂场景下的适用性提供了必要的理论支撑。

第 3 章 基于双重特征约束的动态判断方法

在 SLAM 系统运行时，特征点的质量与数量对系统的定位精度和跟踪稳定性起着至关重要的作用。目标检测生成的检测框内可能同时包含动态物体与静态背景信息，采用更精确的网络分割与动态特征剔除策略，提升 SLAM 在动态遮挡环境下的适用性。

传统的方法直接移除检测框内的所有特征点，但在背景区域占比较大的情况下，该方法可能导致大量静态特征点的误剔除。尽管后续研究提出了多种优化方案，如在检测框内保留部分静态特征点，但动态区域在整体图像中占比较大的情况下，优质静态特征点的误剔除现象仍然较为显著。不仅影响系统的稳定性，还可能导致跟踪失败，进而影响 SLAM 的整体性能。因此，在动态环境下准确判断静态特征点并保留，是提升 SLAM 系统稳定性的关键。

YDO-SLAM 采用基于深度信息的 Z-score 评估与对极几何约束相结合的双重特征点筛选方法，有效提升了系统的鲁棒性，确保其在复杂动态场景中依然能够实现高精度位姿估计与稳定的地图构建。

3.1 局部自适应的去模糊方法

在相机高速移动时，图像容易出现运动模糊，根本原因在于像素在曝光期间发生位移和叠加。当相机与场景物体之间存在较快的相对运动时，模糊现象会更加显著。在视觉 SLAM 任务中，快速移动不仅会导致图像模糊，还会显著降低特征提取与匹配的准确性，进而影响系统的定位精度与地图构建质量。因此采用合适的去模糊方法对于提升 SLAM 的稳定性至关重要。

3.1.1 传统去模糊方法

图像模糊可被划分为若干种类别，其中运动模糊因其在实际应用中的普遍性及处理难度较高，成为研究重点。主要原因在于拍摄时相机的移动或物体的位置变化，造成图像像素在成像过程中产生叠加效应。在相机维持曝光期间，可假定

传感器于每一瞬时所捕获的信号呈累积状态。若拍摄周期保持极短，则可假定在每一瞬时，传感器捕获的信号均是明确无误的。倘若拍摄过程中相机与被摄物体产生相对移动，则会导致最终图像像素失位而显现运动模糊现象。其形成过程，如公式 3-1 所示。

$$I_g = g\left(\frac{1}{T} \int_{t=0}^T S(t) dt\right) = g\left(\frac{1}{M} \sum_{i=0}^{M-1} S[i]\right) \quad (2-1)$$

其中 $S(t)$ 表示在时间 t 下传感器捕获的信号， $S[i]$ 对应着第 i 个清晰信号， I_g 表示生成的模糊图像， M 则表示通过采样得到的清晰信号总数， T 为曝光持续的时间，而 g 则表征了非线性的相机响应函数。此函数转换将把传感器捕获的连续信号 $S(t)$ 转译为离散的数字图像 $\tilde{S}(t)$ ，即 $\tilde{S}(t) = g(S(t))$ ， $\tilde{S}(t) = g(S(i))$ 。把曝光时间段内所有传感器信号涵盖的内容进行叠加，便能生成拍摄所得的图像。若在该过程中相机与物体发生相对移动，则生成的图像将显现为运动模糊的特征。

图像去模糊的目标是恢复模糊图像的清晰度，使其尽可能接近原始未模糊状态。该过程通常依赖于图像运动模糊的数学模型，通过对成像过程中像素模糊机制的解析，推导出合理的去模糊策略。如公式 3-2 所示。

$$g(x, y) = h(x, y) \otimes f(x, y) + n(x, y) \quad (2-2)$$

其中 $g(x, y)$ 代表模糊图像， $h(x, y)$ 是退化系统函数，即点扩散函数， $f(x, y)$ 为清晰图像， $n(x, y)$ 是加性噪声， $\otimes$ 表示卷积运算。图像去模糊是上述过程的逆过程， $h(x, y)$ 和 $n(x, y)$ 越精准，去模糊效果便越好。以下是两种普遍的传统去模糊技术，并对其局限性进行了探讨。

维纳滤波是一种基于频率域的去模糊算法，而 Lucy-Richardson (LR) 算法则在时域内进行运算。维纳滤波旨在让去模糊后的图像 $\hat{f}(x, y)$ 与原清晰图像 $f(x, y)$ 的差距最小化，这种差距以均方差衡量，可实现去模糊与去噪。其均方差表达式，如公式 3-3 所示。

$$e^2 = E\left[f(x, y) - \hat{f}(x, y)\right]^2 \quad (2-3)$$

通过最小化 e^2 求解去模糊后的图像 $\hat{f}(x, y)$ ，其在频率域的表达，如公式 3-4 所示。

$$\hat{f}(u, v) = \frac{k^*(u, v)}{|k(u, v)|^2 + \frac{s_n(u, v)}{s_f(u, v)}} G(u, v) \quad (2-4)$$

其中 $k^*(u, v)$ 是点扩散函数在频率域的复共轭, $S_n(u, v)$ 和 $S_f(u, v)$ 分别是噪声 n 与原清晰图像 f 的功率谱, $G(u, v)$ 为模糊图像的傅里叶变换。从式 3-4 可知, 应用维纳滤波需明确 $S_n(u, v)$ 和 $S_f(u, v)$, 若无先验信息, 这两项难以确定, 因而限制了其应用场景。

LR 算法的去模糊理论基础包括贝叶斯理论、泊松分布以及最大相似估计, 以此对模糊图像进行复原处理。在 LR 算法中, 模糊图像 g 由原始清晰图像 f 经过退化过程生成。根据最大相似理论, 目标是最大化 $P(f|g)$ 的概率, 并通过贝叶斯公式进行求解, 以迭代优化的方式恢复清晰图像, 如公式 3-5 所示。

$$P(f|g) = \frac{P(g|f)P(f)}{P(g)} \quad (2-5)$$

由于 $P(f)$ 与 $P(g)$ 为常数, 所以关键在于求 $P(g|f)$ 的最大值, 这通过泊松统计模型来实现。泊松统计模型, 如公式 3-6 所示。

$$P(x) = \frac{u^x e^{-u}}{x!} \quad (2-6)$$

其中变量 u 代表指定时间段内事件发生频率的平均值。假定模糊图像中各个像素点彼此独立, 依据泊松统计的数学框架, 条件概率分布 $P(g|f)$, 如公式 3-7 所示。

$$P(g|f) = \prod_{(x,y)} \frac{h(x,y)f(x,y)^{g(x,y)} e^{(h(x,y)f(x,y))}}{g(x,y)!} \quad (2-7)$$

通过求解极值, 从而得到迭代求解结果, 如公式 3-8 所示。

$$f^{n+1}(x, y) = \left[\frac{g(x, y)}{h(x, y)f^n(x, y)} h(x, y)^T \right] f^n(x, y) \quad (2-8)$$

L-R 算法存在一定局限性。若未明确指定迭代结束的条件, 将可能导致计算资源的不断消耗。此外, 还会出现迭代算法的震荡现象, 或者收敛于局部极值点, 难以得到真正的全局极值, 因此需要选取合适的初始值。

3.1.2 基于注意力引导的局部自适应去模糊方法

在动态遮挡场景中, 对于图像模糊处理主要分为全局运动模糊和局部动态模

糊，当相机高速运动时，图像会出现全局运动模糊，使得传统特征点检测方法难以准确提取有效信息。当动态目标运动时会引起局部模糊，但如果对所有区域采用相同的去模糊处理，容易导致关键动态目标信息被过度修正，从而影响后续特征匹配和定位。

通过分析实验，构建了一种计算效率高且能有效增强特征点稳定性的去模糊方法。根据 YOLOv6s 分割将图像细分为动态与静态区域。针对这两类区域实施差异化的去模糊策略，在静态区域内实施强度较大的去模糊操作，而在动态区域内则采取较为温和的去模糊操作，旨在保全动态目标的特征信息，同时增强背景的清晰度。采用基于导向滤波的自适应去模糊策略，通过事先构建模糊权重图来调控不同区域的去模糊程度，以此实现对特征提取及匹配过程的优化。首先，利用 YOLOv6s 网络已经对图像进行了动态目标检测，得到了动态目标区域 R_d 以及背景区域 R_s 。随后，为两种不同的区域赋予不同的去模糊权重。

对于动态区域，为避免过度修正导致动态目标特征丢失，选择较低的去模糊强度。对于静态背景区域，通过强去模糊处理来提高图中背景细节清晰度，便于后续特征点提取和匹配。模糊权重 $W(x, y)$ 定义，如公式 3-9 所示。

$$W(x, y) = \begin{cases} \alpha, & \text{if } (x, y) \in R_d \\ \beta, & \text{if } (x, y) \in R_s \end{cases} \quad (2-9)$$

其中 $0 < \alpha < 0.5$ 表示动态区域采用较弱去模糊， $0.5 < \beta < 1$ 则表示背景区域采用较强去模糊。

引导滤波是一种具备边缘保真特性的滤波技术，能够基于局部属性动态调整图像的平滑程度。对模糊图像施以引导滤波处理，并通过先前构建的权重图 $W(x, y)$ 进行强度差异的去模糊操作。假设输入模糊图像为 $I_b(x, y)$ ，先对其进行高斯模糊操作 $F(I_b)$ ，进一步提取图像的低频信息，然后计算去模糊后的图像，如公式 3-10 所示。

$$I_c(x, y) = I_b(x, y) - W(x, y)(F(I_b)(x, y) - I_b(x, y)) \quad (2-10)$$

其中 $W(x, y)$ 较大时判定为静态背景区域， $F(I_b) - I_b$ 的差值被大幅修正，从而得到较强的去模糊效果；当 $W(x, y)$ 较小时判定为动态区域，修正较弱，能够保留动态目标的局部细节，避免特征信息过度丢失。

为进一步提升图像模糊区域的定位精度和去模糊处理的适应性，在

YOLOv6s 网络对动态与静态区域初步划分后，构建了通道注意力（Channel Attention）和空间注意力（Spatial Attention）模块，用于分别挖掘图像在通道和空间维度上的重要性特征。

对于静态区域，注意力机制加强了边缘和纹理特征的响应，使系统对于清晰背景区域的恢复效果更佳；对于动态区域，则适当抑制无效纹理噪声，避免对动态目标进行过度锐化。结合模糊权重图，通过将注意力权重与权重图 $W(x,y)$ 进行加权融合，实现了像素级别的自适应去模糊控制。

本节提出的基于注意力引导的局部自适应去模糊方法通过较强的去模糊处理，解决了静态背景区域模糊严重问题，提高背景区域的细节清晰度，增强特征点提取效果。并且能处理动态区域过度修正的问题，利用较弱的去模糊策略，保留动态目标的特征信息，避免误剔除有效特征点，提升了 SLAM 特征提取的稳定性。通过考虑不同区域的模糊情况，实现自适应去模糊，提高 ORB 特征提取的鲁棒性，进而显著提升 SLAM 系统的定位与追踪准确性。本方法计算量低、易于实现，并能较好的融合现有的 YOLOv6s 分割与背景修复模块中，为动态遮挡场景下的视觉 SLAM 算法提供有效的图像优化。

3.2 基于双重约束筛选动态特征

在 SLAM 跟踪任务中，系统的定位精度和稳定性在很大程度上依赖于特征点的质量和数量。然而，在动态环境中，运动目标的干扰常导致特征点提取出现误匹配，从而引发累积误差，甚至导致跟踪失败。目标检测线程生成的检测框可能覆盖动态物体，也可能包含静态背景中的部分区域。因此，如何精准筛选背景特征点、剔除动态物体特征，并提升 SLAM 系统的鲁棒性，成为亟待解决的关键问题。

传统的动态特征判断方法会剔除动态物体检测框内的全部特征，这种方法虽能有效地屏蔽动态目标的干扰，但在框体范围占比较大的情况下，大量的静态特征点被误剔除，导致系统因特征点数量过少而跟踪丢失。为此，YDO-SLAM 采用结合深度信息的 Z-score 评分与对极几何约束的双重筛选策略，以增强系统在动态环境下的适应能力。在消除动态特征的同时，尽可能的保留背景中的有用特征，以此来增强 SLAM 系统的稳定性和定位准确性。

3.2.1 结合改进 Z-score 评分判断动态特征

传统 SLAM 的定位跟踪主要基于静态特征点，但在动态场景下大量的静态特征点可能被误剔除，特征点数量的显著减少导致系统跟踪过程中位姿估计不准确而跟踪丢失。

针对上述问题，本节算法对剔除特征点的策略作出改进。首先提取特征点、线，对世界坐标系下点的空间坐标定义为 $P(X, Y, Z)^T$ ，其相对的平面像素坐标为 $P_C(u, v)$ 。特征线采用时间复杂度为 $O(n)$ 的 LSD 提取线特征，使用 LBD 描述符进行线特征匹配。对于世界坐标系中，线的空间坐标可定义为 $L_W = (L_S, L_E)$ ，其中 L_S 和 L_E 分别代表线在空间坐标的起点和终点，相对的平面像素坐标为 $l_C = (l_S, l_E)$ ， l_S 和 l_E 分别代表线在平面像素坐标的起点和终点。由此可以得出线中点的空间坐标和平面像素坐标分别为 $L_M = \frac{(L_S + L_E)}{2}$ 和 $l_M = \frac{(l_S + l_E)}{2}$ 。

近年来，根据深度图像的深度信息来筛选动态点的方法较为成熟，例如利用 RANRAC^[45] 算法检测出正常深度值和异常深度值来剔除离群点，而 Dynamic-VINS^[46] 通过检测框四个顶点的深度值与中心深度值进行比较，设定一个标准深度值，当特征点深度值小于标准深度值，则将其视为动态特征点。但这些方法都基于动态物体轮廓占据检测框很大或者动态物体上的深度值整体处于波动不大的区间情况下适用。当相机俯视或仰视时，从动态物体提取的特征点的深度值范围可能较大，导致深度信息无法很好地区分动态点。因此，YDO-SLAM 采用双重特征约束方法，第一阶段基于改进的 Z-score 评分剔除动态点，二阶段通过点对级几何约束的动态点剔除方法（将深度值大于 0 的特征点认为是有效点）。

在第一阶段，首先计算有效特征点的深度值均值 μ 和标准差 σ ，如公式 3-11 所示。

$$\begin{cases} \mu = \frac{1}{n} \sum_{i=1}^n x_i \\ \sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2} \end{cases} \quad (2-11)$$

其中 μ 为特征点的深度均值， σ 为特征点的整体深度标准差， x_i 为检测框内深度值样本， n 为有效特征点的数量。Z-score 的计算公式，如公式 3-12 所示。

$$Z = \frac{x_i - \mu}{\sigma} \quad (2-12)$$

Z-score 评分为最常用的数据异常点评价方法，由于均值与标准差对于异常值都较敏感，假设有一个较大的样本值在样本内，则会直接影响到样本的均值和标准差，因此在动态环境下直接使用 Z-score 评分时是不可靠的。为了让对特征点深度值异常值的检测效果更好，本文对 Z-score 的评分方法作出改进，根据四分位数差的大小，调整 Z-score 的阈值，如果数据集的四分位数差较大，则可能表明数据的分布范围较广，因此可以使用较大的 Z-score 阈值来识别异常值；反之，如果四分位数差较小，则可能表明数据的分布范围较窄，因此可以使用较小的 Z-score 阈值来识别异常值。利用四分位数差 IQR 不受极端值影响的优势结合 Z-score 评分，可以更灵活地检测数据集中的异常值，旨在提升对异常值检测的鲁棒性和准确性。基于 IQR 修正的 Z-score 定义，如公式 3-13 所示。

$$\begin{cases} Z_{QAD} = \left| \frac{x_i - \tilde{X}}{K \times QAD} \right| \\ QAD = (Q3_{AD} - Q1_{AD}) \\ \sigma \approx K \times QAD \end{cases} \quad (2-13)$$

其中 Z_{QAD} 为修正后整个样本特征点不同深度值的评分， $\tilde{X}$ 是整个样本的中位数， $Q3_{AD}$ 为整个样本的上四分位数， $Q1_{AD}$ 为整个样本的下四分位数， QAD 为四分位数绝对偏差， K 值 0.6745 为常用的类型数据分布修正系数。

修正系数是基于正态分布的性质推导得到的，使满足正态分布的数据四分位数绝对偏差与标准差之间的比例差异，通常为正态分布位于均值一个标准差范围内的概率密度函数值的中位数。因此，当样本中特征点的 Z_{QAD} 超出阈值 $Thre$ 范围时，其为特征点深度值异常值，视该特征点为检测框内的动态点并剔除掉。但当特征点的数据分布不满足正态分布时，固定阈值通常会错检或漏检，在根据多次实验调优后，本文采用了一种基于分位数差值的自适应阈值方法，该方法既能适用于满足正态分布的数据集，亦能有效处理非正态分布的数据集。阈值 $Thre$ 的设定，如公式 3-14 所示。

$$\begin{cases} Thre^L = Q1_{AD} - 2QAD \\ Thre^U = Q3_{AD} + 2QAD \end{cases} \quad (2-14)$$

其中设定阈值上限 $Thre^L$ 与阈值下限 $Thre^U$ ，若特征点的数值 Z_{QAD} 超过阈值范围，则将其识别为动态特征点。

在第二阶段，由于检测框内仍可能存在未检测出的动态特征点，因此采用对极几何约束进行二次筛选。设 I_1 和 I_2 分别表示连续两帧图像的成像平面，相机从前一帧到当前帧的运动由旋转矩阵 R 和平移向量 t 描述。空间点 P 在两帧中的投影点分别为 P_1 和 P_2 ，相机光心分别为 O_1 和 O_2 。光心连线 O_1O_2 与成像平面的交点分别为极点 e_1 和 e_2 ，而 P_1e_1 和 P_2e_2 的连线 l_1 和 l_2 被称为极线。对极几何约束模型对于相机在不同角度下观察同一空间点 P 的关系，如图 3-1 所示。

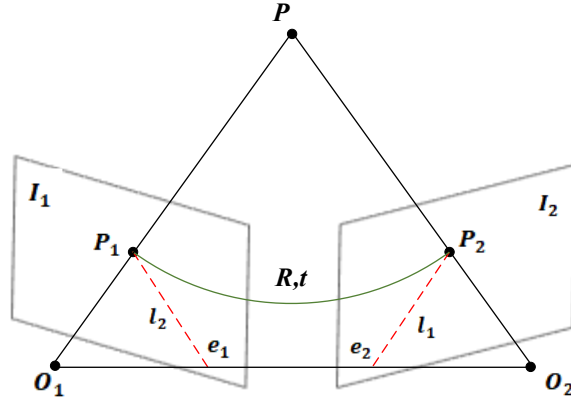


图 3-1 对极几何约束

特征点的齐次坐标 P_1 和 P_2 ，如公式 3-15 所示。

$$P_1 = [u_1, v_1, 1], P_2 = [u_2, v_2, 1] \quad (2-15)$$

其中， u 和 v 表示图像像素坐标系中的特征点的坐标值，当前帧中的极线 l_2 可以根据基础矩阵 F 和 P_1 计算得出，如公式 3-16 所示。

$$l_2 = [X, Y, Z]^T = FP_1 = [u_1, v_1, 1]^T \quad (2-16)$$

其中 X 、 Y 表示线向量。 P_2 点到极线的距离 d ，如公式 3-17 所示。

$$d = \frac{|P_2^T FP_1|}{\sqrt{\|X\|^2 + \|Y\|^2}} \quad (2-17)$$

若距离 d 的数值超出设定的经验阈值，则认定其为动态特征点。在跟踪线程中提取到的部分特征线也在动态物体上，动态特征线也同样被剔除。当特征线的中点坐标 L_M 落在检测框内，就判定该特征线为动态线并剔除掉。

3.2.2 基于点线特征的相机位姿优化

在跟踪线程中提取了关键特征，因此需要对投影误差函数重新定义。相机姿态的相对变换通过矩阵 $E_{(R,t)}$ 来表示，从世界坐标系到相机坐标系的旋转矩阵为

R, t 则对应于相应的平移矢量。该变换关系用于计算特征线在不同帧中的投影误差，从而优化 SLAM 系统的跟踪精度，如公式 3-18 所示。

$$E_{(R,t)} \in SE(3) = \left\{ E_{(R,t)} = \begin{pmatrix} R & t \\ 0 & 1 \end{pmatrix} \mid R \in SO(3) \right\} \quad (2-18)$$

设点 I_k 为世界坐标中的一个 3D 点， s_k 为该点的 2D 像素坐标，点特征的最小化重投影误差 e_p ，如公式 3-19 所示。

$$\begin{cases} e_p = s_k - \pi(I_k, K, E_{(R,t)}) \\ E_{(R,t)} = \exp(e_{(R,t)}) \end{cases} \quad (2-19)$$

其中 π 为投影函数， K 为相机内参， $e_{(R,t)} \in SE(3)$ 为相机姿态的相对变换向量。

世界坐标系下线的空间坐标为 $L_W = (L_S, L_E)$ ， L_S 和 L_E 分别是线的空间坐标的起点和终点。对应的平面像素坐标为 $l_C = (l_S, l_E)$ ， L_M 是 L_S 和 L_E 的中点， l_S^h 、 l_E^h 、 l_M^h 是相应的齐次坐标，以此定义线系数，如公式 3-20 所示。

$$I_{(R,t)} = \frac{l_S^h \times l_E^h}{|l_M^h|^2} \quad (2-20)$$

将线特征重投影误差 e_l 定义为投影线段的起点、中点和终点到与图像平面中检测到的线的距离总和。得到线重投影函数，如公式 3-21 所示。

$$e_l = I_{(R,t)}^T \pi \left((l_S, K, E_{(R,t)}) + (l_M, K, E_{(R,t)}) + (l_E, K, E_{(R,t)}) \right) \quad (2-21)$$

通过整合点特征和线特征的误差函数得到一个总的代价函数，如公式 3-22 所示。

$$\{R, t\} = \arg \min \sum_i^A \varepsilon \left(e_p^T(e_{(R,t)}) \sum_p^{-1} e_p(e_{(R,t)}) \right) + \sum_j^B \varepsilon \left(e_l^T(e_{(R,t)}) \sum_l^{-1} e_l(e_{(R,t)}) \right) \quad (2-22)$$

其中 A 和 B 分别为特征点、线的集合，分别包含 i 个点和 j 个线， ε 为惩罚因子， $\sum p \sum l$ 为 R, t 的协方差矩阵。

3.3 实验结果与分析

本文的实验平台配置如下：CPU 为 Intel Core i5-12400F，内存为 16 GB；GPU 为 RTX 3060 显卡，12 GB 显存；在 Ubuntu 18.04 平台上运行本文算法。为了检验其有效性，实验采用 TUM 公开数据集进行了测试及评估。

3.3.1 图像去模糊实验

本实验针对动态遮挡场景下的图像模糊问题,验证基于注意力引导的局部自适应去模糊方法的有效性,如图 3-2 所示。

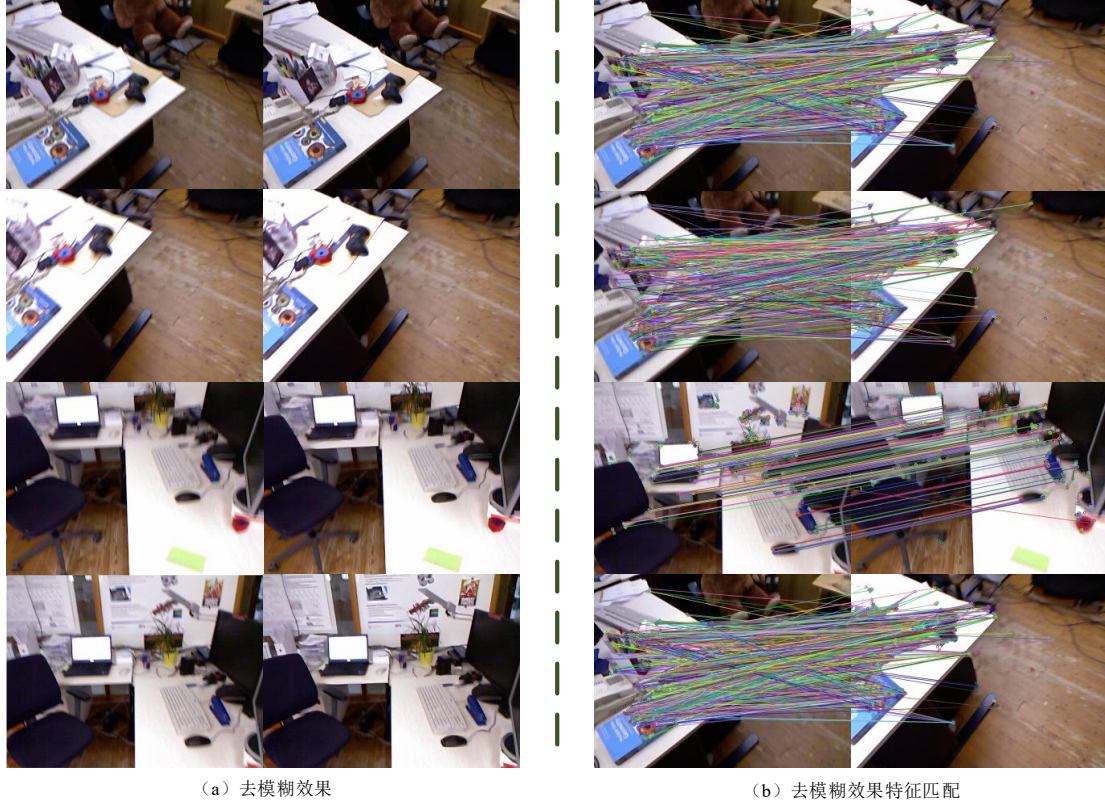


图 3-2 去模糊效果及特征点匹配图

图 3-2(a)为去模糊后的图像效果,而图(b)为去模糊后的特征点匹配情况。从图(a)可以观察到,基于所提方法处理后的图像在边缘细节、纹理恢复方面取得了显著提升,物体的轮廓更加清晰,场景结构得到较完整的复原。表明采用的局部自适应去模糊策略能够有效解决动态场景中的运动模糊问题,并提高图像中特征信息的可辨识度。图(b)进一步验证了去模糊方法对特征点匹配的影响。可见去模糊后图像中的特征点匹配数量显著增加,且匹配线条更为均匀分布,表明提取的特征点质量和匹配的正确率更高。表明注意力引导机制的有效性,使网络能赋予图像中运动模糊影响区域不同权重,并通过局部去模糊增强特征点的可识别性。

综上所述,本实验结果表明所提出的去模糊方法在动态遮挡场景下能够有效提升图像清晰度,并提升特征点匹配效果,为后续视觉 SLAM 任务中的特征提取、跟踪及地图构建提供了更加稳定的图像信息。为了进一步验证上述方法的合理性,本节通过特征点提取实验去对比不同去模糊策略下 ORB 特征点的数量和

匹配精度，如图 3-3 所示。

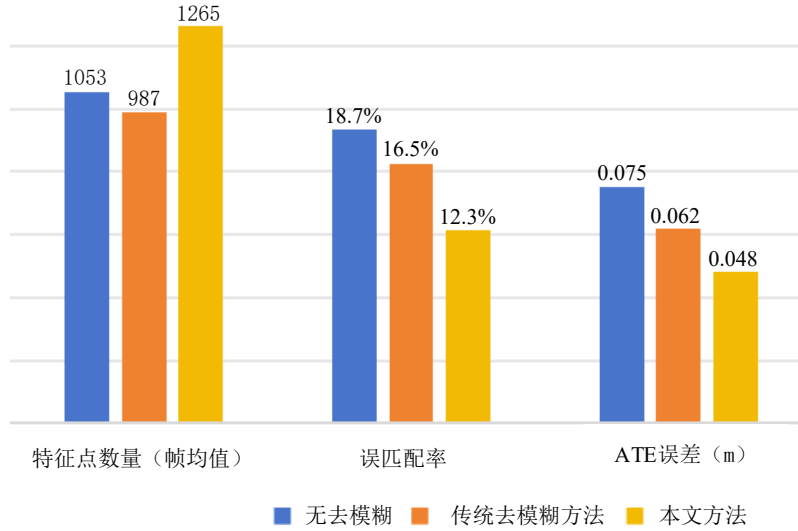


图 3-3 匹配精度对比图

实验结果显示，背景区域的特征点数量显著增加，同时动态区域的误匹配率降低。实验中采用绝对轨迹误差（ATE）评估定位精度。实验数据表明，该方法较传统去模糊方法能够进一步降低定位误差，提高 SLAM 系统整体稳定性。结果验证了基于注意力引导的局部自适应去模糊方法在动态遮挡场景下的有效性，该方法能在不改变整体 SLAM 系统架构的前提下，提高特征提取和定位精度。

3.3.2 动态特征剔除实验

实验采用 TUM RGB-D 数据集进行算法性能评估。TUM 数据集是由德国慕尼黑工业大学发布的一个高质量 RGB-D 图像序列数据集，广泛应用于 SLAM 与三维重建等研究中。数据集提供了丰富的场景，包括静态和动态环境，并包含高精度的姿态真值信息，适用于评估视觉里程计和轨迹估计算法的准确性。本文选取了其中包含动态干扰的序列用于验证方法的鲁棒性。

本节选取自 TUM 数据集的高动态序列包括 W_halfsphere、W_rpy、W_static、W_xyz，这些序列中包含不同程度的动态遮挡，如人员行走或坐姿移动。选取这些序列是因为它们在动态性上具有典型性，能覆盖从低遮挡密度到高遮挡密度的不同情境，具有代表性地反映算法在复杂动态环境下的鲁棒性表现。旨在评估对动态遮挡物运动状态的判定能力以及动态特征的去除效果，如图 3-4 所示。图 3-4 直观展示动态物体剔除的效果，选取 TUM 数据集中任意三个原始动态帧图 3-4(a)、(e)、(i)，随后，通过特征提取并将图 3-4(b)、(f)、(j)将点和线特征可视化，

并通过 YDO-SLAM 算法进行动态特征剔除，图 3-4(c)、(g)、(k)所示为动态特征剔除图像帧。



图 3-4 点、线特征动态物体剔除效果

对比图 3-4(d)、(h)、(l)，可以明显观察到 YDO-SLAM 成功识别并剔除了动态物体（蓝色表示动态特征，绿色代表静态特征），显著降低了动态物体对地图构建过程的影响。这一过程不仅提升了系统的定位与跟踪精度，有效减少了轨迹误差。

3.3.3 轨迹对比与建图实验

本文选取 TUM 数据集进行验证评估，图为 ORB-SLAM3、DS-SLAM、

DynaSLAM 和 YDO-SLAM4 种算法在 TUM 数据集下的 4 个动态序列 W_halfsphere、W_rpy、W_static、W_xyz 中运行的轨迹图对比，图上的蓝色曲线代表在 4 种算法作用下的相机运动路径，黑色曲线对应实际的相机轨迹，红线则展示了这 4 种算法的操作路径与真实轨迹之间的轨迹偏差，轨迹对比图如图 3-5 所示。

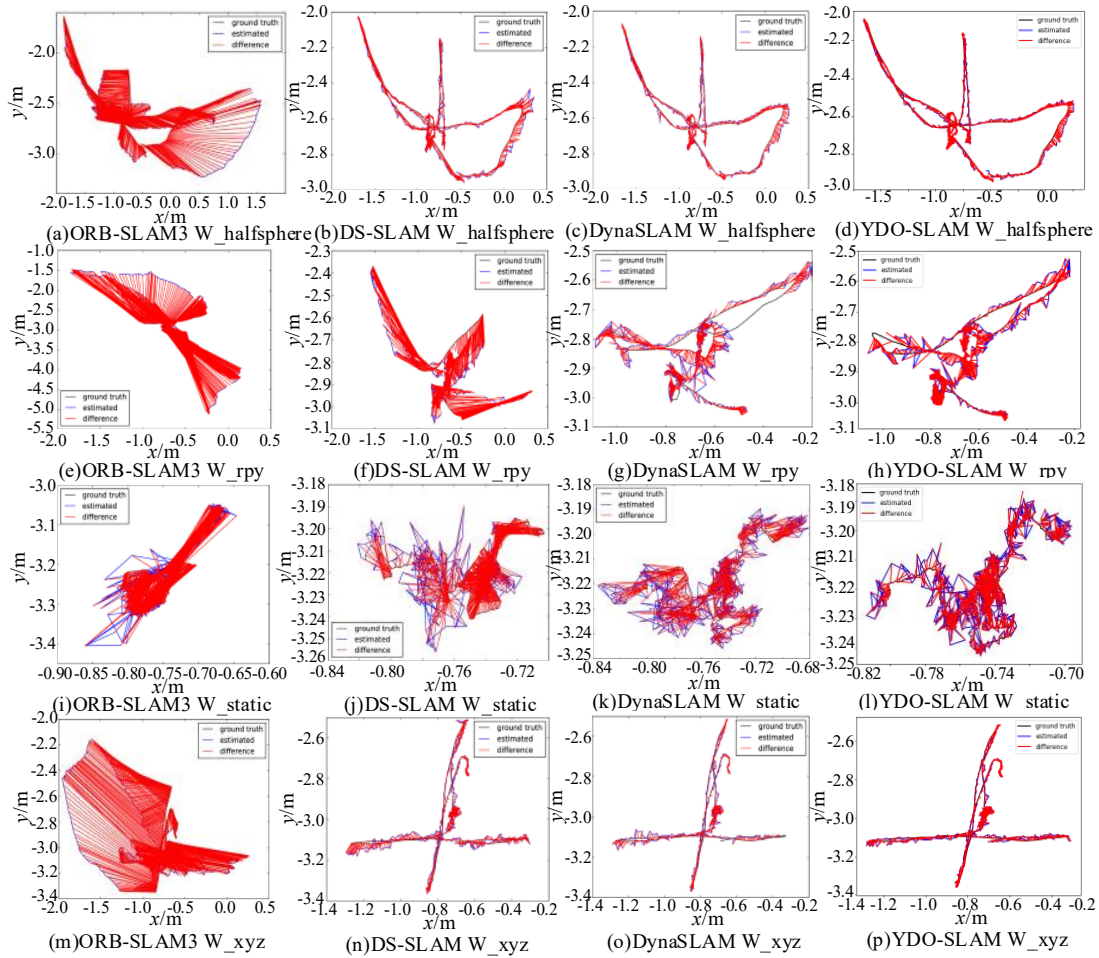


图 3-5 TUM 数据集中四种算法的轨迹图

在 W_xyz 序列中，遮挡物体为 1~2 名持续行走的人体，遮挡区域密度中等偏高；而 W_static 中的动态遮挡频率更低，遮挡密度较低，便于对比算法在不同遮挡复杂度下的表现。其中 ORB-SLAM3 为静态环境中的 SLAM 算法，DS-SLAM、DynaSLAM 和 YDO-SLAM 为动态环境中的 SLAM 算法。从图 3-5(a)、(e)、(i)、(m)可以看出，由于 ORB-SLAM3 无法识别动态场景下的运动物体，其生成的轨迹与真实轨迹误差较大；相比之下，图 3-5(d)、(h)、(l)、(p)显示出本文算法能有效地解决在运动背景中存在的相互遮挡目标边缘难以识别的问题，提高目标运动

特征判定的精度。实验结果表明，在去除运动目标后，所得到的运动目标与其它三种方法相比，具有较小的误差。图 3-5(e)~(p)为四种算法在另外 3 组动态序列的轨迹对比图，可以看出与 ORB-SLAM3，DS-SLAM，DynaSLAM 等 3 种算法相比，本文提出的算法能够更好地适应复杂场景下的动态目标去除，从而降低跟踪误差。

本实验以绝对位姿误差(APE)为评价指标，适用于评估 SLAM 系统的全局一致性。为减小随机误差的影响，每个序列执行了 5 次重复实验，并通过计算均方根误差（RMSE）、平均值（Mean）以及标准差（SD）作为评价指标。所比较的算法包括 ORB-SLAM3、DS-SLAM 以及 DynaSLAM。如表 3-1 所示的实验结果中，以粗体黑字突出显示了最佳数值。研究表明，所提出的算法在大部分动态场景中展现出优越性能，显著降低了动态对象对位姿估计的干扰，从而为 SLAM 系统的稳定定位与地图构建提供了坚实的基础。

表 3-1 4 种算法在 TUM 数据集下的绝对位姿误差对比（单位：m）

TUM 序列	ORB-SLAM3 ^[20]	DS-SLAM ^[23]	DynaSLAM ^[24]	Our
	RMSE Mean SD	RMSE Mean SD	RMSE Mean SD	RMSE Mean SD
fr3_walking_halfsphere	0.312 0.279 0.077	0.148 0.137 0.112	0.141 0.113 0.111	0.137 0.109 0.102
fr3_walking_rpy	0.947 0.822 0.499	0.258 0.316 0.125	0.253 0.248 0.127	0.247 0.212 0.121
fr3_walking_static	0.415 0.324 0.278	0.067 0.055 0.038	0.050 0.037 0.031	0.041 0.031 0.018
fr3_walking_xyz	0.744 0.579 0.457	0.077 0.059 0.051	0.061 0.056 0.045	0.057 0.049 0.033
平均值	0.605 0.501 0.328	0.138 0.142 0.082	0.126 0.114 0.079	0.121 0.100 0.069

为评估本文所提出动态物体剔除算法的性能，与 3 种主流开源动态 SLAM 算法进行了对比分析。在 TUM 数据集内，本研究选择了 4 个动态序列来进行绝对轨迹误差的对比分析，详情如表 3-2 所示。表中以加粗黑体标识的数值代表最佳性能指标，分析结果显示在多数动态序列及整体平均值方面，相较于其他算法，本方法均有一定优势。

DS-SLAM、DM-SLAM 和 Dyna-SLAM 算法均采用几何约束方法剔除动态点，但对于深度值异常的动态点，这些方法难以有效去除，导致动态物体无法被完整剔除。本文算法结合改进的 Z-score 评分与对极几何约束，通过双重特征约束剔除动态物体，并进一步进行背景修复。这种方法不仅减少了深度值异常点对

动态剔除的影响，同时保留了大量静态特征信息，从而提升了 SLAM 系统的鲁棒性。在 4 个实验序列中，本文方法的 ATE 均值优于对比算法，相较于 DS-SLAM、DM-SLAM 和 Dyna-SLAM 算法，绝对轨迹误差分别降低 86.73%、20.75%及 17.21%，这证实了本方法在动态物体剔除方面具有显著优势，成功降低了深度异常点对系统运行的负面影响。

表 3-2 4 种算法在 TUM 数据集下的绝对轨迹误差对比（单位：m）

TUM 序列	DS-SLAM	DM-SLAM ^[47]	Dyna-SLAM	Our
fr3_walking_halfsphere	0.030	0.027	0.026	0.023
fr3_walking_rpy	0.439	0.033	0.036	0.027
fr3_walking_static	0.022	0.017	0.016	0.011
fr3_walking_xyz	0.079	0.009	0.007	0.012
平均值	0.143	0.022	0.021	0.018

3.3.4 轨迹精度对比试验

如图 3-6 所示，本文选取 TUM 数据集进行验证评估，图为 DynaSLAM 和 YDO-SLAM4 种算法在 TUM 数据集下的 4 个动态序列 S_rpy、S_xyz、W_static、W_xyz 中运行的轨迹图对比。

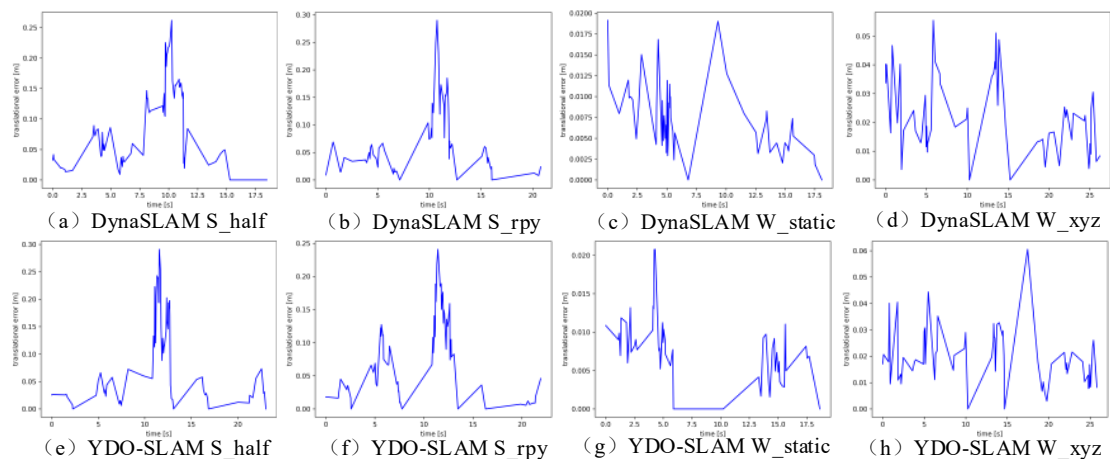


图 3-6 TUM 四种算法在数据集部分序列的定位误差对比图（单位：m）

从图 3-6(b)、(f)可以看出 YDO-SLAM 算法整体误差趋势与 DynaSLAM 很接近，在 7.5s-12.5s 本文算法定位精度更准确，从图 3-6(c)、(g)可以看出整体误差相比于 DynaSLAM，本文算法定位精度更高，跟踪效果更好。与 DynaSLAM 算

法相比,本文提出的算法能够更好地适应复杂场景下的动态目标去除,从而降低跟踪误差。

3.4 本章小结

针对传统视觉 SLAM 算法在面对动态遮挡环境时分割动态物体与剔除其特征上表现欠佳,在高速移动场景中常出现图像模糊的问题,本章采用基于注意力机制的局部自适应去模糊策略,并通过结合深度信息评估与对极几何约束的双重特征约束方法,提升 SLAM 在动态场景下的鲁棒性。实验基于 TUM 公开数据集进行。验证本章算法有效提升系统在复杂动态环境中的性能。

第 4 章 基于网络实例分割与双向映射的背景修复方法

在动态遮挡条件下，传统视觉 SLAM 算法的适用性较差。由于动态物体频繁出现在相机视野，且遮挡情况复杂多变，传统方法难以精准区分动态目标与被遮挡物体的边缘。在特征点提取过程中，动态物体的干扰容易导致特征点数量不足且质量下降，使得 SLAM 系统难以获取高精度特征点信息。这一问题不仅降低了位姿估计精度，影响机器人对自身位置和姿态的判断，还会降低定位跟踪的准确性。并且生成的地图未能精确呈现环境布局，进而限制了视觉 SLAM 技术在动态遮挡环境下的应用。传统的算法在区分动态特征点时，经常错误地将静态特征点归类为动态点而予以去除，该方法不仅未能解决特征点匮乏的问题，反而导致特征点的丢失更多，从而造成跟踪效果的不稳定性，甚至引发跟踪失败的情况。原始动态图像与特征点提取效果不佳，对后续 SLAM 任务带来诸多困难，严重制约了其在实际复杂场景中的应用与发展。原始动态图像帧与特征点提取效果图如图 4-1 所示。



图 4-1 原始动态图像帧与特征点提取效果图

本文针对移动机器人运行于大视角且动态目标和物体存在遮挡关系时易出现目标边缘分割模糊的问题，提出一种基于动态遮挡场景改进 YOLOv6s 网络的 SLAM 算法 YDO-SLAM。该算法改进包括 YOLOv6s 网络架构的语义信息融合模块和点线特征的双重特征约束两个环节。首先通过相机获取 RGB 信息和深度信息，利用改进的 YOLOv6s 网络，实现平面特征及其参数的提取。随后，通过基于多尺度的高、低层次特征融合网络来提取图像中的纹理信息。并改进多尺度神经网络获取不同尺度图像，结合多尺度图像特征更好的分割目标边缘，同时提取特征点和特征线。针对视觉 SLAM 系统在处理动态遮挡情境下，难以准确排

除动态特征点又可能误删静止特征点的问题,采用了双重特征约束的动态特征点判断方法。第一阶段为修正的 Z-score 评分方法,第二阶段为对极几何约束方法,两种判断方法相结合的方式来判断动态特征点并剔除动态点和物体。根据静态的点线特征优化相机位姿,利用双向映射模型的背景修复方法来补全被剔除区域的 RGB 图像背景,最后通过对特征点的再提取补充高质量特征点,提高 SLAM 在动态遮挡环境下的定位和建图能力。总体技术路线如图 4-2 所示。

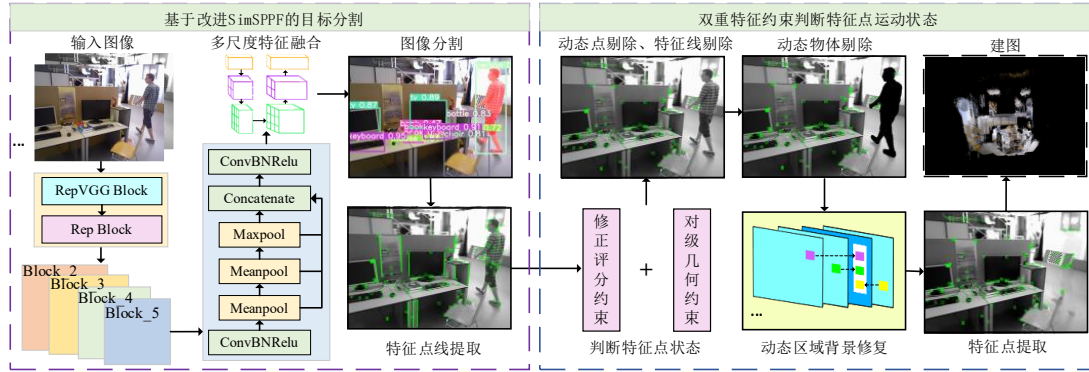


图 4-2 系统结构框图

4.1 基于改进 YOLOv6s 的物体分割

4.1.1 改进 simspfp 模块

YOLOv6s 网络结构由主干、颈部以及头部四个部分构成。与先前的版本相比, YOLOv6s 在多维度上实现了性能提升。通过引入 Mosaic 的数据,提升了运算能力,进而实现了包括任意放大、剪切、排列在内的多种功能。在 Backbone 中,设计了一个更有效的骨干网络 EfficientRep Backbone,该网络可以重参数化,所以效率更高。相较于 YOLOv5 所搭载的 CSP-Backbone,本算法网络的 Backbone 展现出更为卓越的特征表示能力,同时能高效地调配硬件资源。在 Neck 组件中,引入了 Rep-PAN,旨在构建一个更为高效的特征融合网络架构。Head 部分采用了解耦检测头结构,并对其进行了优化与简化设计。在这一框架下, YOLOv5 采用了融合分类与回归分支的策略构建其检测模块,而 YOLOv6s 则通过分离分类与回归分支的方法进行设计,并进一步引入了两层 3×3 卷积网络,旨在提升目标识别精度。YOLOv6s 以其最快的运行速度和最少的参数,在识别具有先验信息的动态目标和潜在动态目标方面表现出色,满足了 SLAM 系统对目标检测网络

精度的要求。因此，本文选择了 YOLOv6s 作为目标检测网络。

在多种传统的池化方式中，最大池化通过捕捉并保留局部区域中的最大值，从而突出全图的重要特征，特别适用于保留强信号区域的完整性。平均池化则更注重提取局部区域的全局信息，在网络的多尺度特征融合时有助于对动态目标进行分割，尤其在动态场景下能够增强网络对背景变化的鲁棒性。为使 YOLOv6s 网络在相对复杂动态场景下的目标分割更加高效，对 YOLOv6s 的 Backbone 主干网络中的 SimSPPF 模块进行改进，将 SimSPPF 模块中的池化方式替换为两次平均池化和一次最大池化，通过结合最大池化和平均池化，能够同时保留局部的强信号和全局信息，使得网络能更好地识别边缘信息并适应复杂动态遮挡场景。将 YOLOv6s 的 Neck 层中的 RepBlock 融合特征方式替换为 FPNB 多尺度卷积模块，使整个网络架构不仅融合了高低层语义特征信息，同层内部也实现了多尺度特征融合，提升动态场景中的人物和背景的分割效果，从而增强了被遮挡部分图像修复的效果，YOLOv6s 网络架构的各部分改进后的结构示意图如图 4-3 所示。

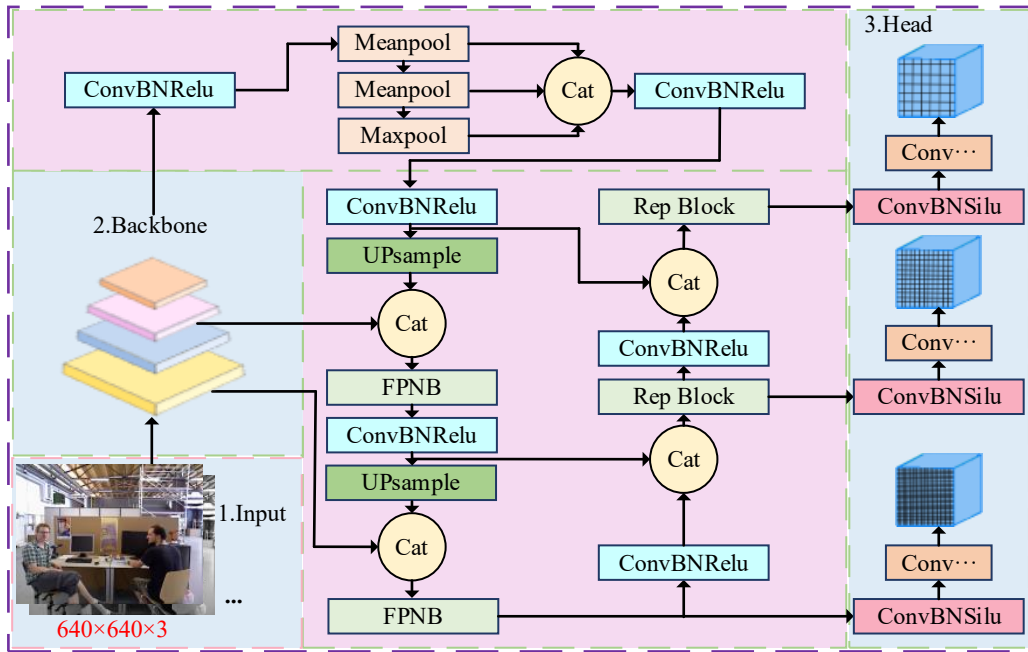


图 4-3 改进的 YOLOv6s 网络架构

4.1.2 双分支 FPNB 融合特征

针对分割时未能充分利用全局语义信息的问题，通过多尺度神经网络中采用了跨多个尺度的高低层特征融合策略和内部双分支特征融合的网络结构来处理不同尺度的图像特征。该网络通过卷积对输入图像进行上下采样，生成不同尺度

和语义通道的图像，不仅减少了传统下采样过程背景信息的损失，同时增加了网络感受野。为了将不同尺度的图像表征特征更好地融合，防止被遮挡物体和动态物体边缘模糊。改进网络对输入图像识别预处理的 **simsppf** 模块和多尺度融合语义通道信息的 **neck** 层，在不同尺度上生成出估计的清晰图像，并通过跨尺度融合多尺度图像特征，最大程度地保留图像的纹理信息，逐步排除无用的像素信息。将图像中被遮挡物体和动态物体边缘特征放大，使动态物体轮廓分割的更加精准，提高特征点提取精度。通过内部多尺度特征融合处理高层特征图像，该方法将权值在不同的尺度上进行分配，采用并行优化的网络结构能够显著缩减系统的运行时间，降低了训练参数的数目，降低了系统的不稳定性。

针对 **Neck** 层融入了 **FPNB** 下采样模块，该模块有显著的泛化能力和深度特征提取能力，作为多级特征整合框架，其通过调节最小与最大梯度，从而强化网络的学习效能与抗干扰性。**FPNB** 模块集成了两条路径，其中一条路径由一个 1×1 卷积层紧随其后连接一个 3×3 卷积层组成，另一路径则结合一个 1×1 卷积层与 3 个依次串联的 3×3 卷积层。

在 **FPNB** 模块中使用的多个 1×1 卷积和 3×3 卷积组合，其中 1×1 卷积有助于进行通道数变换，减少参数量， 3×3 卷积则用来提取更高层次的特征。通过两者的结合，能够更加精确地提取多尺度信息，并有效地压缩特征空间，提高了网络的表征能力，从而增强了网络在动态遮挡环境中的表现。对于多尺度目标和动态场景中的遮挡问题，深度表征能力较为重要，该改进可提升网络区分被遮挡物体和动态目标。**FPNB** 的结构通过多分支的方式处理不同尺度的信息，使得网络能够适应更多样的输入数据，提高了网络的鲁棒性和泛化能力。由于动态遮挡场景下的动态物体的速度、方向、外观等信息变化较大，而 **FPNB** 能够帮助网络更好地捕捉这些变化。最终通过对两条分支的特征进行融合，获得完整的图像特征表达。这种融合方式使 **FPNB** 模块能够更有效地实现多尺度特征提取，增强目标特征的表达能力，提升对目标特征的捕捉能力和更精确的位姿估计，如图 4-4 所示。

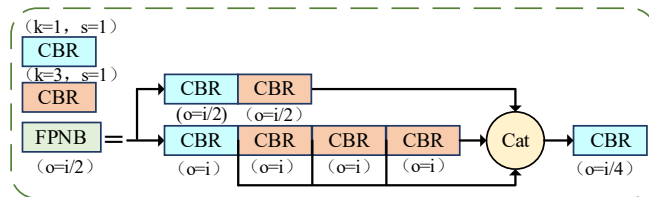


图 4-4 网络架构中 FPNB 模块图

为了在多尺度的特征融合中估计目标物体的位姿,本方法采用了不同尺度下采样的图像作为训练目标。每次迭代生成的中间图像输出形成了一个清晰边缘图像的高斯金字塔,并为每个尺度应用了欧氏损失。鉴于这种情况,提出了一种加权损失函数,如公式 4-1 所示。

$$loss = \frac{1}{N} \sum_{i=1}^n \sum_{j=1}^Q \left\| \frac{\tilde{\omega}_j}{H_i \times W_i \times C_i} (x_{ji} - \tilde{x}_{ji}) \right\|^2 \quad (3-1)$$

其中 n 表示了样本的数量, x_{ji} 表示第 j 个尺度上第 i 个训练样本, $\tilde{x}_{ji}$ 表示网络预测的第 j 个尺度上第 i 个训练样本。而 Q 则表示了网络的尺度数量, $\tilde{\omega}_j$ 表示了第 j 个尺度的权重, H_j 、 W_j 和 C_j 分别是训练样本在第 j 个尺度上的长度、宽度和通道数。其中权重计算如公式 4-2 所示。

$$\begin{cases} \omega_j = \left(\frac{1}{n} \sum_{i=1}^n \|x_i\|_1 \right)^{-1} \\ \tilde{\omega}_j = \frac{\omega_j}{\sum_{i=1}^n \omega_j} \end{cases} \quad (3-2)$$

4.2 基于改进双向映射的背景修复

在剔除动态目标后,图像中的多重特征信息随之减少,特别是在存在较多动态物体的场景中,此策略可能导致定位准确度下降,从而对闭环检测及构建静态地图造成不利影响。为解决这一问题,通过一种基于双向映射的背景修复模型修复背景。该模型充分利用了前、后两帧的静态信息,对剔除的动态目标遮挡的 RGB 图像进行复原。在对静止背景进行修复的基础上,对图像中的特征点进行再提取、匹配,从而提升 SLAM 系统在多动态物体的动态遮挡环境下的稳定性。

针对特定场景,传统 SLAM 在特征提取方面表现出的不稳定性,使得系统的定位准确度有所降低。本文在背景修复过程中首先剔除了动态物体,随后通过特征再提取,显著降低了动态物体对系统跟踪定位的干扰,进而提升了位姿估计的准确性,并有效缩减了轨迹误差。

4.2.1 背景修复原理

在视觉 SLAM 系统中，背景修复的核心目标是通过利用时间相邻帧的信息来推测并填充当前帧中被动态物体遮挡的区域。常见的背景修复方法可以分为基于时序信息的插值法、基于图像修复的补全法和基于深度映射的重构法。本文采用改进双向映射的方法，即在前后帧之间建立静态像素点的对应关系，并结合光流信息和深度约束，推测被遮挡区域的真实背景内容。双向映射的背景修复模型如图 4-5 所示。

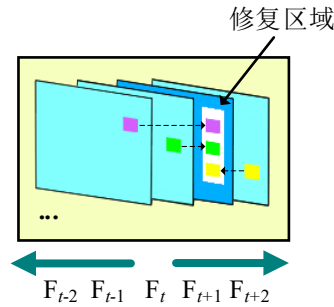


图 4-5 基于双向映射的背景修复模型

假设在时刻 t ，输入图像为 F_t ，该图像中包含了动态目标 D_t 和静态背景 B_t 。目标检测网络会识别并剔除 D_t ，但会留下遮挡区域 M_t 。用前后帧 F_{t-1} 和 F_{t+1} 中的信息，填充 M_t 以恢复完整的背景图像 B_t^* 。

双向映射的背景修复模型分为前向映射和后向映射两个部分，前向映射的基本原理是将前帧 F_{t-1} 中的静态像素投影到当前帧 F_t ，以预测被遮挡区域的内容。假设像素点 p 在 F_{t-1} 和 F_t 之间的变换关系由单应性矩阵 H 确定，则其投影变换如公式 4-3 所示。

$$p_t = H_{t-1}^t \cdot p_{t-1} \quad (3-3)$$

其中 H_{t-1}^t 可通过特征点匹配与 RANSAC 估计获得。对于被动态目标遮挡的区域 M_t ，如果 p_t 的值缺失，则从 F_{t-1} 中找到对应的 p_{t-1} 进行填充，如公式 4-4 所示。

$$B_t^*(p_t) = F_{t-1}(p_{t-1}) \quad (3-4)$$

反向映射方法则是从后续帧 F_{t+1} 中寻找合适的像素填充当前帧 F_t 的遮挡区域 M_t 。同理，反向映射的定义，如公式 4-5 所示。

$$p_t = H_{t+1}^{t-1} \cdot p_{t+1} \quad (3-5)$$

同样，若 p_t 落入遮挡区域 M_t ，则可以从 F_{t+1} 中找到对应的 p_{t+1} 进行填充，如

公式 4-6 所示。

$$B_t^*(p_t) = F_{t-1}(p_{t-1}) \quad (3-6)$$

4.2.2 约束条件与优化策略

由于动态场景下的光照、运动模糊等因素影响，单纯依靠前后帧的信息进行插值可能会引入较大的误差。为了提高背景修复的精度，本文采用光流约束和深度一致性约束来优化填充过程。

光流描述了像素在图像序列中的运动关系，可以用于估计遮挡区域的背景运动情况。利用光流场 I_t ，可以建立像素点 p_t 在相邻帧之间的运动轨迹如公式 4-7 所示。若 $p_t \in M_t$ ，则采用光流估计的方法填充，填充原理同上，如公式 4-7 所示。

$$p_{t+1} = p_t + I_t(p_t) \quad (3-7)$$

对具有深度信息的图像帧，可以进一步利用可以进一步利用深度约束来增强背景修复的准确性。假设像素点 p_t 在深度图中的深度为 $Z_t(p_t)$ ，通过三角测量法可以估计其世界坐标系中的真实位置，如公式 4-8 所示。

$$p_t = K^{-1} p_t Z_t(p_t) \quad (3-8)$$

其中 K 是相机内参。当深度变化较大时，说明该区域可能受到动态物体的遮挡，因此可以采用局部深度平滑的方法来填充缺失区域，如公式 4-9 所示。

$$Z_t^*(p_t) = \frac{1}{N} \sum_{i=1}^N Z_{t-1}(p_i) + Z_{t+1}(p_i) \quad (3-9)$$

其中 p_i 是当前帧 p_t 在前帧和后帧中对应的参考像素点，用于深度插值计算。最终，通过综合光流约束和深度一致性约束，得到完整的修复背景 B_t^* 。

4.3 实验结果与分析

4.3.1 动态遮挡物体分割算法实验

四种算法在 TUM 数据集部分序列上实例分割如图 4-6 所示，为了更好的验证本文算法的优势，采用 TUM 数据集中的动态场景序列，选择了不同的算法（Mask-Rcnn、Transform、Yolov6s 和 YDO-SLAM）进行性能对比。

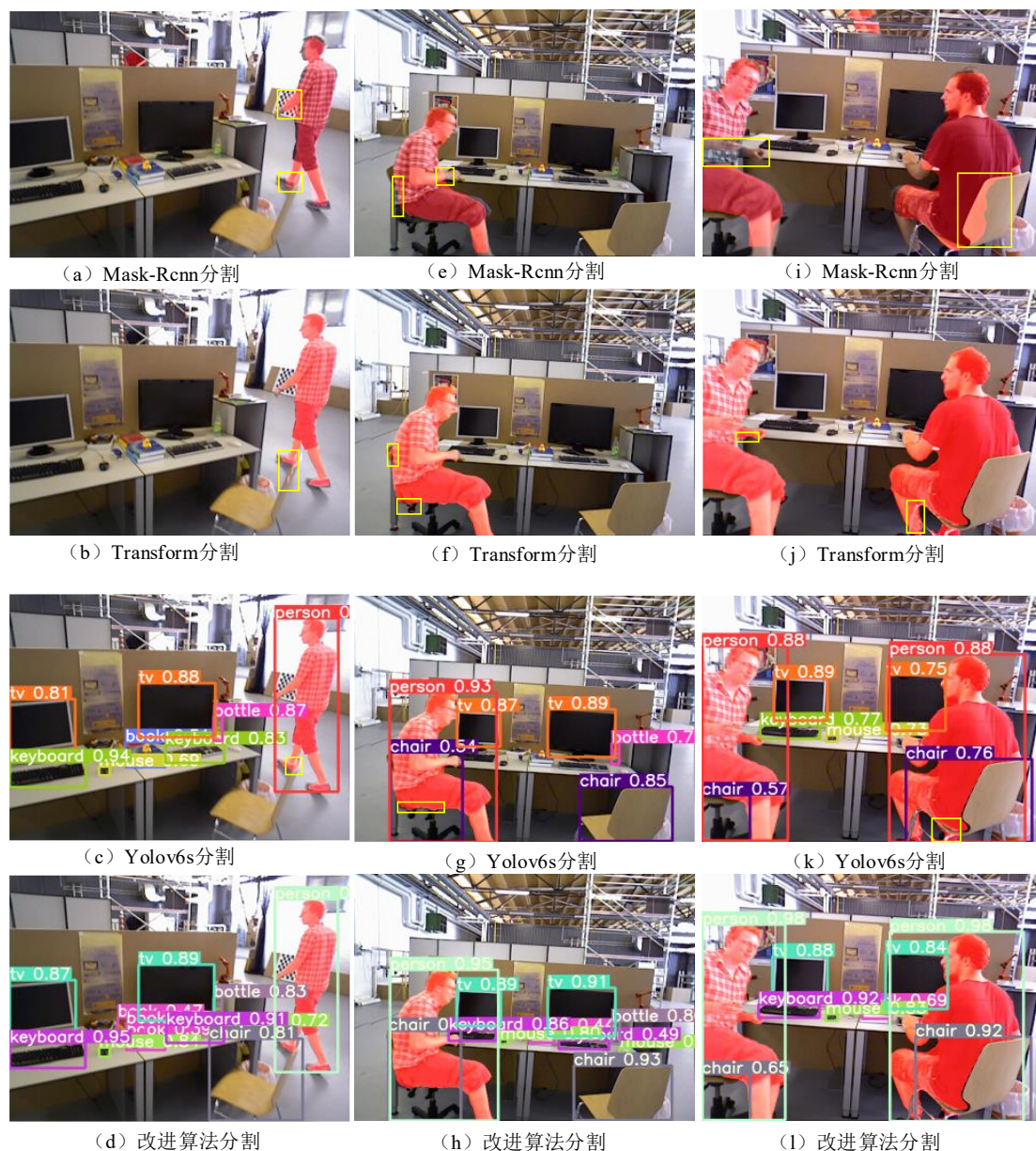


图 4-6 四种算法在 TUM 数据集部分序列上实例分割对比图

在图 4-6(a)-(l)的场景中，存在物体相互遮挡的情况。例如，在图 4-6(a)中，Mask-Rcnn 未能完全分割靠近被遮挡物体的部分；在图 4-6(i)中，左侧人物的实例分割严重缺失，右侧人物则过度分割了椅背。这表明 Mask-Rcnn 受限于卷积网络的感受野和卷积核大小，难以获取全局的语义信息，在动态遮挡场景下表现不佳，导致物体分割不完全、过度分割甚至失效。通过图 4-6(b)~(c)、(f)~(g)、(j)~(k)可知 Transform、Yolov6s 在遮挡情况下对物体实例分割效果显著提升，但由于物体边界的语义特征模糊、缺少全局特征的捕获，物体边界分割都有不同程度分割不准确的情况，而 YDO-SLAM 与前三种算法相比，改进了网络主干层中的

SimSPPF 模块并添加多尺度语义通道的融合，将物体的边界纹理信息最大保留，增加被遮挡物体边界语义性同时抑制了无用信息。从而提升了算法的实例分割效果，因此在相对复杂的动态遮挡场景下，YDO-SLAM 的分割性能效果更为突出。

为了准确地评估本文提出的算法相对优势，采用不同交并比（IoU）下的平均准确率（AP）作为实例分割效果的衡量标准。表 4-1 显示了 Mask-Rcnn、Transform、Yolov6s 和 YDO-SLAM 在 AP、AP50（IoU>50%）、AP75（IoU>75%）方面的对比结果。

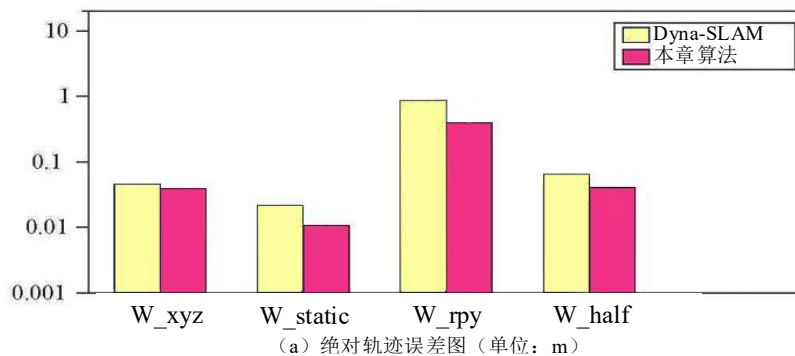
表 4-1 四种算法 AP 值对比

算法	AP	AP ⁵⁰	AP ⁷⁵
Mask-Rcnn	36.44	54.21	33.41
Transform	37.76	59.07	39.12
Yolov6s	38.27	61.86	40.85
YDO-SLAM	39.62	64.38	43.74

根据表 4-1 数据，相比于 Mask-Rcnn，YDO-SLAM 的平均精度相比于 Mask-Rcnn 提高了 16.35%，与 Transform 相比提高了 7.93%，与改进前的 Yolov6s 相比提高了 4.69%。

4.3.2 系统定位精度实验

为验证本章提出的算法在动态场景下的鲁棒性，选取了 TUM 数据集中 W_xyz、W_static、W_rpy 以及 W_half 这四个动态序列，本研究对比了所提出算法与 Dyna-SLAM 在轨迹误差上的性能差异。采用均方根误差（Root-Mean-Square Error, RMSE）和标准差（Standard Deviation, S.D.）作为指标，指标数值越小，表明系统的鲁棒性越强，轨迹误差如图 4-7 所示。



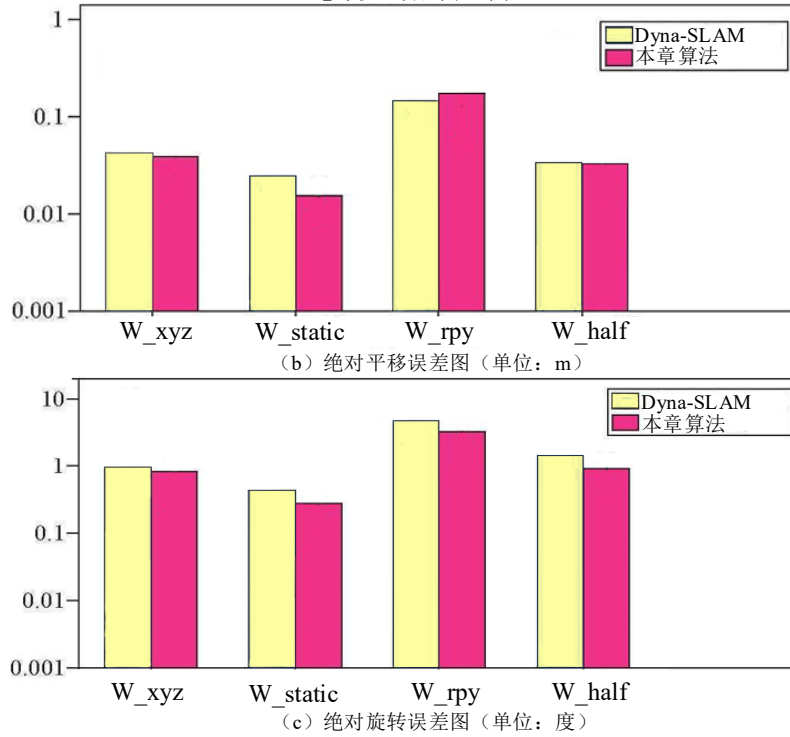


图 4-7 轨迹误差图

图 4-7 为 Dyna-SLAM 算法和本章算法在轨迹误差方面的 RMSE 对比结果, 其中误差线代表标准差 (S.D.)。从图 4-7 (a) 可以明显看出相较于 Dyna-SLAM, 本章算法的绝对轨迹误差减少了 13.68%。由图 4-7 (b) 可以得知本章算法在相对平移误差的 RMSE, 相较于 Dyna-SLAM 减少了 8.23%。本章算法在处理平移相关的动态变化时具备更出色的表现, 能够更准确地确定物体的平移轨迹, 降低误差。最后, 从图 4-7 (c) 可知, 本章算法在绝对旋转误差的 RMSE 上, 相较于 Dyna-SLAM 减少了 11.51%, 说明在处理旋转相关的动态场景时, 本章算法的轨迹误差得到了极大的优化, 能够更精准地对旋转情况进行定位和计算。无论是在旋转还是平移等综合因素影响下的动态场景中, 本章算法在整体的轨迹误差控制上, 与 Dyna-SLAM 相比, 表现出更优的鲁棒性及更高的精度。这表明本章节所述算法在动态场景应用中展现出的有效性和优势。

4.3.3 背景修复实验

背景修复与特征点提取的效果如图 4-8 所示, 图 4-8 (a)~(e) 展示了从 TUM 数据集中挑选的 5 个动态原始彩色图像帧, 这些图像帧中包含各种程度的动态遮挡现象及运动模糊; 图 4-8 (f)~(j) 对应于经过背景修正后的图像帧, 由图可

见双向映射模型在缺失区域填充静态背景信息的效果，保证了修复区域的合理性和统一性。图 4-8 (k) ~ (o) 展示了背景修复后的点、线、面特征提取帧，静态特征点分布更加均匀，能为后续 SLAM 任务提供更高质量的匹配点集。本章方法通过剔除动态物体、修复背景并提取点线面特征，有效降低了动态物体对 SLAM 系统的干扰，减少因动态目标带来的特征匹配误差。此外，背景修复后的特征点分布更为完整，提高了系统位姿估计精度，从而减少轨迹误差。本方法在动态遮挡场景下，能够在保证定位精度的同时，优化地图构建质量，为 SLAM 在复杂动态遮挡场景下的应用提供更可靠的技术支撑。

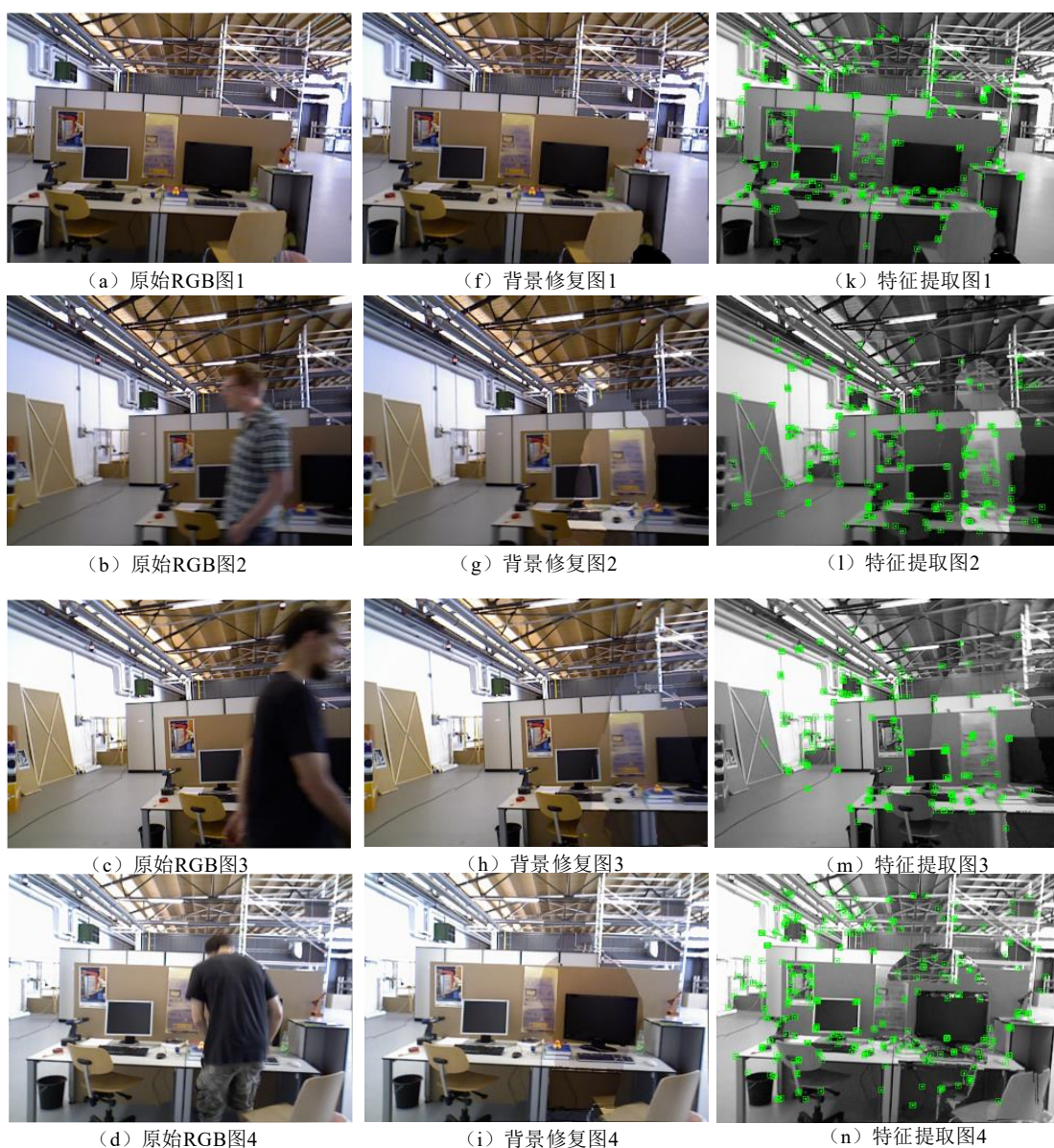




图 4-8 背景修复与特征提取图

4.4 硬件平台设计及真实场景实验

4.4.1 机器人硬件平台

为验证算法对动态遮挡物体语义分割以及动态遮挡物体剔除区域背景修复可行性测试实验，本节基于 Husky 机器人平台构建巡检机器人，图像数据通过视觉传感器来采集。实验中需要处理动态遮挡物体以及网络分割等相关数据，故本实验采用分布式计算架构进行测试。此架构有助于减少机器人端的计算负担，提高系统的实时性和整体性能，Husky 机器人如图 4-9 所示。

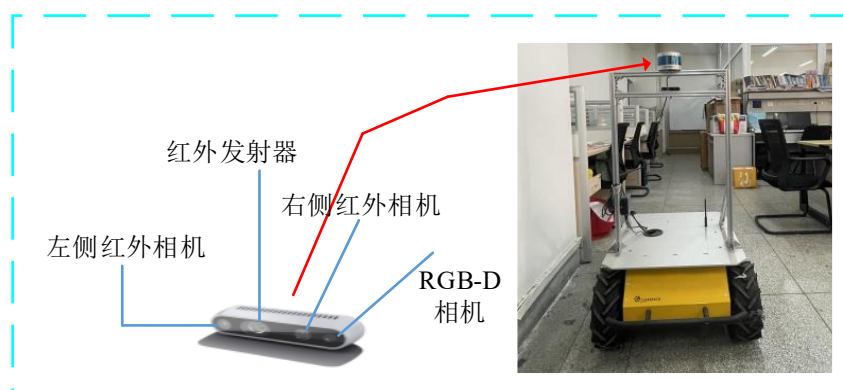


图 4-9 Husky 电力巡检机器人平台

4.4.2 真实场景实验

利用 Husky 机器人进行动态遮挡场景下的定位精度验证，由于实验室的设备条件有限且满足真实环境中出现相互遮挡动态物体干扰因素，故本节实验验证以实验室环境作为模拟场景来进行。真实实验室场景环境如图 4-10 (a) 所示，大小为 $10\text{m} \times 6\text{m}$ ，真实场景的平面布局图如图 4-10 (b) 所示，图中黄色和蓝色长方形平面为实验室的工作台，紫色线条为机器人的实际运行路线。

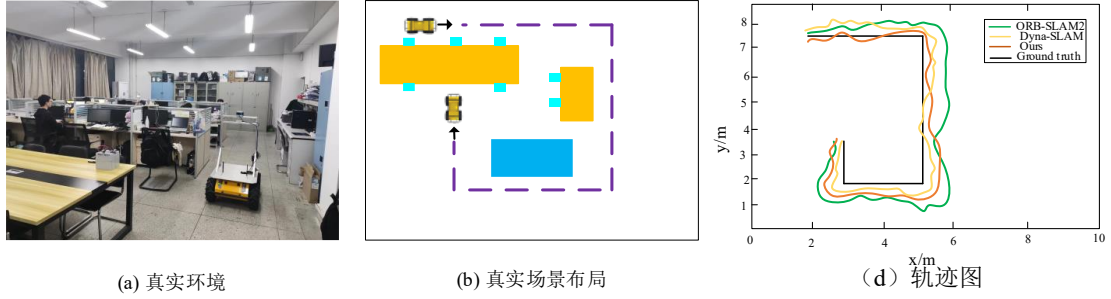


图 4-10 真实场景环境布局图

图 4-10 (d) 为本章算法与 ORB-SLAM2、DynaSLAM 算法的真实轨迹的对比。经测试，YDO-SLAM 的绝对轨迹误差为 0.127m，DynaSLAM 的绝对轨迹误差为 0.169m。由于 ORB-SLAM2 算法无法在具有动态物体的场景下运行，其轨迹误差较大。DynaSLAM 算法加入了 MASK-RCNN 算法识别动态物体，其轨迹误差相对于 ORB-SLAM2 算法有所降低，但比本章算法轨迹误差大。本章算法通过结合改进 YOLOv6s 网络能有效的识别动态物体和背景的边缘，从而更好的分割动态遮挡物体。

通过结合双重特征约束策略剔除动态特征，减少静态特征点的误剔除。采用双向映射的背景修复模型修复背景并补充特征点，防止特征点过少导致 SLAM 跟踪丢失。改进方法显著提升了特征匹配精度，减小了轨迹误差。表明本章算法在真实动态遮挡场景下鲁棒性更强，为视觉 SLAM 在动态遮挡环境下稳定、高效地定位跟踪提供了有力保障。机器人运行中某一帧图像如图 4-11 所示。



图 4-11 机器人运行时某一帧图像

选取机器人运行过程中有动态遮挡物体的一帧图像进行实验分析。其中图 4-11 (a) 为原始 RGB 图，图 4-11 (b) 为背景修复图。图中可以看出，本章算法在动态遮挡场景下可准确识别动态物体并剔除，并且使用前后帧的静态信息修复当前修复区域。特征再提取图如 4-11 (c) 所示，红色框辅助标记被修复的动态物体区域，可见背景修复后补充了高质量的特征点，为 SLAM 在动态遮挡场景

下的定位提供了更多的信息。

4.5 本章小结

本章针对动态遮挡场景下视觉 SLAM 算法存在动态遮挡物体干扰及背景信息缺失问题,提出了基于改进 YOLOv6s 网络的 SLAM 算法。通过结合双重特征约束的动态特征剔除策略,精准识别异常深度特征点。采用基于前后帧静态信息的双向映射方法,建立剔除区域与前后帧的对应关系,恢复被动态物体遮挡的 RGB 背景,修复后的背景补充了高质量特征点,为位姿估计与地图构建提供完整信息,提升了系统在动态遮挡场景下的鲁棒性。

第 5 章 全文总结展望

5.1 全文总结

本文围绕动态遮挡条件下的视觉 SLAM 问题，引入基于改进 YOLOv6s 网络的 SLAM 算法。第一章论述了 SLAM 技术的理论基础，分析了在动态遮挡环境下运行的困难，阐述了国内外前沿的最新学术进展与未来发展方向。第二章介绍视觉 SLAM 及 YOLOv6s 目标检测网络的基本理论框架，对 SLAM 系统中的视觉里程计、后端优化、闭环检测以及地图构建的原理进行叙述，详细解析了 YOLOv6s 网络的体系结构及其目标检测的工作流程。第三章和第四章针对相互遮挡物体边缘模糊导致分割效果不佳等问题，和筛选动态特征时误剔除静态特征的问题。提出结合了双重特征约束策略的基于改进 YOLOv6s 网络的 SLAM 算法，并在公开数据集 TUM 和真实场景上进行实验，验证算法的可行性。

主要研究内容与创新点概括如下：

(1) 针对动态遮挡环境下物体相互遮挡造成边界模糊的问题，采用基于注意力引导的局部自适应去模糊方法，考虑不同区域的模糊情况构建模糊权重图，根据模糊权重图来调控不同区域的去模糊程度，以此实现对特征提取及匹配过程的优化。提高 ORB 特征提取的稳定性，提升 SLAM 特征提取的稳定性和定位跟踪精度。

(2) 针对物体遮挡条件下边缘分割效果差的问题，改进了 YOLOv6s 的动态目标检测方法。该算法在 Backbone 层通过结合最大池化和平均池化，能够同时保留局部的强信号和全局信息，使得网络能更好地识别边缘信息，适应复杂动态遮挡场景。在 Neck 层，采用 FPNB 模块进行特征融合，以增强多尺度目标检测能力。

(3) 针对移除动态物体时静态特征误剔除和静态背景缺失的问题，提出基于双重特征约束的动态特征剔除方法。采用改进的 Z-score 评分方法结合四分位数调整阈值，更精准地识别异常深度值，提高动态特征点剔除的准确性。采用双向映射的背景修复模型，通过前帧与后帧的静态特征填充剔除动态物体区域，旨在保留的更多特征用于系统定位与地图构建。

5.2 展望

尽管本文所提出的方法在动态遮挡场景下取得了较好的实验效果,但仍存在一些不足和未来需要进一步探索的问题:

(1) 复杂动态环境下的适应性问题。在复杂动态遮挡环境中(如人群密集场景),动态目标的密度较高,可能导致误剔除静态特征点,影响系统的稳定性。未来可以考虑结合光流分析、时间序列建模等方法,进一步提高算法的适用性。

(2) 计算复杂度与实时性问题。改进后的 YOLOv6s 网络在提高检测精度的同时,也增加了一定的计算复杂度。未来可以探索轻量级网络结构结合知识蒸馏技术,提高 SLAM 系统的实时性能。

(3) 背景修复方法的鲁棒性。目前的双向映射背景修复方法依赖于前后帧的静态信息,在场景剧烈变化或光照条件变化较大的情况下,可能出现修复误差较大。未来可以引入生成对抗网络(GAN)以提升修复质量。

参考文献

- [1] 陈孟元,郭行荣,钱润邦,程浩.基于区域映射深度优化的动态多特征 RGBD-SLAM 算法[J].中国惯性技术学报,2024,32(01):42-51.
- [2] 谢郭蓉,曲毅,蒋镭圻.抗遮挡目标跟踪的模型学习综述[J].计算机工程与应用,2022,58(02):43-56.
- [3] 陈孟元,陈何宝,刘金辉,韩朋朋,方鑫,郭俊阳.动态场景下基于双重特征约束的视觉惯性 SLAM 算法[J].中国惯性技术学报,2023,31(04):381-389.
- [4] 陈孟元,韩朋朋,刘金辉,张玉坤,江浩玮,丁陵梅.动态遮挡场景下基于改进 Transformer 实例分割的 VSLAM 算法[J].电子学报,2023,51(07):1812-1825.
- [5] 严瀚宇,孙博,马天力,高嵩.动态场景下融合深度信息的实时语义 SLAM 方法[J].传感器与微系统,2025,44(02):139-142.
- [6] 张鹏飞,李宏伟,赵亚帅,张彭昱,冯彬彬.基于改进 SLAM 框架的动态场景三维语义地图构建方法研究[J].计算机应用与软件,2024,41(11):234-240.
- [7] 隆良梁,魏小源.动态场景感知下的移动机器人视觉定位与建图[J].机床与液压,2024,52(21):79-86.
- [8] 陈孟元,丁陵梅,张玉坤.基于改进关键帧选取策略的快速 PL-SLAM 算法[J].电子学报,2022,50(03):608-618.
- [9] 白克强,朱亚兰,杨秀清,向勇,邓子犇,姜官武.融合动态目标跟踪的视觉 SLAM 算法[J].信息与控制,2024,53(05):574-584.
- [10] 李金,张荣芬,罗朝阳,刘宇红.复杂环境下的视觉同时定位与建图[J].制造业自动化,2024,46(10):142-150.
- [11] 王伟,汤琴琴,汪先伟.基于特征点改进的视觉 SLAM 定位研究[J].计算机测量与控制,2024,32(02):219-226.
- [12] 黄晓涛,冯桑,张宁,何鸿杰,杨润彬.一种基于 ORB-SLAM2 的动态视觉 SLAM 方法[J].机械工程与自动化,2025,54(01):20-24.
- [13] 王险峰,刘赵杰,赵通,杨浩伟,汪柏彤.动态环境下基于深度相机与 IMU 的动态目标剔除[J].计算机与数字工程,2025,53(01):90-95.
- [14] Newcombe R A, Lovegrove S J, Davison A J. DTAM: Dense tracking and mapping in real - time[C]//2011 international conference on computer vision (ICCV 2011).

- Barcelona, Spain: IEEE, 2011: 2320 - 2327.
- [15]Engel J, Schops T, Cremers D. LSD - SLAM: Large scale direct monocular SLAM[C]//Proceedings of the European Conference on Computer Vision (ECCV 2014). Berlin: Springer, IEEE, 2014: 834 - 849.
- [16]Engel J, Koltun V, Cremers D. Direct Sparse Odometry[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(03): 611 - 625.
- [17]Davison A J, Reid I D, Molton N D, et al. MonoSLAM: Real - time single camera SLAM[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(06): 1052 - 1067.
- [18]Klein G, Murray D. Parallel Tracking and Mapping for small AR workspaces[C]//2007 6th IEEE and ACM international symposium on mixed and augmented reality (ISMAR 2007). Nara, Japan: IEEE, 2007: 225 - 234.
- [19]T. Qin, P. Li, S. Shen, VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator[J]. IEEE Transactions on Robotics, 2018, 34(04): 1004-1020.
- [20]Feng D ,Yin Z ,Wang X , et al.YLS-SLAM: a real-time dynamic visual SLAM based on semantic segmentation[J].The Industrial Robot,2025,52(01):106-115.
- [21]Mur - Artal R, Montiel J M M, Tardos J D. ORB - SLAM: a versatile and accurate monocular SLAM system[J]. IEEE Transactions on Robotics, 2015, 31(05): 1147 - 1163.
- [22]Mur - Artal R, Tardos J D. Orb - slam2: An open - source slam system for monocular, stereo, and rgb - d cameras[J]. IEEE Transactions on Tobotics, 2017, 33(05): 1255 - 1262.
- [23]Yu C,Liu Z,Liu X, et al.DS-SLAM:a semantic visual SLAM towards dynamic environments[C]//2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).Spain:IEEE Press,2018:1168-1174.
- [24]Bescos B, Fácil J M, Civera J, et al. DynaSLAM: tracking, mapping and inpainting in dynamic scenes[J]. IEEE Robotics and Automation Letters, 2018, 3(04): 4076 - 4083.
- [25]赵诗伦,韩东晓.动态场景下基于语义信息的视觉 SLAM 研究[J].计量与测试技术,2024,51(10):57-62.
- [26]Badrinarayanan V, Kendall A, Cipolla R. SegNet: A Deep Convolutional Encoder-

Decoder Architecture for Image Segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481–2495.

[27] Hornung A, Wurm K M, Bennewitz M, et al. OctoMap: An efficient probabilistic 3D mapping framework based on octrees[J]. Autonomous Robots, 2013, 34(02): 189-206.

[28]Xiao L,Wang J,Qiu X,et al.Dynamic-SLAM:Semantic monocular visual localization and mapping based on deep learning in dynamic environment[J].Robotics & Autonomous Systems,2019,117(12):1-16.

[29]Cheng J,Wang Z,Zhou H,et al.DM-SLAM:a feature-based SLAM system for rigid dynamic scenes[J].International Journal of Geo-Information,2020,9(04):202.

[30]Ji T,Wang C,Xie L.Towards Real-time semantic RGB-D SLAM in dynamic environments[C]//2021 IEEE International Conference on Robotics and Automation (ICRA).China: IEEE Press, 2021:11175-11181.

[31]Huang T,Zhang R,et al.PLD-SLAM:a new RGB-D SLAM method with point and line features for indoor dynamic scene[J].ISPRS International Journal of Geo-Information,2021,10(03):163.

[32]张磊,徐孝彬,曹晨飞等.基于动态特征剔除的图像与点云融合的机器人位姿估计方法[J].中国激光,2022,49(06):58-69.

[33]Wang Y, Xu K, Tian Y, et al. DRG-SLAM: A Semantic RGB-D SLAM Using Geometric Features for Indoor Dynamic Scene[C]//2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Japan: IEEE Press, 2022: 1352-1359.

[34]刘泽峰,冉腾,肖文东,袁亮.基于残差的场景流动态目标跟踪视觉 SLAM 算法[J].电子测量技术, 2023, 46(01): 1-9.

[35]喻擎苍,董根阳,方才威,孙树森.基于目标检测和多视图几何的动态 SLAM 算法[J].现代电子技术,2025,48(01):135-143.

[36]朱旭,徐熙平,张宁,孟志,谭红蔚,张路晴,胡一鸣,张嘉旭.基于改进的 ELSED 线检测算法的双目视觉 SLAM[J].激光与光电子学进展, 2023, 60(01): 1-17.

- [37]李嘉辉,范馨月,张干,张阔.基于背景修复的动态 SLAM[J].数据采集与处理,2024,39(05):1204-1213.
- [38]王爽,刘云平,张柄棋,陆旭春,徐梁.基于全景分割与多视图几何的动态 SLAM 方法[J].电子测量技术,2024,47(24):149-159.
- [39]靳德利,仇新,张旭阳,朱文辉,左依林.动态场景下基于 DeepLabv3+的语义视觉 SLAM[J].通信与信息技术,2024,(06):29-34.
- [40]仇新,朱文辉,张旭阳,靳德利,左依林.基于 YOLOv5 的动态场景视觉 SLAM 研究[J].通信与信息技术,2024,(05):35-40.
- [41]Liu G, Zeng W, Feng B, et al. DMS-SLAM: A general visual SLAM system for dynamic scenes with multiplesensors[J]. Sensors, 2019, 19(17): 3714.
- [42]张恒,张晓强,姜官武,张智鑫,何阳,王旭亮.基于实例分割与光流特征聚类的动态 SLAM 算法[J].激光与光电子学进展, 2023, 60(01): 1-11.
- [43]蒋鹏程,邱俊武,陈衡锋,章旭国,陈佳鑫,田壮.基于 YOLOv4 的室内动态场景下 ORB-SLAM3 优化方法[J].时空信息学报,2024,31(05):596-604.
- [44]张喜,张鹏,刘鹏.基于垂直约束的紧耦合激光惯性 SLAM 方法[J].国外电子测量技术,2024,43(09):1-7.
- [45]Chum O, Matas J, Kittler J. Locally optimized RANSAC[C]//Pattern Recognition: 25th DAGM Symposium, Magdeburg, Germany, September 10-12, 2003. Springer Berlin Heidelberg, 2003: 236-243.
- [46]Liu J, Li X, Liu Y, et al. RGB-D inertial odometry for aresource-restricted robot in dynamic environments[J]. IEEE Robotics and Automation Letters, 2022, 7(04): 9573-9580.
- [47]Cheng, J.; Wang, Z.; Zhou, H.; Li, L. DM-SLAM: A Feature-Based SLAM System for Rigid Dynamic Scenes[J]. ISPRS International Journal of Geo-Information, 2020, 9(04), 202.

攻读学位期间发表的学术论文的科研成果

一、发表学术论文

- [1] 陈孟元, 许瑞珩, 杨苏朋, 丁帅. 动态遮挡环境下基于改进 YOLOv6s 网络的 SLAM 算法[J]. 中国惯性技术学报, 2025, 33(03) (已录用)。
- [2] 陈孟元, 杨苏朋, 许瑞珩, 等. 改进 OneFormer 分割网络的动态物体实时跟踪 SLAM 算法[J]. 中国惯性技术学报, 2025, 33(03): 257-266.

二、参与的科研项目

- [1] 安徽省重点研究与开发计划项目 (参与)
- [2] 安徽省高校协同创新项目 (参与)
- [3] 安徽省高校杰出青年科研项目 (参与)

三、科技竞赛

- [1] 第二届马鞍山专利检索大赛二等奖

致谢

行文至此，意味着我的论文即将完成。回首这段学习与探索的时光，心中感慨万千，也充满感恩。

首先，感谢我的导师陈孟元。从研究方向的确定到论文框架的搭建，您始终以严谨的治学态度给予我悉心指导。在遇到困难时，您的耐心点拨让我豁然开朗；在思路停滞时，您的鼓励让我重拾信心。您对学术的专注与热忱，让我深刻体会到科研的意义，这份教诲将令我终身受益。

感谢课题组的同学们。在实验设计和数据分析中的宝贵建议，让我少走了许多弯路；日常的讨论与交流，也让我感受到思维碰撞的火花。与你们共事时光，是这段旅程中不可或缺的温暖记忆。

感谢我的家人。无论我专注于学业还是陷入低谷，你们始终默默支持我的选择，用包容与理解为我筑起坚实的后盾。你们的信任与陪伴，是我坚持的动力。

纸短情长，唯愿将这份感激化作前行的动力。未来定当继续努力，不负所期。