分类号:TP242.6　　学校代码:10363
密　级:公开　　学　　号:2210310105

安徽工程大学
Anhui Polytechnic University

硕士学位论文

题目　基于深度学习的电力巡检机器人视觉SLAM算法研究

论文作者　郭行荣
指导教师　陈孟元 教授
学科（专业）　电气工程
研究方向　电力系统及其自动化

论文提交日期：2024 年 5 月 31 日

分类号：TP242.6　　单位代码：10363
密　级：公 开　　学　　号：2210310105

题　目　基于深度学习的电力巡检机器人视觉SLAM算法研究

题　目　Research on Deep Learning based Visual SLAM Algorithm for Electric Power Inspection Robots

学生姓名： 郭行荣

导　　师： 陈孟元（教授）

专业学位类别： 电气工程

研究方向（领域）： 电力系统及其自动化

论文答辩日期：2024 年 05 月 15 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：郭行荣

日期：2024 年 5 月 30 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：郭行荣　　指导教师签名：

日期：2024 年 5 月 30 日　　日期：2024 年 5 月 30 日

基于深度学习的电力巡检机器人视觉 SLAM 算法研究

摘　要

电力巡检机器人巡检时需按照固定路线进行巡航，其关键技术在于同时定位与地图构建（Simultaneous Localization and Mapping，SLAM），其巡检过程可以描述成电力巡检机器人通过自身搭载的传感器在没有任何先验信息输入的情况下，在运动中建立环境的模型，同时估计自己的位姿，并构建完整静态地图。根据电力巡检机器人使用的传感器不同，分为激光和视觉两大类，而早期的 SLAM 技术大多使用激光雷达来实现，但由于其高昂价格的局限性，很难大规模应用于各种工业场景下。随着技术的发展，视觉传感器逐步进入大众视野，由于视觉传感器具有使用场景广、能够获取丰富的信息等优点，使得视觉传感器逐渐成为 SLAM 研究的主流。然而视觉 SLAM 算法在遭遇动态物体时容易出现定位失败，因此如何减小动态物体对视觉 SLAM 定位的影响是本文待解决问题。

针对传统视觉 SLAM 算法在动态场景下难以对动态物体进行准确识别以及剔除动态物体后无法快速高效的修复等问题，提出基于语义分割和背景修复的动态视觉 SLAM 算法。该算法包含潜在动态物体分割和背景修复两个环节。潜在动态物体分割环节通过本文所提的语义分割网络对图像信息中存在的物体进行语义分割，并提取图像特征点。在背景修复环节，通过物体运动判断获取动态点并剔除动态物体，利用最优近邻像素匹配对已经剔除动态物体区域进行相似搜索、

像素匹配，并进行特征点的提取，为 SLAM 系统的定位提供丰富信息。

针对 RGBD-SLAM 在特征提取环节，深度图易出现深度不连续、单一特征点提取特征信息不足以及突然出现动态物体造成轨迹误差增大的问题，提出基于区域映射深度优化的动态多特征 RGBD-SLAM 算法。该算法包含深度图优化和动态物体剔除两个环节。深度图优化环节通过轻量级的 AFFormer 网络提取语义信息，并将提取的语义信息映射到其对应的深度图上进行近邻修复，从而完成深度图的优化。动态物体剔除环节包含点特征剔除动态物体线程和点线面特征提取线程。首先利用点特征信息，联合优化图像序列与多视图几何完成动态物体的剔除，并利用双向映射背景修复模型补全剔除区域的静态信息，然后进行点线面特征的提取，为 SLAM 系统稳定高效运行提供了良好的保障。

为验证本文所提算法有效性，在公开 TUM 数据集分别测试基于语义分割和背景修复的动态视觉 SLAM 算法和基于区域映射深度优化的动态多特征 RGBD-SLAM 算法，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：同时定位与地图构建，深度学习，深度映射，运动判断，背景修复

RESEARCH ON DEEP LEARNING BASED VISUAL SLAM ALGORITHM FOR ELECTRIC POWER INSPECTION ROBOT

ABSTRACT

The power inspection robot needs to cruise along a fixed route during inspection, and its key technology lies in Simultaneous Localization and Mapping(SLAM). The inspection process can be described as the power inspection robot establishes an environment model in motion, estimates its position and posture, and builds a complete static map through its own sensors without any prior information input. According to the different sensors used by power inspection robots, divided into laser and visual two categories, and the early SLAM technology mostly uses LIDAR to achieve, but due to its high price limitations, it is difficult to large-scale application in various industrial scenarios. With the development of technology, vision sensors have gradually entered the public's field of vision. Due to the advantages of wide application scenarios and rich information acquisition, vision sensors have gradually become the mainstream of SLAM research. However, visual SLAM algorithm is prone to localization failure when encountering dynamic objects, so how to reduce the influence of dynamic objects on visual SLAM localization is the problem to be solved in this paper.

Aiming at the problems that traditional Visual SLAM algorithm is difficult to accurately identify dynamic objects in dynamic scenes and can't be quickly and efficiently repaired after removing dynamic objects, improved dynamic Visual SLAM algorithm based on semantic segmentation and background repair. The algorithm consists of two parts: potential dynamic object segmentation and background restoration. The potential dynamic object segmentation segment uses the semantic segmentation network proposed in this paper to segment the existing objects in the image information and extract the image feature points. In the background repair process, dynamic points are obtained by judging the movement of objects and dynamic objects are deleted. The optimal neighbor pixel matching is used to carry out similar search and pixel matching for the deleted dynamic object region, and the feature points are extracted, which provides rich information for the localization of SLAM system.

Aiming at the problems that RGBD-SLAM in feature extraction, depth of the depth map is easy to appear discontinuous and single feature points extraction feature information is insufficient and sudden dynamic object trajectory error increases the problem, based on regional mapping depth optimal dynamic characteristics more RGBD-SLAM algorithm. The algorithm includes depth map optimization and dynamic object elimination. In depth map optimization, semantic information is extracted through

lightweight AFFormer network, the extracted semantic information is mapped to its corresponding depth map for nearest neighbor repair, so as to complete the optimization of depth map. The process of dynamic object removal includes the thread of point feature removal and the thread of point, line and surface feature extraction. Firstly, the point feature information is used to jointly optimize the image sequence and multi-view geometry to complete the elimination of dynamic objects, the bidirectional mapping background repair model is used to complete the static information of the eliminated region, and then the point, line and surface features are extracted, which provides the stable and efficient operation of the SLAM system.

In order to verify the effectiveness of the proposed algorithm, the dynamic Visual SLAM algorithm based on semantic segmentation and background restoration and the dynamic multi-feature RGBD-SLAM algorithm based on region mapping depth optimization were tested respectively on the open TUM data set. Finally, a hardware experiment platform is built to verify the feasibility of the proposed algorithm in real scenarios.

Guo Hangrong (Electrical engineering)

Supervised by Professor Chen Mengyuang

KEY WORDS: Simultaneous Localization and Mapping, Deep learning, Depth Mapping, Motion Judgment, Background Repair

目 录

第 1 章 绪论

1.1 本文研究背景及意义

电，作为现代生活的重要组成部分，极大改变了人类的生活方式，更在一定程度上塑造了我们的世界。电力的广泛应用推动了社会发展，电力系统稳定安全运行是工业稳定生产和居民安全生活的重要保障，配电房作为工业稳定生产和居民安全生活用电的核心区域，其稳定安全运行对工业稳定生产和人们安全生活有着极其重要的社会作用。为保证配电房正常使用，对其进行定期安全巡视检查是必不可少的环节。但传统人工巡检存在智能自动化程度低、成本代价高、效率低、缺乏数据的收集和交流的问题，当从一个配电房巡检到另一个配电房时，由于距离较远，人工巡检花费时间长效率低[1][2]。

近些年来随着机器人技术的快速发展，语义识别、目标检测、三维建图、自动避障以及数据融合等新型技术的引入，使得机器人拥有了替代人类高效准确工作的能力。机器人执行配电房的巡检任务能够很好地解决传统人工巡检方式中遇到的问题，但电力巡检机器人易受复杂环境的影响，当电力巡检机器人从一个配电房巡检到另一个配电房的过程中遭遇动态物体时容易出现电力巡检机器人定位失败，从而导致巡检路线紊乱无法完成巡检任务[3]。

电力巡检机器人巡检时需按照固定路线进行巡航，其关键技术是 SLAM，巡检过程可以描述成电力巡检机器人通过搭载的传感器，在没有任何经验信息获得的情况下建立地图模型，同时估计自身的位姿，并构建完整静态地图[4][5]。根据搭载传感器类型的不同，主要分为激光和视觉两大类，早期的 SLAM 技术主要采用激光雷达来实现，但由于其高昂价格的局限性，很难大规模应用于各种工业场景。随着技术的发展，视觉传感器逐步进入大众视野，由于视觉传感器具有使用场景广、价格便宜、能够获取丰富的信息等优点，使得使用视觉传感器的 SLAM 技术逐渐成为研究的主流，其中利用 RGB-D 相机作为传感器的 RGBD-SLAM 技术越来越受到研究人员的关注[6][7]。

为提高电力巡检机器人对巡检场景的理解能力，引入深度学习技术。基于深度学习的语义分割网络不仅实现了对物体的检测，还对每类物体标注了语义信息，

因此结合语义分割网络的视觉 SLAM 算法是目前视觉 SLAM 方向的研究热点之一[8][9][10]。

本文将针对视觉 SLAM 算法在遭遇动态物体时容易出现电力巡检机器人定位失败，从而导致巡检路线紊乱无法完成巡检任务的技术难题展开研究，设计适合于巡检过程中可以完成规定巡检路线任务视觉 SLAM 和 RGBD-SLAM 系统，这对解决人工巡检问题有着极大的作用。

1.2 国内外研究现状

1.2.1 电力巡检机器人研究现状

电力巡检机器人按照不同产品类型主要分为轮式电力巡检机器人和轨道式电力巡检机器人。轮式电力巡检机器人示意图如图 1-1（a）所示，轮式电力巡检机器人主要采用两轮驱动+万向轮的底盘，它能做到定位与建图、智能感知及自动避障等功能，可在灵活自如的规避各种障碍物[11][12][13]。轨道式电力巡检机器人示意图如图 1-1（b）所示，它是由巡检机器人、轨道路线、通讯网络和监视平台组成。它具有自动感知和续航时间长等能力，但是它灵活度极低，只能按布置的轨道路线进行巡检。随着各种深度学习算法应用落地，结合深度学习算法的智能电力巡检机器人被开发，它具有智能感知，自动避障，自动精准定位，能够实时信息共享，自动报警等功能。目前智能电力巡检机器人被大量应用于社会的各行各业[14][15]。

（a）轮式巡检机器人　　（b）轨道式巡检机器人

图 1-1 巡检机器人示意图

为完成从一个配电房到另一个配电房的巡检任务，结合目前所拥有的实验设

备及器具，综合轮式电力巡检机器人和轨道式电力巡检机器人的优缺点，选定图1-1（a）所示的轮式巡检机器人作为本论文中的电力巡检机器人[16]。

2006 年加拿大提出首个用于配电站巡检轮式巡检机器人，通过红外热像仪和可见光图像采集系统对设备进行巡检，同时通过电脑进行实时监控，但是由于同年 SLAM 相关技术发展不完备，只能通过机械遥控方式对机器人进行控制，大大降低了机器人的自主性。我国开始于上世纪末对变电站巡检机器人进行技术研究，并于 2004 年研制出落地产品。随着十几年的发展我国的巡检机器人产品越来越多，比如国网智能、江苏亿嘉和、深圳朗驰等[17]。各类巡检机器人产品如图 1-2 所示。

（a）国网智能巡检机器人 （b）江苏亿嘉和巡检机器人（b）深圳朗驰巡检机器人

图 1-2 各类巡检机器人产品

1.2.2 视觉 SLAM 研究现状

视觉 SLAM 有搭载单目、双目和 RGB-D 相机三种主流视觉传感器的算法。其中基于 RGB-D 相机的 SLAM 算法通过红外结构光或 TOF 原理直接测出图像中各像素离相机的距离[18]，因此它比单目和双目相机能够提供更加丰富的图像信息以及更准确的深度信息。而且单目相机或双目相机在计算深度时有一定的时间成本，所以基于 RGB-D 相机的 SLAM 得到了广泛的研究[19][20]。

目前基于特征点法的 RGBD-SLAM 系统主要分为以单一点特征和以多特征两类为主。Mur-Artal 等人[21]提出基于单一特征点的经典的 ORB-SLAM2 算法，通过提取角点信息完成特征匹配，并通过特征数据关联得到物体运动位姿，但此系统在稀疏纹理场景下极易出现因特征数量过少而跟踪丢失的问题。Campos 等人[22]提出的ORB-SLAM3 依赖最大后验估计，并利用所有先前信息进行重定位，使系统可以在特征稀少场景中长期运行，但依然无法有效解决稀疏纹理场景下跟踪易丢失的问题。Von Gioi 等人[23]提出了线段检测器（Line Segment Detector,

LSD），该算法统计所有像素的梯度大小和方向，将梯度方向变化小且相邻的像素作为一个连通域，并筛选连通域为直线检测结果，基于此系统引入了多特征提取的方法。Gomez-Ojeda 等人[24]提出了 PL-SLAM 算法，在图像点特征稀少或分布不均匀的情况下引入 LSD 检测器，并在所有线程中使用两种特征，采用两种特征的组合描述能力作为闭环检测的依据，使得该算法在稀疏纹理场景下取得了不错的效果，但当环境中出现动态物体时容易出现特征提取紊乱的问题。PL-SLAM 系统框架如图 1-3 所示。

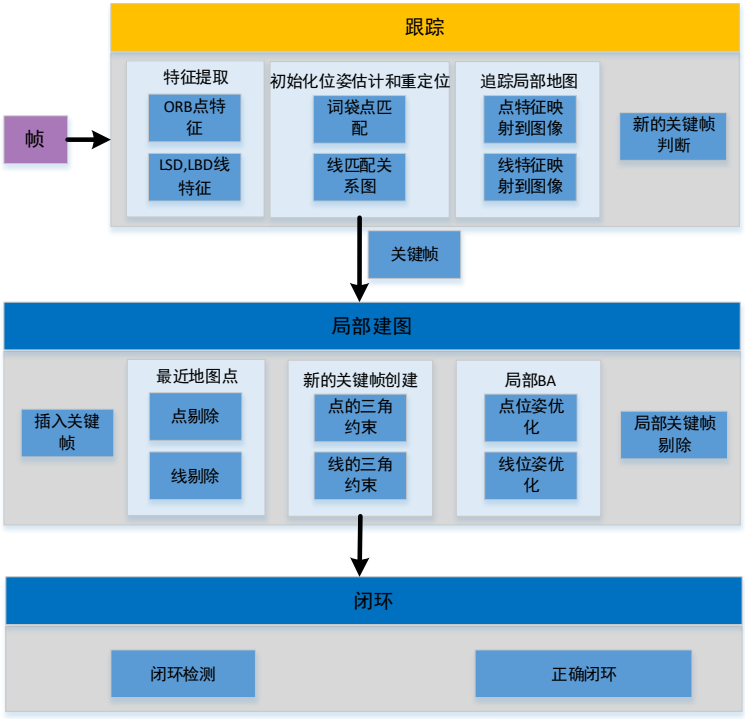


图 1-3 PL-SLAM 系统框架

Li 等人[25]针对稀疏结构化场景下，传统 SLAM 无法有效提取特征信息进行准确跟踪的问题，提出了基于点、线和面的解耦-精化方法，并在位姿估计模块中使用基于曼哈顿关系的 RGBD-SLAM[26]。该算法通过从周围环境中提取几何特征来提高跟踪和映射精度，但在特征提取环节由于深度相机得到的深度图在物体边界出现深度不连续以及场景中闯入动态物体的问题，使得 RGBD-SLAM 算法无法稳定高效运行。为了减少动态物体的影响，Zhang 等人[27]利用点和线特征计算摄像机姿态，并结合语义分割网络与 K-均值聚类算法[28]来检测潜在的动态特征，采用两种一致性检查策略，检测和过滤动态特征，该算法能够有效解决低动态稀疏场景下的动态物体，但易因深度信息的不准确以及特征信息提取过少而出现动态物体剔除不完全的问题。

1.2.3 动态场景下视觉 SLAM 研究现状

多数经典的视觉 SLAM 算法都实现了静态环境中的定位与构图，然而，当场景中存在动态物体，会导致定位精度减低和构建的静态地图可利用率降低，从而使系统难以进行准确的位姿估计与定位，因此在动态环境中能稳定运行的 SLAM 是目前研究的热点之一[28]。

针对环境中动态物体的问题，Sun 等人[29]提出了一种基于深度相机的去除动态物体的算法，该算法利用静态特征信息检测平面区域，将检测结果与重投影误差相结合，细化了运动与静态区域。随着人工智能的快速发展，基于图像目标检测的网络取得了巨大进步。因此基于深度学习方法的视觉 SLAM 被广泛的提出。Kaneko 等人[30]提出的 Mask-SLAM 通过语义分割网络 DeepLabV2 分割图像，并通过语义标签来识别动态物体，但未考虑已移动的物体。Zhong 等人[31]提出的 Detect-SLAM 算法对潜在目标点进行运动传播然后检测语义信息，并保留阈值之下的静态点，进而提高动态物体的识别精度。Bescos 等人[32]使用卷积神经网络对潜在的动态物体进行分割，通过联合 Mask RCNN 和多视几何对动态物体分割剔除，且将之前的彩色以及深度图映射到待修复帧上完成修复，该方法可以有效的判断动态物体，但易受到噪音的影响。Yu 等人[33]提出的 DS-SLAM 通过 SegNet 网络分割目标物体并联合运动一致性检测方法来降低动态目标的影响，从而大大提高了动态环境下的定位精度，但该方法易因剔除区域特征点数量过少而导致系统运行失败。Liu 等人[34]提出的实时可视化动态 RDS-SLAM 算法，通过添加语义分割和优化线程实现动态环境中动态物体实时跟踪，但该算法受限于卷积网络的分割能力，易导致因分割不完全而失败。Xiao 等人[35]提出 Dynamic-SLAM 算法，该算法利用 SSD 方法进行先验目标检测，根据目标区域特征识别动态物体并剔除动态特征点，但该系统中难以精准分割动态区域。

1.3 研究内容与章节安排

本文针对传统视觉 SLAM 算法在动态场景下难以对动态物体进行准确识别以及剔除动态物体后无法快速高效的修复等问题，提出一种基于语义分割[36][37]和背景修复的动态视觉 SLAM 算法。该算法包含潜在动态物体分割和背景修复

两个环节。潜在动态物体分割环节通过本文所提的 MSNET 语义分割网络对图像帧中存在的物体进行语义分割，并提取特征点。在背景修复[38]环节，通过物体运动判断获取动态点并剔除动态物体，利用最优近邻像素匹配[39]对已经剔除动态物体区域进行相似搜索、像素匹配，并进行特征点的提取，为 SLAM 系统的定位提供丰富信息。

针对 RGBD-SLAM 在特征提取环节，深度图易出现深度不连续、单一特征点提取特征信息不足以及突然出现动态物体造成轨迹误差增大的问题，本文提出基于区域映射深度优化的动态多特征 RGBD-SLAM 算法。该算法包含深度图优化和动态物体剔除两个环节。深度图优化环节通过轻量级的 AFFormer[40]网络提取语义信息，并将提取的语义信息映射到其对应的深度图上进行近邻修复，从而完成深度图的优化。动态物体剔除环节包含点特征剔除动态物体线程和点线面特征提取线程。首先点特征联合多视图几何完成动态物体的剔除，并利用双向映射背景修复模型补全剔除区域的静态信息，最后进行点线面特征的提取。该算法引入区域映射深度优化以解决深度不连续的问题，通过联合深度图与多视图几何剔除动态物体，改善了现有 RGBD-SLAM 算法深度信息不准确的问题，提高了动态场景下系统运行的稳定性，为 SLAM 系统稳定高效运行提高了良好的保障。论文总体技术路线如图 1-4 所示。

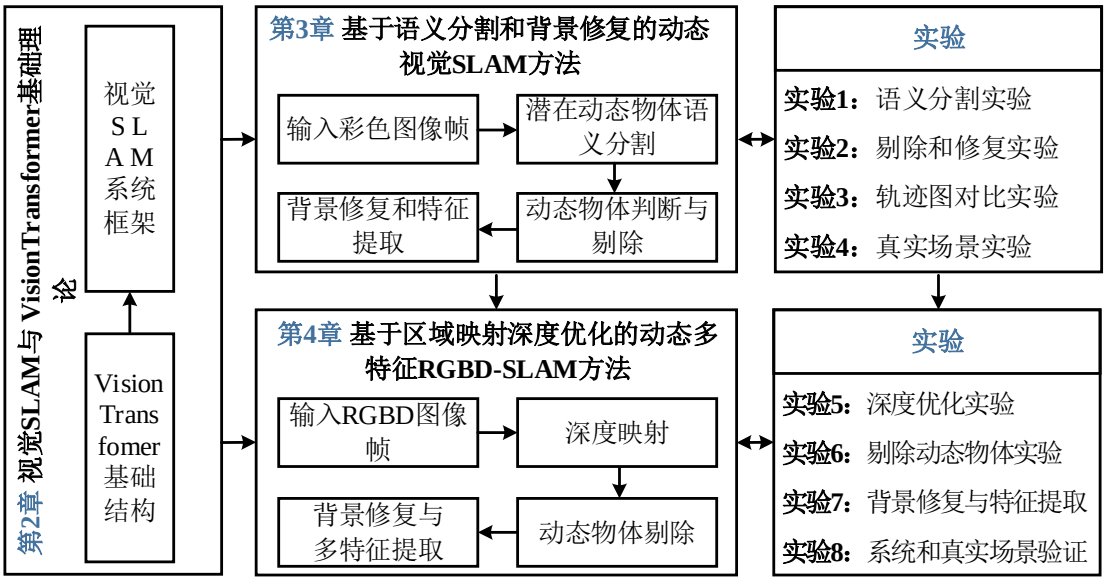


图 1-4 论文总体技术路线

各章节的主要内容如下：

第 1 章：绪论。首先介绍研究背景及意义，引出巡检机器人对人类的作用，

然后介绍电力巡检机器人的分类，介绍电力巡检机器人的核心技术视觉 SLAM 研究现状，特别是视觉 SLAM 和 RGBD-SLAM 技术的研究热点内容，最后重点介绍了动态场景下视觉 SLAM 和 RGBD-SLAM 算法的定位与构图技术，并对本文各章节之间的关联进行了介绍。

第 2 章：视觉 SLAM 与 Vision Transformer 基础理论。首先分别对视觉 SLAM 系统的视觉里程计、后端优化、回环检测和建图以及传统深度优化这几部分的基础理论和流程进行分析和模型构建。然后探究 Vision Transformer 的图像块的处理和工作原理，最后建立 Vision Transformer 网络分割模型，为电力巡检机器人视觉 SLAM 算法能在动态场景中稳定运行奠定基础。

第 3 章：基于语义分割和背景修复的动态视觉 SLAM 算法。针对传统视觉 SLAM 算法在动态场景下难以对动态物体进行准确识别以及剔除动态物体后无法有效修复剔除区域等问题，本算法设计一种残差-双重金字塔主干网络对动态物体区域特征进行多重提取，同时提出多类特征 Transformer 分割模块提高潜在动态物体像素权重与增强全局语义信息，精确标记出潜在动态物体。通过多视图几何判断出动态物体并进行剔除，根据最优近邻像素匹配方法完成剔除区域静态背景修复，并提取修复区域特征点用于位姿估计和定位。

第 4 章：基于区域映射深度优化的动态多特征 RGBD-SLAM 算法。针对 RGBD-SLAM 在特征提取环节，深度图易出现深度不连续、单一特征点提取特征信息不足以及突然出现动态物体造成轨迹误差增大的问题，提出一种基于区域映射深度优化的动态多特征 RGBD-SLAM 算法。利用轻量级的语义分割网络获取 RGB 帧中语义信息，将得到的语义掩码映射到其对应的深度图像中，并在掩码映射后的区域内完成深度图优化。为减少动态物体对 SLAM 系统轨迹精度的影响，通过迭代最近点求解相机位姿，结合多视图几何得到物体位姿、运动估计矩阵和动态视觉误差，估计物体运动状态并剔除动态物体。根据双向映射背景修复补全剔除区域的静态信息，并进行点线面特征提取以完成电力巡检机器人视觉 SLAM 算法稳定高效运行。

第 5 章：全文总结展望。对本文所有的研究内容进行分析与总结，并展望研究过程中新出现的待解决问题。

第 2 章 视觉 SLAM 与 Vision Transformer 基础理论

本章主要介绍基于语义分割的视觉 SLAM 基础理论。首先分别对视觉 SLAM 系统的视觉里程计、后端优化、回环检测和建图以及传统深度优化几个部分相关基础知识叙述。然后对 Vision Transformer 语义分割模型基础进行分析，利用语义分割网络对环境中的物体进行类别信息的区分，为后续动态场景下视觉 SLAM 的研究奠定理论基础。

2.1 视觉SLAM系统框架

视觉 SLAM 主要由视觉里程计、后端优化、回环检测和建图四个模块组成。视觉 SLAM 基础框架图如图 2-1 所示。本节主要对视觉里程计、后端优化以及传统深度优化进行分析与介绍，对回环检测和建图进行简单阐述[41]。

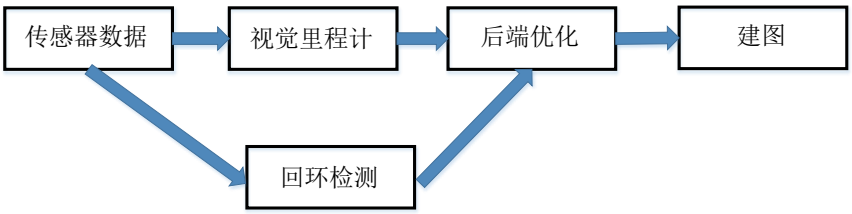


图 2-1 视觉 SLAM 基础框架图

2.1.1 视觉里程计

视觉里程计主要根据相邻帧关系估计运动路径，为后端优化提供先验数据。根据跟踪图像帧方法的不同分为特征点法和直接法，本文主要使用特征点法。基于特征点法的视觉 SLAM 系统有可分为以单一特征和以多特征为主两大类，提取的特征包含点特征、线特征以及面特征[42][43]。

点特征通常选取 ORB 特征，ORB 特征因其具高效率和本地性的性质被广泛应用于视觉 SLAM 中，ORB 特征是由关键点和描述子两部分组成。ORB 特征的关键点是在 FAST 角点的基础上通过计算特征点方向得到的，描述子采用 BRIEF（Binary Robust Independent Elementary Feature）。FAST 角点检测过程如下（如图 2-2 所示）：

1、在彩色图像中选取像素 p，假设它的亮度为 I_p。

2、设置比较阈值 T（例如，取亮度值 I_p 的 20%）。

3、绘制一个以像素点 p 为几何中心，半径为 3 个像素的虚拟采样圆形，选取与采样圆形边缘相交的 16 个采样像素点。

4、若被选取的 16 个采样像素点中，存在 N 个连续的像素点的亮度小于 I_p-T 或大于 I_p+T，则该像素点 p 可被定为像素关键点（本文 N 取 12，通常为了快速检测，一般先检测第 1 和 9，如果符合再检测第 5 和 13）。

5、循环以上四步并对每一个像素执行相同操作。

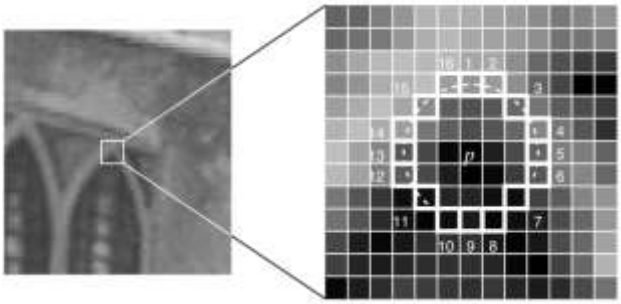


图 2-2 FAST 角点的检测

通过构建 8 层图像金字塔，按 1.2 倍率缩放图像，并按每层金字塔提取特征点，从而实现 ORB 特征空间尺度不变特性。旋转不变特性实现方法如下。

1、在一个小的像素块 B 中，定义像素块的矩如公式（2-1）所示：

$$m_{pq}=\sum_{x,y\in B}x^p y^q I(x,y),\quad p,q=\{0,1\} \tag{2-1}$$

2、通过矩可以找到像素块的质心，如公式（2-2）所示：

$$C=\left(\frac{m_{10}}{m_{00}},\frac{m_{01}}{m_{00}}\right) \tag{2-2}$$

3、连接像素块的几何中心 O 与质心 C，得到一个方向向量 $\overrightarrow{OC}$，于是特征点的方向可以定义为公式（2-3）：

$$\theta=\arctan(m_{01},m_{10}) \tag{2-3}$$

BRIEF 描述子的核心思想是在关键点 p 的周围以一定模式选取 M 个点对，把这 M 个点对的比较结果组合起来作为描述子。通常情况下采用 0 和 1 的描述

向量来进行关键点附近两个随机像素点的比较。选取的可视化如图 2-3 所示。

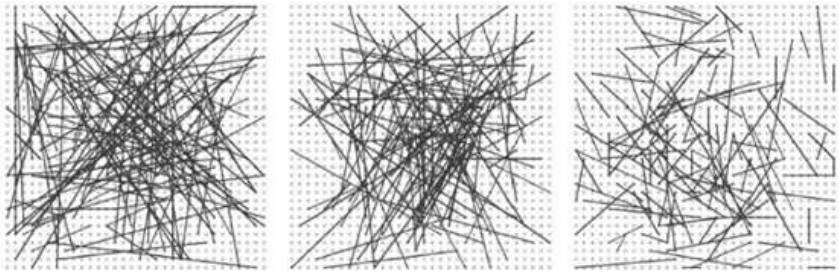


图 2-3 可视化

视觉 SLAM 中通常使用 LSD（Line Segment Detector）作为线特征的提取方法，LSD 算法是基于图像边缘的检测和分析，通过边缘像素的方向和强度信息来提取直线段。其基本思想是在边缘像素上构建全局分支和模拟部分分支图，然后基于这个图来提取直线段。LSD 算法的主要步骤如下：

1、边缘检测：提取图像的边缘。

2、边缘像素排序：将边缘像素按照强度降序进行排序。

3、全局分支：在边缘像素上构建全局分支图。从强度最高的边缘像素开始，计算与其相邻的边缘像素的方向差，并根据方向差构建全局分支图。

4、模拟部分分支：在全局分支图的基础上，模拟部分分支图。从最强的全局分支开始，根据边缘像素的方向和强度信息，计算模拟部分分支图。

5、非极大抑制：根据边缘像素的方向和强度信息，进行非极大抑制操作，得到最终的直线段。

面特征的提取由空间平面方程表示，使用点法式来求解平面。设 $\boldsymbol{n}=[a,b,c]^{\mathrm{T}}$ 为平面法向量，图像内任意两像素点组成的向量 $\boldsymbol{xx}_0=[x-x_0,y-y_0,z-z_0]^{\mathrm{T}}$，由垂直向量点乘得到公式（2-4）：

$$\begin{aligned}\boldsymbol{n}^{\mathrm{T}}\boldsymbol{xx}_0 &= a(x-x_0)+b(y-y_0)+c(z-z_0)\\ &= ax+by+cz+(-ax_0-ay_0-az_0)\\ &= 0\end{aligned} \tag{2-4}$$

像素点 $\boldsymbol{p}=[x_1,y_1,z_1]^{\mathrm{T}}$ 到平面内的距离 D 由相关证明可得到公式（2-5）：

$$D=\frac{ax_1+by_1+cz_1+d}{\sqrt{a^2+b^2+c^2}} \tag{2-5}$$

若法向量为单位向量，则可写为公式（2-6）：

$$D = \boldsymbol{n}^{\mathrm{T}}\boldsymbol{p} + d \tag{2-6}$$

对于像素点 $\boldsymbol{p}$ 的参数化，已知图像平面点 $\boldsymbol{m} = [u,v,1]^{\mathrm{T}}$ 以及逆深度 $\rho = \frac{1}{h}$，则点的坐标为公式（2-7）：

$$\boldsymbol{p} = [uh,vh,h]^{\mathrm{T}} = h\boldsymbol{m} = \frac{\boldsymbol{m}}{\rho} \tag{2-7}$$

如果点 $\boldsymbol{p}$ 在平面内，则得到公式（2-8）：

$$h \cdot \left(\boldsymbol{n}^{\mathrm{T}} / d\right)\boldsymbol{m} = 0 \tag{2-8}$$

因此逆深度可以写为公式（2-9）：

$$\rho = \frac{1}{h} = \frac{-\boldsymbol{n}^{\mathrm{T}}}{d}\boldsymbol{m} = \theta^{\mathrm{T}}\boldsymbol{m} \tag{2-9}$$

平面可以转化为公式（2-10）：

$$\theta = [\theta_1,\theta_2,\theta_3]^{\mathrm{T}} = -\frac{\boldsymbol{n}}{d} \tag{2-10}$$

RGBD 相机进行 3D-3D 点机器人位姿的解算时通常采用迭代最近点（Iterative Closest Point, ICP）算法。假设两帧图像间关联的 3D 特征点分别为 p_i 和 p_i'，则两特征点之间关系计算方式如式（2-11）所示：

$$p_i = \boldsymbol{R}p_i' + \boldsymbol{t} \tag{2-11}$$

其中，$\boldsymbol{R}$ 为旋转矩阵，t 为平移矩阵。求解相机位姿 $\boldsymbol{T}$ 计算方式如式（2-12）所示：

$$\boldsymbol{T} = \min_{\boldsymbol{R},\boldsymbol{t}} \frac{1}{2}\sum_{i=1}^{n} \| (p_i - (\boldsymbol{R}p_i' + \boldsymbol{t})) \|_2^2 \tag{2-12}$$

2.1.2 后端优化

SLAM 系统中前端视觉里程计只有短暂的记忆，仅能在极短的时间段保证系统的运动轨迹保持最优状态。但环境变化因素易造成图像帧间特征点匹配失误，导致错误的关联信息造成系统误差，并且随着 SLAM 系统运行时间的增加误差逐渐积累变大，对构建静态地图精度产生影响。后端优化是 SLAM 系统的重要

一环，可以提高地图精度，优化机器人位姿，解决数据关联问题以及增强系统鲁棒性。

后端优化方案主要有两种，一种是用长时间内状态估计，它使用过去以及未来的信息来更新当前的状态，常用的方法主要是卡尔曼滤波优化方法。另一种是当前状态只由过去的信息来决定，常用的方法主要是图优化方法。

本文主要采用的是优化方法是卡尔曼滤波，卡尔曼滤波的基本思想是结合系统的先验知识和传感器测量信息，使用递推的方式进行状态估计，它分为两个主要步骤。

1、预测步骤（Predict）。根据系统的动态方程，对系统当前的状态进行预测。这一步骤利用先前的状态估计 $\hat{x}_{k-1}$ 和运动模型，通过状态转移方程预测系统的下一步状态 $\breve{x}_k$，并计算出预测的状态的协方差 $\breve{P}_k$，如公式（2-13）所示：

$$\breve{x}_k = A_k\hat{x}_{k-1} + u_k, \breve{P}_k = A_k\hat{P}_{k-1}A_k^{\mathrm{T}} + R \tag{2-13}$$

2、更新步骤（Update）。在收到传感器的测量数据后，根据测量模型，对预测的状态进行调整。通过比较预测的状态与传感器测量值之间的差异，计算出卡尔曼增益 K，并将其应用于预测的状态和协方差，得到更新后的状态估计 $\hat{x}_k$ 和协方差 $\hat{P}_k$，最后计算后验概率的分布，如公式（2-14）和公式（2-15）所示：

$$K = \breve{P}_k C_k^{\mathrm{T}}\left(C_k\breve{P}_k C_k^{\mathrm{T}} + Q_k\right)^{-1} \tag{2-14}$$

$$\begin{cases} \hat{x}_k = \breve{x}_k + K\left(z_k - C_k\breve{x}_k\right) \\ \hat{P}_k = \left(I - KC_k\right)\breve{P}_k \end{cases} \tag{2-15}$$

2.1.3 回环检测和建图

回环检测指的是检测相机是否经过同一个地方，是为了解决整个 SLAM 系统出现的累计误差。相机判定没有回环会导致无法得到精确轨迹和可重复使用的静态地图，相机判定回环能够提高当前数据与所有先验数据的关联，而且可以利用回环检测进行重定位。一般情况下希望不借助其它设备直接使用自身的传感器数据进行回环检测，通常使用两帧图像之间的相似性确定是否回环。

建图是构建传感器所处地点的局部静态地图以及全局的静态地图的过程。视觉 SLAM 建图主要是为了进行定位、导航、避障、重建与人机交互。视觉 SLAM

算法通常有稀疏点云地图，半稠密地图，稠密地图和八叉树地图这几种建图效果。

2.1.4 传统深度优化

深度相机由于环境和设备自身的影响在生成深度图像时容易出现深度信息不连续的问题，在后续的应用过程中会降低 SLAM 系统位姿估计精度。传统的深度优化算法包含双边滤波算法、联合双边滤波算法以及融合超像素分割 RGB 图像和深度图像配准的深度图像修复算法，本文基于融合超像素分割 RGB 图像和深度图像配准的深度图像修复算法进行改进。基础算法流程如图 2-4 所示。

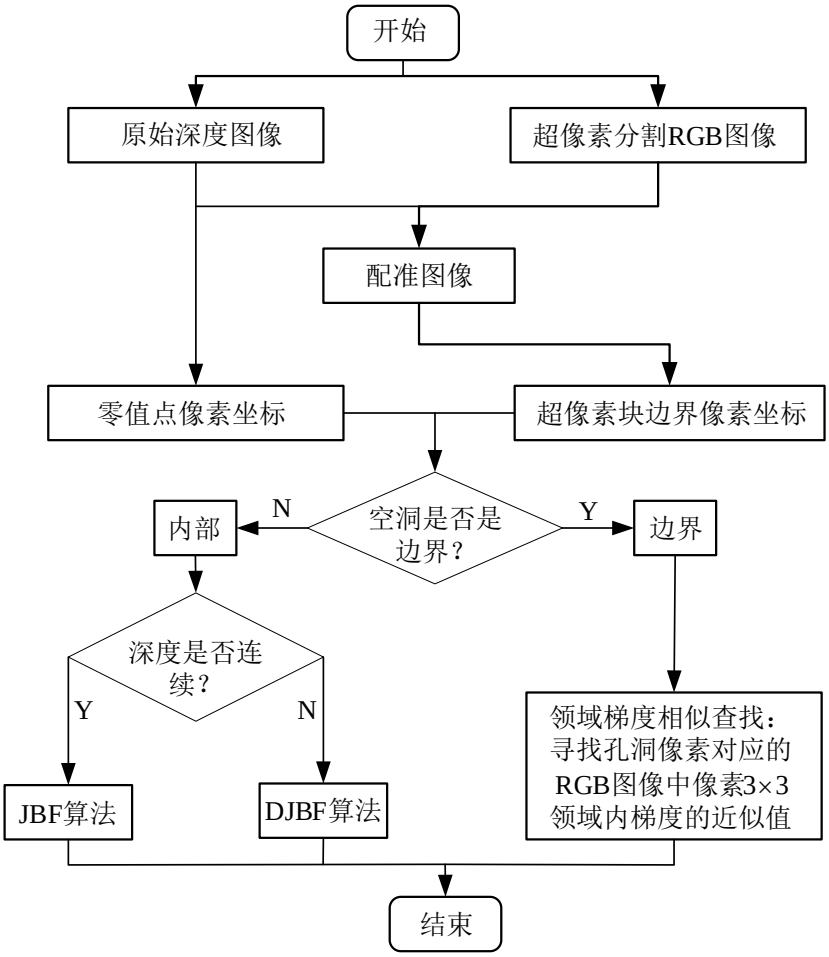


图 2-4 基础算法流程

2.2 Vision Transformer 基础结构

Vision Transformer 是一种基于 Transformer 的图像分割网络，主要功能为图像识别和图形分类。它将一张图像切割成一些固定尺寸大小的图像块，并对每个图像块进行线性嵌入，同时添加位置嵌入，然后将产生的向量序列输入到标准的

Transformer Encoder，并通过 Transformer 的注意力机制以及激活函数进行下游任务。Vision Transformer 结构如图 2-5 所示。

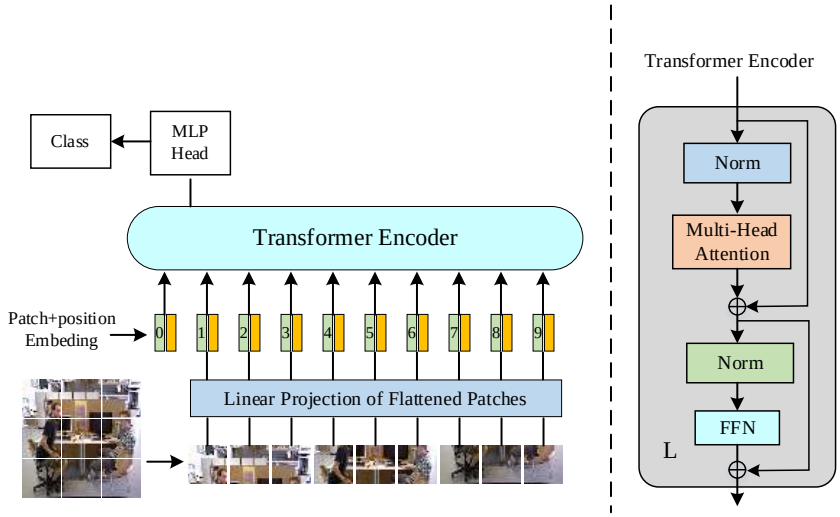


图 2-5 Vision Transformer 结构

2.2.1 图像块的处理

Vision Transformer 可直接对输入的图像进行处理。将尺寸为 $S\times S\times 3$ 图像划分为若干个块(patch)，每个块大小相同，均为 $n\times n\times 3$，输入图像经过划分后得到$(S/n)^2$ 个块，将每个块映射到一维向量中，得到的块尺寸大小为 $n\times n\times 3$。通过映射得到一个长度为 $3n^2$ 的词元（Tokes），Transformer 的自注意机制无法标记输入词元的顺序，为此引入了位置编码（Position Embedding），位置编码采用的是一个可用于训练的参数，将其直接加到词元上，再通过一个丢弃层，得到输入的二维矩阵序列，并将这些二维矩阵序列作为 Transformer Encoder 的输入。图像块处理如图 2-6 所示。

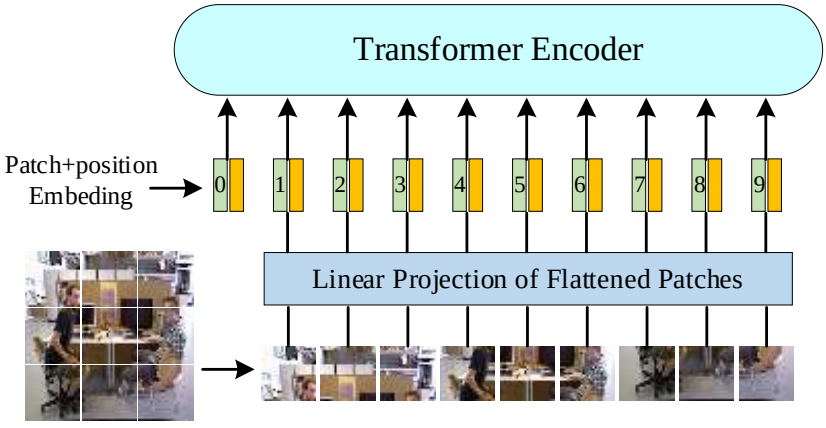


图 2-6 图像块处理

2.2.2 工作原理

多头自注意力机制是 Vision Transformer 的核心工作原理，多头自注意力机制主要是由多个单头注意机制构成，单头注意机制包含转换层和注意力层。

转换层输入词元 $\boldsymbol{A}\in\mathbb{R}^{n_a\times d_a}$、$\boldsymbol{B}\in\mathbb{R}^{n_b\times d_b}$ 通过线性层进行线性变换为查询向量 $\boldsymbol{Q}$、键向量 $\boldsymbol{K}$ 和值向量 $\boldsymbol{V}$，输入词元长度为 n、维度为 D。通过向量得到矩阵，其计算方法如式（2-16）所示：

$$\begin{cases}\boldsymbol{Q}=\boldsymbol{A}\boldsymbol{W}^{\boldsymbol{Q}}\\ \boldsymbol{K}=\boldsymbol{B}\boldsymbol{W}^{\boldsymbol{K}}\\ \boldsymbol{V}=\boldsymbol{B}\boldsymbol{W}^{\boldsymbol{V}}\end{cases} \tag{2-16}$$

其中，$\boldsymbol{W}^{\boldsymbol{Q}}\in\mathbb{R}^{D_a\times D^k}$，$\boldsymbol{W}^{\boldsymbol{K}}\in\mathbb{R}^{D_b\times D^k}$，$\boldsymbol{W}^{\boldsymbol{V}}\in\mathbb{R}^{D_b\times D^v}$，$D_k$ 是查询向量和键向量的维度，D_v 是值向量的维度。查询向量由词元 $\boldsymbol{A}$ 线性变换而来，键向量和值向量从词元 B 线性变换而来，编码层和解码层是自注意力机制主要应用的地方。

注意力层对三种向量进行交互，首先对查询向量和键向量进行聚合计算，然后将产生的结果与值向量再进行聚合计算，其计算方式如式（2-17）所示：

$$\text{Attention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\text{Softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{D_k}}\right)\boldsymbol{V} \tag{2-17}$$

其中，Softmax 为激活函数，$\sqrt{D_k}$ 为变换维度。

注意力权重是由查询向量和键向量生成，并进行变换维度，然后使用激活函数归一化。并将归一化后的权重分配给值向量中相应的词元，进而得到最终的特征信息的输出。归一化计算方式如式（2-18）所示：

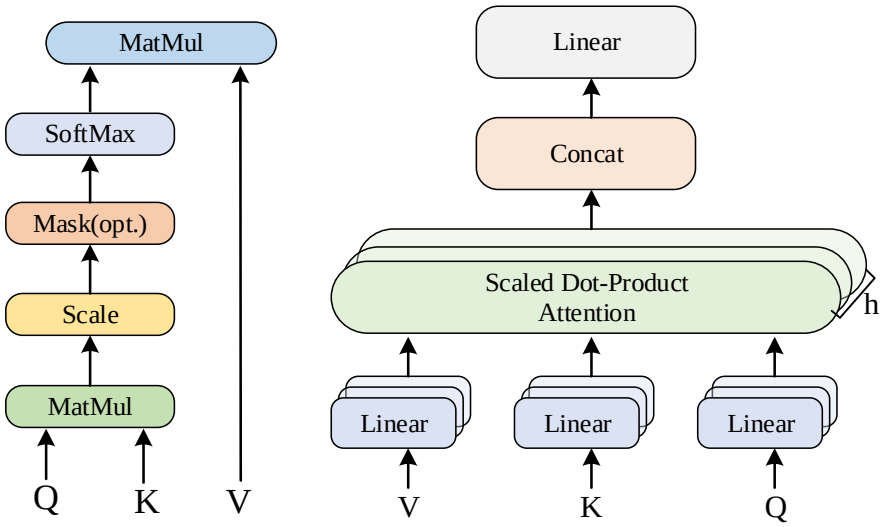
$$y_i=\frac{e^{z_i}}{\sum_j e^{z_j}} \tag{2-18}$$

其中，z_i 为 i 位置缩放后的值，y_i 表示归一化值。

单头注意力机制仅能关注单一或不同位置的特征信息，且只能采用一组矩阵获取输入特征信息，当输入多元特征信息时，单头注意力机制难以进行精确的输入特征信息表达。于是将多个头获取的信息联合，得到与输入词元同一维度的特征信息表达。多头自注意力机制输出矩阵计算方式如式（2-19）所示：

$$\begin{cases} \text{MultiHead}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = \text{Concat}(\text{head}_1,\cdots,\text{head}_n)\boldsymbol{W}^{O} \\ \text{head}_i = \text{Attention}(\boldsymbol{Q}\boldsymbol{W}_i^{Q},\boldsymbol{K}\boldsymbol{W}_i^{K},\boldsymbol{V}\boldsymbol{W}_i^{V}) \end{cases} \tag{2-19}$$

其中，参数矩阵 $\boldsymbol{W}_i^{Q} \in \mathbb{R}^{D_{model} \times D_k}$，$\boldsymbol{W}_i^{K} \in \mathbb{R}^{D_{model} \times D_k}$，$\boldsymbol{W}_i^{V} \in \mathbb{R}^{D_{model} \times D_v}$，$\boldsymbol{W}^{O} \in \mathbb{R}^{D_{model}}$ 为线性变换矩阵，自注意力与多头自注意力如图 2-7 所示。



（a）自注意力图　　（b）多头自注意力图

图 2-7 自注意力与多头注意力

2.3 本章小结

本章主要介绍了视觉 SLAM 与 Vision Transformer 分割模型的基础原理。首先分别对视觉 SLAM 系统的视觉里程计、后端优化、回环检测和建图以及传统深度优化这几部分的基础理论和流程进行分析和相关基础模型构建。然后探究 Vision Transformer 的图像块的处理和工作原理，最后建立 Vision Transformer 网络分割模型，为电力巡检机器人视觉 SLAM 算法的设计了奠定基础。

第 3 章 基于语义分割和背景修复的动态视觉 SLAM 算法

电力巡检机器人运行于动态场景时，由于动态物体的影响易出现特征点提取过少，使得电力巡检机器人视觉 SLAM 算法难以提取到足够的特征点信息，极大的影响视觉 SLAM 系统位姿估计精度以及定位任务。原始动态图像与特征点提取效果如图 3-1 所示。

图 3-1 原始动态图像与特征点提取效果

针对传统视觉 SLAM 算法在动态场景下难以对动态物体进行准确识别以及剔除动态物体后无法有效修复剔除区域等问题，提出一种基于语义分割和背景修复的动态视觉 SLAM 算法（Improved Dynamic Visual SLAM Algorithm Based on Semantic Segmentation and Background Repair，DSB-SLAM）。本章算法设计一种残差-双重金字塔主干网络对动态物体区域特征进行多重提取，同时提出多类特征 Transformer 分割模块提高潜在动态物体像素权重与增强全局语义信息，精确标记出潜在动态物体。通过多视图几何判断出动态物体并进行剔除，根据最优近邻像素匹配方法借助相邻帧中静态信息完成动态剔除后静态背景修复，并提取修复区域特征点用于位姿估计。本章技术路线如图 3-2 所示。

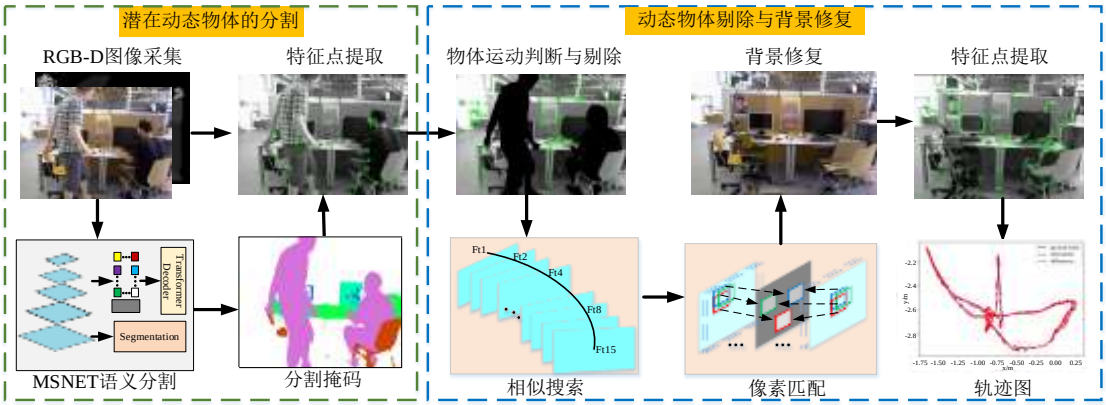


图 3-2 本章技术路线图

3.1 基于改进Vision Transformer的物体语义分割

为解决传统语义分割算法对动态物体分割时由于物体快速移动造成分割模糊或分割错误等问题，本章提出一种多类特征增强和多类特征引导的语义分割网络 MSNET(Multivariate Feature Fusion and Multivariate Feature Estimation Semantic Segmentation Network)。该网络包含动态物体区域特征多重提取的主干网络模块以及提高潜在动态物体像素权重和增强全局语义信息的多类特征 Transformer 分割模块。首先利用主干网络模块得到具有丰富信息和抗干扰能力强的多重特征图 C5，然后对特征图进行多类特征增强，并将融合特征输入到 Transformer Decoder 中，经 MLP head 得到 Class head 和 Mask head，同时对特征图 C5 进行多类特征引导，将估计特征输入多层上采样中并与 Class head 做点乘后输出类别特征图，再将类别特征图与 Mask head 做点乘，最后得到 Segmentation Map。本章提出的 MSNET 分割网络结构图如图 3-3 所示。

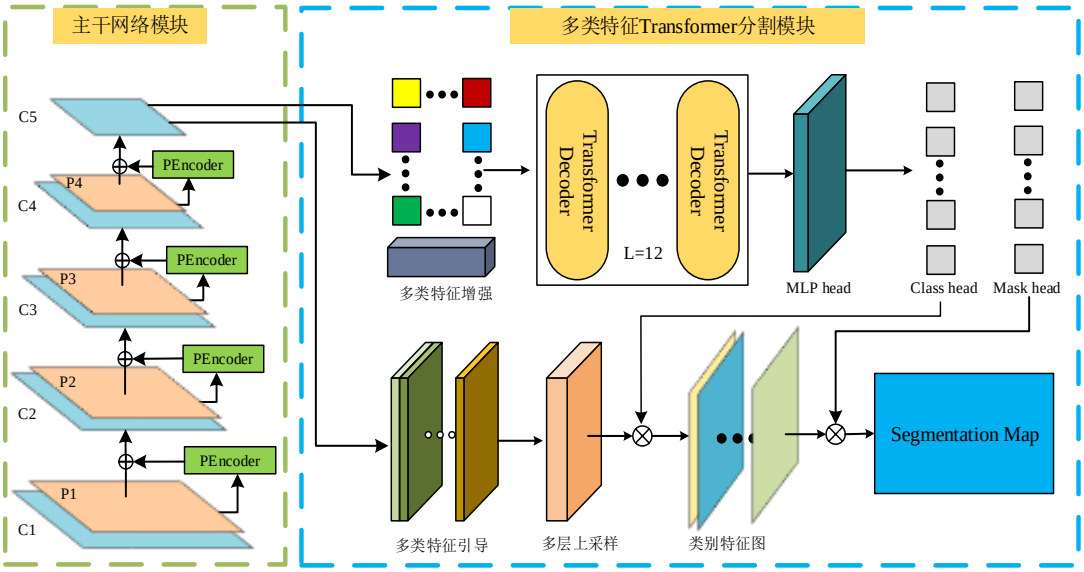


图 3-3 MSNET 分割网络结构图

3.1.1 主干网络模块

本章语义分割算法在 Vision Transformer Base 16 的基础上增加了采样金字塔构建残差-双重金字塔网络（Residual - Double pyramid network），同时为了更加有效提取来自图像的多重特征，将具有丰富特征信息和边界信息的特征图 C1、C2、C3、C4 分别采样至下一级特征金字塔的维度得到下采样特征图 P1、P2、

P3、P4，将经过 PEncoder 模块的特征图和下采样特征图进行 concat 融合得到特征信息图 C5。

PEncoder 模块如图 3-4 所示。PEncoder 模块的作用为提取全局图像特征信息。本算法将特征图（C1、C2、C3、C4）输入到 PEncoder 模块中，对特征图进行全局 Patch Embedding 和 Position Embedding，将其依次输入到 MSA 和 DML 层中，从而得到特征图中有关全局的特征联系信息。经过 PEncoder 模块获得的特征图计算方式如式（3-1）、（3-2）和（3-3）所示：

$$\mathrm{MSA}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = \mathrm{softmax}(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{D}})\boldsymbol{V} \tag{3-1}$$

$$\boldsymbol{a}_{i-1} = \mathrm{MSA}(\mathrm{LN}(\boldsymbol{b}_{i-1})) + \boldsymbol{b}_{i-1} \tag{3-2}$$

$$\boldsymbol{b}_{i} = \mathrm{DML}(\mathrm{LN}(\boldsymbol{a}_{i-1})) + \boldsymbol{a}_{i-1} \tag{3-3}$$

其中，$\boldsymbol{Q}$ 表示查询向量，$\boldsymbol{K}$ 表示键向量，$\boldsymbol{V}$ 表示值向量，D 表示维度。其中 $i \in \{1, \ldots, L\}$，$\boldsymbol{b}_{i-1}$ 是输入的 tokens，$\boldsymbol{a}_{i-1}$ 是经过 MSA 层后得到的 tokens，$\boldsymbol{b}_{i}$ 是经过 PEncoder 模块得到的 tokens。

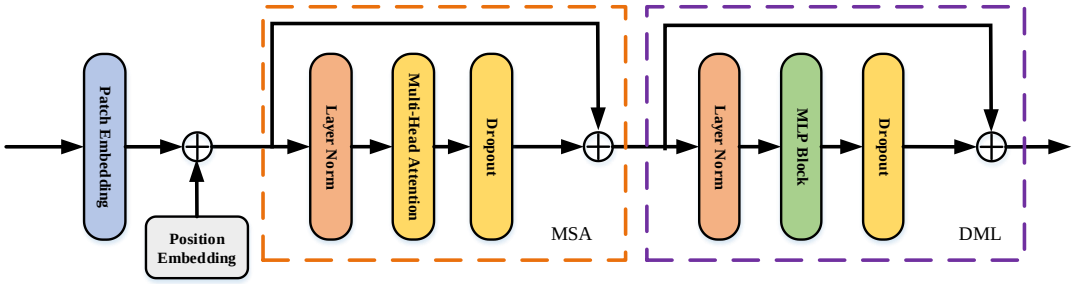


图 3-4 PEncoder 模块

3.1.2 多类特征 Transformer 分割模块

视觉 SLAM 系统在运行过程中容易受到动态目标的干扰，不能准确识别其中的动态目标会影响 SLAM 系统的鲁棒性。本章设计的多类特征 Transformer 分割模块由多类特征增强和多类特征引导两个子网络组成。首先对特征图 C5 进行多类特征增强，将具有多特征信息的多特征图 C5 输入到多类特征增强模块中，通过具有单类潜在动态特征信息的 m 个滤波器进行过滤，得到 m 个单类潜在动态特征信息并融合在一起，最终融合成多类潜在动态特征信息。该算法对潜在的动态目标像素权值进行增强，并将融合后的特征输入到 Transformer Decoder 中，

再进行 MLP 头处理，得到 Class 头和 Mask 头。同时，对特征图 C5 进行多类特征引导，将特征图 C5 输入到多类特征引导模块中，对 C5 进行通道维度和空间维度的关注机制，分别对各维度的特征信息进行平均和最大池化，形成具有全局语义信息的特征地图，并对特征地图进行融合，得到具有密集全局语义信息的特征地图，可以改善全局语义信息。将密集的全局语义信息特征图输入到多层上采样中，与 Class 头相乘输出类别特征图，再与 Mask 头相乘得到 Segmentation map。多类特征 Transformer 分割模块利用多类特征增强和引导的结合来生成像素权重，其中包含了单类强特征信息，并改善了感兴趣区域特征信息的提取。该方法增强了运动目标的像素权重，改善了全局语义信息，同时减轻了噪声干扰的影响。

多类特征增强网络如图 3-5 所示。多类特征增强网络作用为提高单类特征信息以及去除冗余的特征信息。本算法设计 m 个经过预训练具有单类特征信息的滤波器$\{f_1, \ldots, f_m\}$分别对特征图 C5 进行过滤，从而去除无关的、重叠的特征信息。具体来说，m 个滤波器是包含 60 个常见潜在动态对象特征信息的特征映射，使用 m 个滤波器分别对特征映射 C5 进行卷积，得到 m 个特征映射$\{\boldsymbol{t}_1, \ldots, \boldsymbol{t}_m\}$，并与潜在动态特征信息进行融合，进一步得到保留大量潜在动态特征的特征信息图，从而去除大量其他不相关特征信息的干扰，从而增强图像中的潜在动态特征信息。本章设计的滤波器决策方式如式（3-4）所示：

$$\begin{cases} \boldsymbol{y} \rightarrow \boldsymbol{\omega}_j \\ \boldsymbol{F}(\boldsymbol{\omega}_j) = \max\limits_{k=1,\cdots,c} P(\boldsymbol{\omega}_k \mid \boldsymbol{y}_1, \cdots, \boldsymbol{y}_m) \end{cases} \tag{3-4}$$

其中，$\boldsymbol{y}$ 表示特征图 C5 所具有的特征信息，$\boldsymbol{\omega}_j$ 表示第 j 类特征，$P(\boldsymbol{\omega}_k \mid \boldsymbol{y})$ 表示第 k 类的后验概率，$\boldsymbol{F}(\boldsymbol{\omega}_j)$ 为第 j 类特征信息。经由滤波器生成具有强烈单类特征的特征图$\{\boldsymbol{t}_1, \ldots, \boldsymbol{t}_m\}$，将其进行 concat 融合得到融合的特征图 $\boldsymbol{F}_t$。融合的特征图 $\boldsymbol{F}_t$ 具有单类别信息丰富，抗干扰能力强等优点。

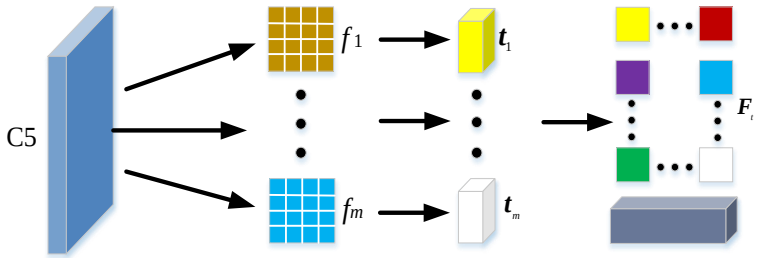


图 3-5 多类特征增强网络

多类特征引导网络如图 3-6 所示。多类特征引导网络作用为增大感兴趣区域通道权重，获取感兴趣区域空间位置，引导神经网络关注动态区域特征信息。将特征图 C5 在通道维度和空间维度进行连接，并将获取的特征图 $\boldsymbol{F}_c$ 与特征图 $\boldsymbol{F}_s$ 进行 concat 融合得到输出 $\boldsymbol{F}_e$。利用注意机制感知全局语义信息的能力，使生成的特征图 $\boldsymbol{F}_e$ 成为一个具有密集全局语义信息的特征图。一方面，通道注意机制在激励部分对不同的通道特征分配权重，为了对潜在动态对象的识别更加的准确，为此我们在分配权重时使用 sigmoid 函数作为激活函数来增强潜在动态特征信息的权重；另一方面，空间注意机制可以自动捕获图像中的重要区域，空间和通道注意机制的结合使得多类特征引导模块在工作时更加关注图像中的重要区域，进一步使得多类特征引导模块更加关注图像重要区域的潜在动态特征信息，进而引导模型关注潜在动态目标特征所在的区域。

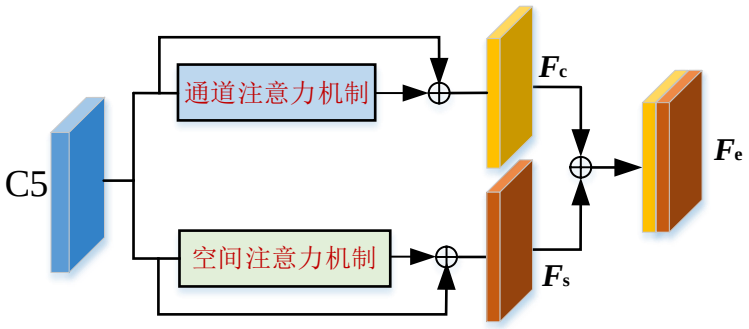


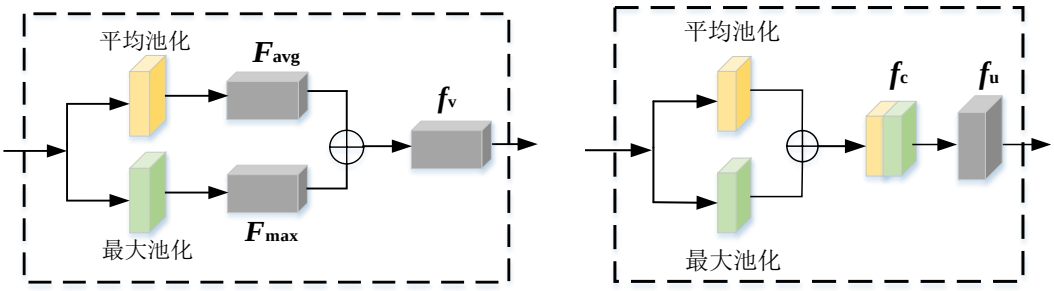
图 3-6 多类特征引导网络

注意力机制如图 3-7 所示。其中通道注意力机制作用是对各层通道分配相应的特征权重，经平均和最大池化操作得到特征 $\boldsymbol{F}_{avg}$ 与 $\boldsymbol{F}_{max}$，再将特征 concat 融合得到具有强关联性的特征信息 $\boldsymbol{f}_v$。空间注意力机制作用为增加特征图中动态区域像素值权重，将特征图输入空间注意力机制，通过平均和最大池化后进行 concat 融合形成特征图 $\boldsymbol{f}_c$，再通过 3×3×1 卷积层和激活函数得到权重分配特征图 $\boldsymbol{f}_u$，进而引导网络关注动态物体所在区域。经过通道注意力机制和空间注意力机制获得的输出 $\boldsymbol{f}_v$， $\boldsymbol{f}_u$ 计算方式如式（3-5）和（3-6）所示：

$$\boldsymbol{f}_v = \sigma(\eta(\boldsymbol{F}_{avg} + \boldsymbol{F}_{max})) \tag{3-5}$$

$$\boldsymbol{f}_u = \sigma(c(\boldsymbol{f}_c)) \tag{3-6}$$

其中，σ 表示 sigmoid 激活函数，η 表 tanh 函数，c 为 3×3×1 卷积层。



（a）通道注意力机制　　（b）空间注意力机制

图 3-7 注意力机制

3.2 动态物体剔除与背景修复

3.2.1 动态物体剔除

在潜在动态物体分割环境使用语义分割网络仅仅只能获取图像中的潜在动态物体，无法具体判断出动态物体。针对此问题本章引入自适应几何阈值进行动态物体判断，相机中任意相邻帧图像间匹配的特征点结合随机抽样一致（Random Sample Consensus, RANSAC）得到基础矩阵 $\boldsymbol{F}$，通过基础矩阵计算当前帧的极线。假设上一帧中的特征点像素位置为 $\boldsymbol{L}_1=[x_1,y_1,1]$，该点在当前帧中对应的位置为 $\boldsymbol{L}_2=[x_2,y_2,1]$，求解出点 $\boldsymbol{L}_1$ 投影到当前帧中的极线 $\boldsymbol{I}_1$，其计算公式如式（3-7）所示：

$$\boldsymbol{I}_1=\begin{pmatrix}\boldsymbol{X}\\ \boldsymbol{Y}\\ \boldsymbol{Z}\end{pmatrix}=\boldsymbol{F}\boldsymbol{L}_1=\boldsymbol{F}\begin{pmatrix}x_1\\ y_1\\ 1\end{pmatrix} \tag{3-7}$$

其中，$\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z}$ 代表基础矩阵中的向量。计算匹配的特征点到它对应极线的距离 d，如果这个距离超过阈值则被视为移动点，反之为静态点，进而确定动态物体。其计算公式如式（3-8）所示：

$$d=\left(\|\boldsymbol{X}\|^2+\|\boldsymbol{Y}\|^2\right)^{-\frac{1}{2}}\left|\boldsymbol{L}_2{}^{\mathrm{T}}\boldsymbol{L}_1\right| \tag{3-8}$$

3.2.2 背景修复

传统动态视觉 SLAM 算法剔除动态物体后，直接在剔除区域进行特征点的提取，会出现提取特征点数量过少、位姿估计不准确等问题，继而影响回环检测

和静态地图的构建。因此，本章提出一种基于近似最近邻匹配（PatchMatch）的最优近邻像素匹配的背景修复算法，借助先验帧中信息完成动态区域剔除后彩色当前帧和深度当前帧图像静态背景的修复。最优近邻像素匹配的背景修复算法包括相似搜索和像素匹配。其中相似搜索包含信息估计、参考帧选取两个步骤。首先通过信息估计找出需要选择的关键帧，然后通过相邻帧之间的运动连接得到判断阈值位移和旋转变化量，然后通过判断阈值得到参考帧，最后结合近似最近邻匹配算法使用参考帧进行背景恢复，其具体的步骤如下。

步骤 1：信息估计

本章采用帧间特征匹配来判断和获取图像信息，以待修复帧为起点沿着箭头方向移动一个 15 帧图像作为修复窗口。对每帧图像进行 ORB 特征点提取，计特征点数量为$\sum M_i$，对每帧与待修复帧进行特征匹配，匹配数记为$\sum N_i$。若$\sum M_i$与$\sum N_i$满足式（3-9）条件则认为此帧可进行下一步。

$$\left|\sum M_i - \sum N_i\right| < 60 \tag{3-9}$$

通过对极几何估计相邻帧之间的运动联系，得到高效的修复关系。假设$\boldsymbol{P} = \left[\mathrm{X}，\ \mathrm{Y}，\ \mathrm{Z}\right]^{\mathrm{T}}$为空间中的一点，在参考帧和目标帧中的投影分别为$p_1$和$p_2$，由相机模型可得像素点$p_1$和$p_2$的像素位置，如公式（3-10）所示：

$$\begin{aligned} s_1 p_1 &= \boldsymbol{K}\boldsymbol{P}, \\ s_2 p_2 &= \boldsymbol{K}(\boldsymbol{R}_{21}\boldsymbol{P} + \boldsymbol{t}_{21}). \end{aligned} \tag{3-10}$$

其中，s_1、s_2为尺度因子，$\boldsymbol{K}$为相机内参矩阵，$\boldsymbol{R}_{21}$和$\boldsymbol{t}_{21}$分别为目标图相对于参考图的旋转和平移矩阵，可由基础矩阵或本质矩阵求解恢复相机运动得到，之后将三维向量投影到二维平面，继而得出任意两帧之间的位移变化量Δp和旋转变化量$\Delta\theta$。

步骤 2：参考帧选取

良好的参考帧是进行近似最近邻匹配的必要前提，因此结合图像信息与位姿具体过程为：步骤 1 取正逆向 15 帧的时间窗口作为一组进行特征提取并匹配得到满足条件的待选参考帧，并将其加入待选参考帧库，当Δp超过阈值τ以及$\Delta\theta$大于阈值γ时将此帧加入待选参考帧库，当满足步骤 1 条件时将待选参考帧加入参考帧库。参考帧选取方法的数学表达式如式（3-11）所示：

$$rf=\left\{f_i\left|\left([\Delta p>\tau]\cup[\Delta\theta>\gamma]\cup\left[\left|M_i-N_i\right|<100\right]\right)\right.\right\} \tag{3-11}$$

由相似搜索得到参考帧库，像素匹配通过改进传统的近似最近邻匹配算法进行背景修复，其具体包括初始化，传播，搜索这三个过程。首先初始化，将待修复的目标图作为 A 图，参考帧为 B 图，在 A 图中随机选取一个像素块作为一个匹配块并随机赋予一个偏移量且在 B 图中找到一个匹配块与之对应；第二步传播，计算 A 图中的匹配块与 B 中匹配块的偏移量差，找到偏移量最小的值；第三步搜索，B 中每一个像素点在以现在的匹配块为中心的同心圆内找一个更加匹配的偏移量来代替当前偏移量，搜索的半径开始为图片的尺寸，然后以一半的收敛速度减少半径，直到结束，重复第 2 和第 3 步直到每个像素块都找到最合适最准确的偏移量，将此偏移量所对应的像素值赋予 A 图中相对应的像素块。偏移量计算公式如式（3-12）示：

$$R=\frac{\sum(\boldsymbol{A}-\boldsymbol{B})^2+\sum\left(\overline{\boldsymbol{A}}-\overline{\boldsymbol{B}}\right)^2}{n^2} \tag{3-12}$$

其中，$\boldsymbol{A}$、$\boldsymbol{B}$、$\overline{\boldsymbol{A}}$、$\overline{\boldsymbol{B}}$、$n^2$ 分别为 A 图中的原始矩阵块，B 图中的原始矩阵块，A 图中的偏移矩阵块，B 图中的偏移矩阵块，图像的大小。以两帧图像为例，最优近邻像素匹配示意图如图 3-8 所示。

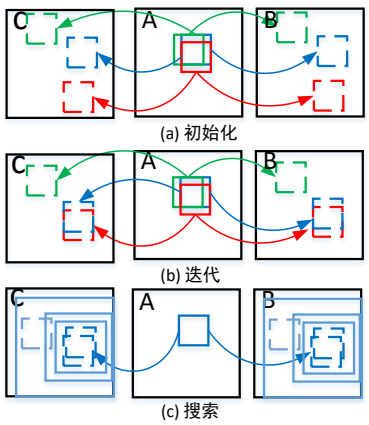


图 3-8 最优近邻像素匹配示意图

3.3 公开数据集实验及分析

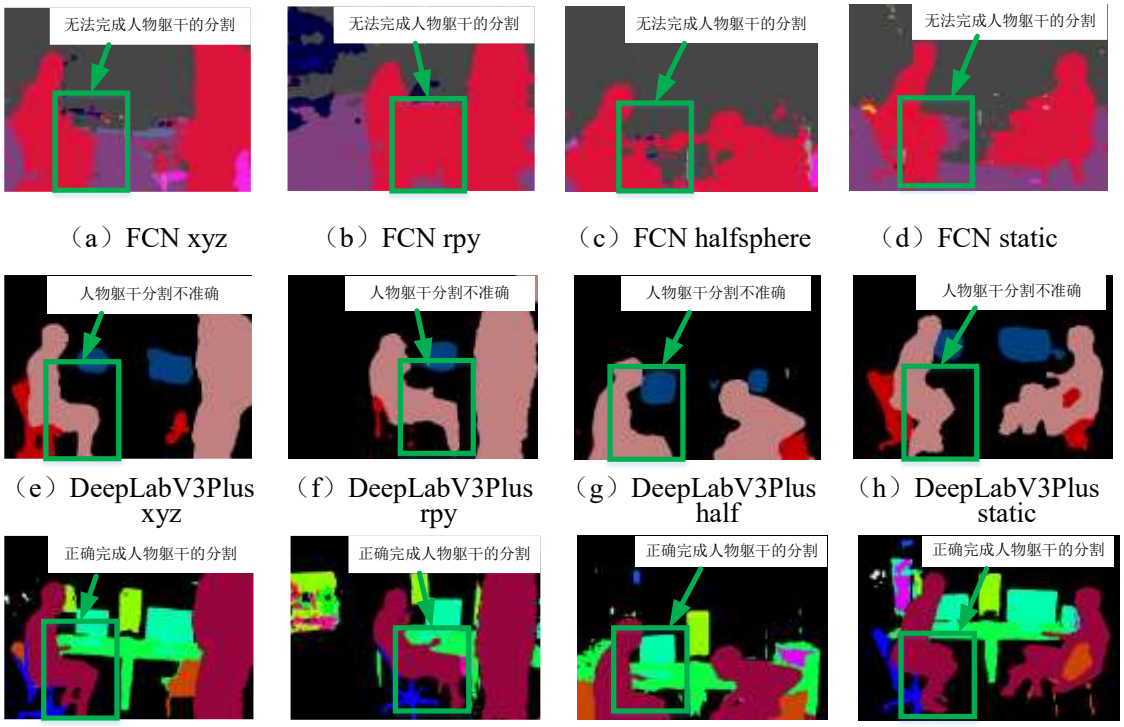
本章从语义分割实验、静态背景修复实验、轨迹图对比实验和定位精度四个方面对本章算法进行有效性分析。本章使用的电脑配置为：CPU 为 11th Gen

Intel(R) Core(TM) i5-11400F 处理器，主频 2.6GHZ，内存 16.00 GB (2667 MHz)，GPU 为 NVIDIA GeForce RTX 3060 (12288MB)，显存 24GB，系统为 Ubuntu18.04。

3.3.1 语义分割实验

本章使用 COCO2017 数据集作为训练，实验软件环境为 python3.8.6、pytorch1.11.1、cuda11.3，编译器为 pycharm2020.1，模型训练参数为：learning rate 设置为 0.001,epoch 设置为 128, batch size 设置为 4, dropout 概率为 0.55。

为了验证本章所提算法 MSNET 的语义分割性能，本章在公开数据集 TUM[44] 数据集中选择所需场景。FCN[45]、DeepLabV3Plus[46]和本章算法三种算法在 TUM 数据集 fr3_walking_xyz(xyz)、fr3_walking_static(static)、fr3_walking_rpy(rpy)、fr3_walking_halfsphere(half)四个序列中语义分割结果对比图如图 3-9 所示，绿色框为本章所作辅助标记。



（a）FCN xyz　（b）FCN rpy　（c）FCN halfsphere　（d）FCN static

（e）DeepLabV3Plus xyz　（f）DeepLabV3Plus rpy　（g）DeepLabV3Plus half　（h）DeepLabV3Plus static

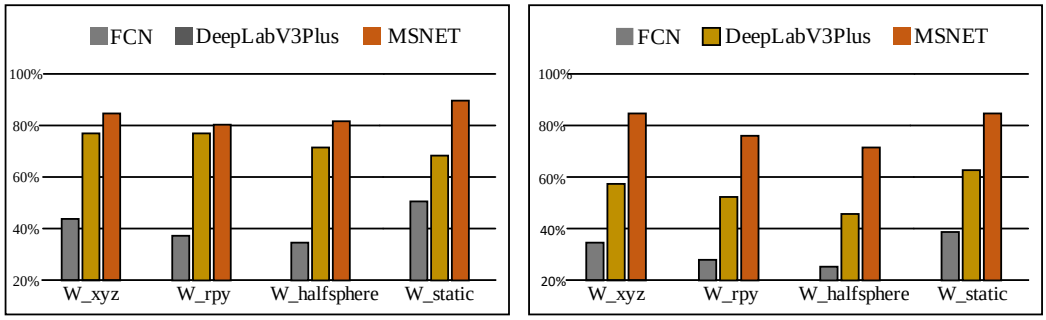
（i）本章算法 xyz　（j）本章算法 rpy　（k）本章算法 half　（l）本章算法 static

图 3-9 语义分割结果对比图

图 3-9（a）~（d）、3-9（e）~（h）、3-9（i）~（l）分别为 FCN、DeepLabV3Plus 以及本章算法在 TUM 数据集上的 4 种不同序列的分割效果图，通过 4 幅图语义分割效果对比可知，由于 FCN 与 DeepLabV3Plus 采用的是卷积网络结构，而卷积网络易到受感受野与卷积核大小的影响导致卷积网络仅针对局部特征有效，对

于人物躯干的分割情况较差。传统的 FCN 算法采用的是全卷积网络结构，对于类别丰富的图像极易出现分割错误的情况，如图 3-9（b）、3-9（c）所示。本章所提出的 MSNET 算法采用改进 Vision Transformer 的主干结构直接对图片进行特征纹理的提取，而采用结合 Transformer Decoder、多类特征增强模块和多类特征引导作为像素级别分割的依据，提高了物体识别与分割的准确率，因此，MSNET 可以有效准确的进行分割。

3 种语义分割算法的评测指标比较图如图 3-10 所示。图 3-10（a）为 3 种分割算法平均交并比(MIoU)，图 3-10（b）为 3 种分割算法类别平均像素准确率（MPA），MIoU 是指对每个类计算真实值于预测值结果的交并比然后对所有类别的 IOU 求均值，MPA 是每个类被正确分类像素数的比例，是标准的准确率度量方法，是语义分割中常用的评价指标之一，其值越大表示分割效果越好。本章所提的 MSNET 算法在四个序列上的平均交并比都达到了 80%以上，类别平均像素准确率都达到了 70%以上且都高于其它两种算法。

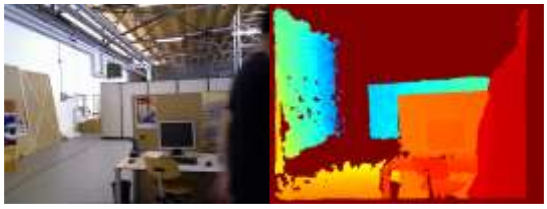


（a）3 种算法 MIoU 值对比　　（b）3 种算法 MPA 值对比

图 3-10 评测指标比较图

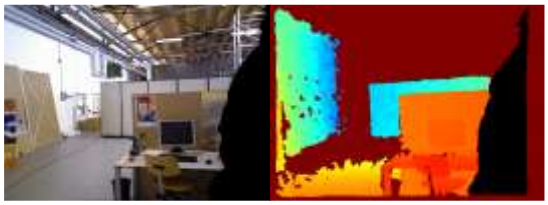
3.3.2 动态物体剔除和静态背景修复实验

本章采用最优近邻像素匹配对剔除区域进行补全。背景修复过程效果图如图 3-11 所示。其中图 3-11（a）、（b）所示为原始图像帧（彩色图像和对应的深度图像）。图 3-11（c）、（d）为剔除后的图像帧，图 3-11（e）、（f）所示为经过本章算法修复后的图像帧，本章算法利用自适应几何阈值剔除动态物体，并根据最优近邻像素匹配进行图像缺失区域的修复工作，从而建立完整全静态图像。因此，本章所提出的背景修复算法能够快速有效的修复剔除动态物体区域的静态背景，为提取到大量有效的特征信息做好了基础。



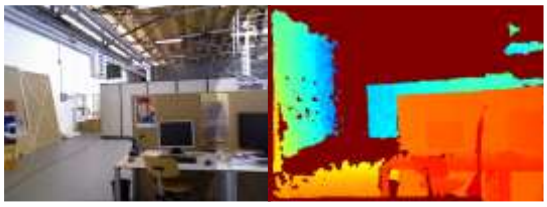
（a）原始图像帧 1

（b）原始图像帧 2



（c）剔除后的图像帧 1

（d）剔除后的图像帧 2



（e）修复后的图像帧 1

（f）修复后的图像帧 2

图 3-11 背景修复效果图

TUM 数据集上 3 种算法特征点提取对比结果图如图 3-12 所示。图 3-12（a）~（d）为 ORB-SLAM2 算法的四个特征点提取效果图，由于 ORB-SLAM2 无法剔除动态物体上的特征点，所以它的位姿估计精度较差。图 3-12（e）~（h）为 DynaSLAM 算法的四个特征点提取效果图，该算法可以剔除动态物体，但未进行剔除区域的修复，使得用于定位任务的特征点数量极少。图 3-12（i）~（l）为本章算法的四个特征提取效果图，本章算法剔除动态物体后进行了背景修复，从而提取更多的特征点，为定位任务提供了良好的环境特征信息。

（a）ORB-SLAM2 提取 1

（b）ORB-SLAM2 提取 2

（c）ORB-SLAM2 提取 3

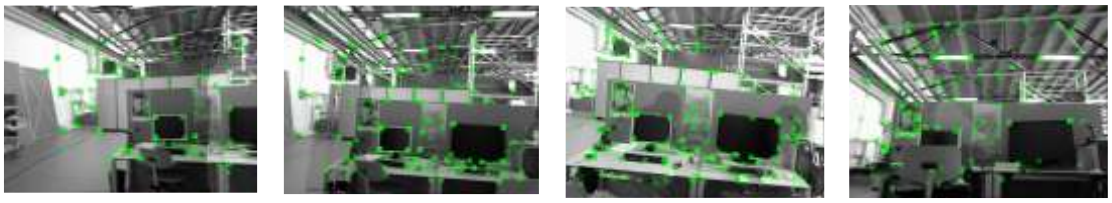
（d）ORB-SLAM2 提取 4

（e）DynaSLAM 提取 1

（f）DynaSLAM 提取 2

（g）DynaSLAM 提取 3

（h）DynaSLAM 提取 4

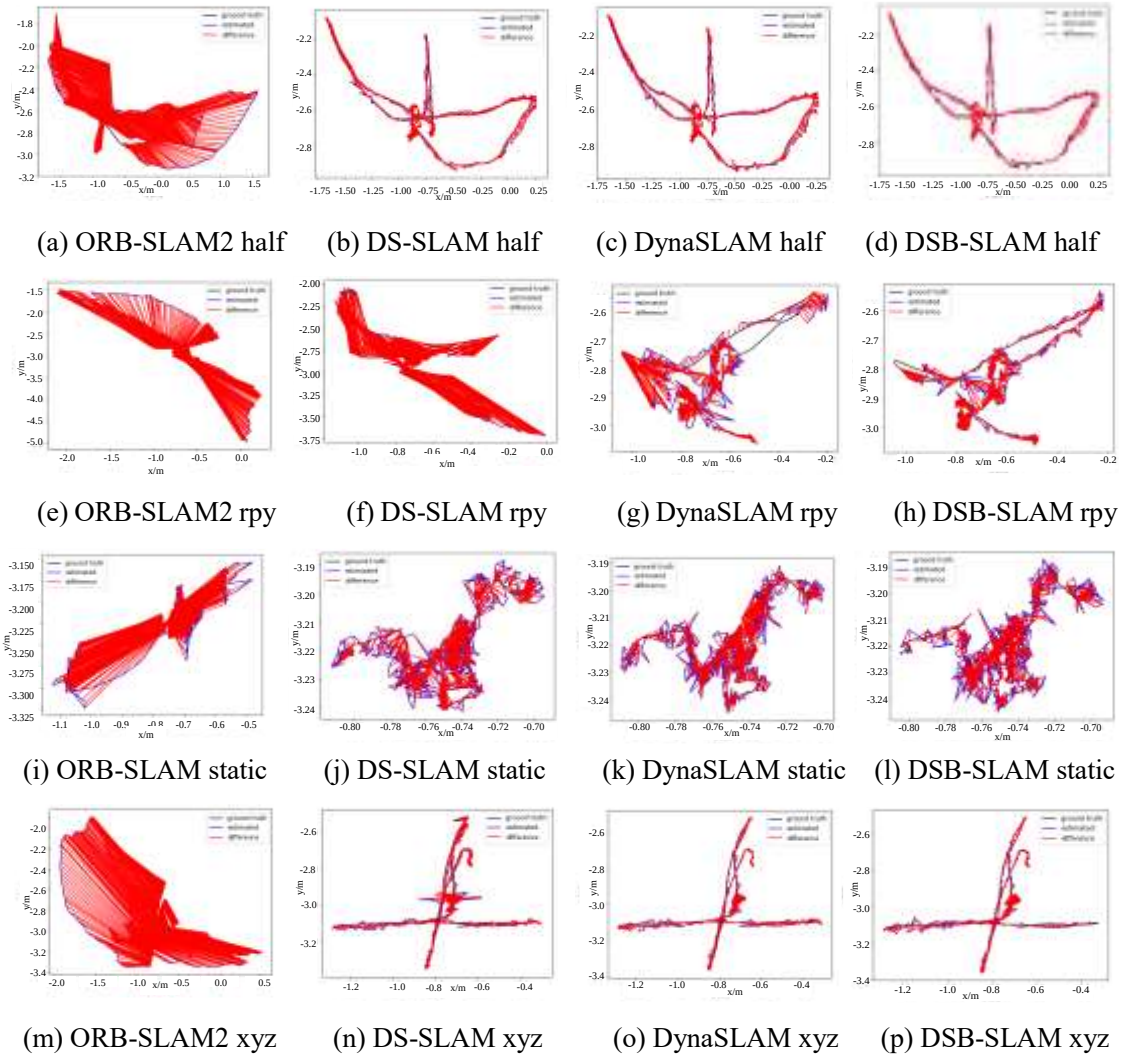


（i）本章提取 1　（j）本章提取 2　（k）本章提取 3　（l）本章提取 4

图 3-12 3 种算法特征点提取对比结果图

3.3.3 轨迹图对比实验和定位精度

本章选取 TUM 数据集四个高动态序列 fr3_walking_xyz(xyz)、fr3_walking_static(static)、fr3_walking_rpy(rpy)、fr3_walking_halfsphere(half)进行轨迹验证。4 种算法在不同序列场景下的轨迹图如图 3-13 所示。从图 3-13 中可以看出，由于 ORB-SLAM2 只能用于静态环境，不能识别动态对象，因此在动态环境中运行时轨迹误差较大，对后端优化构成影响较大。DS-SLAM 算法通过 Segnet 实例分割网络和运动特征点检测对环境中的动态目标进行识别和剔除，但 Segnet 实例分割网络只能识别 20 种目标，在种类信息丰富的室内环境中运行时，很容易忽略除去 20 种之外的动态物体的分割，导致其轨迹误差大于本章所提算法的轨迹误差。DynaSLAM 算法使用的是 MASK-RCNN 网络，受卷积网络感受野和卷积核大小的限制，使得 MASK-RCNN 网络无法获得全局语义信息，在分割大型物体时容易出现分割错误和分割不完整，因此其产生的轨迹误差仍然大于本章所提的算法。上述两种分割算法在准确、清晰地分割动态目标方面都存在局限性。此外，这两种方法都不能在动态目标剔除后进行背景修复，并且都不能从被剔除动态物体的区域中提取特征点。因此，这两种算法都减少了用于位姿估计和地图构建的静态特征点的数量，这将影响 SLAM 算法的定位精度和完整静态地图构建。本章提出的 DSB-SLAM 算法结合了 MSNET 语义分割算法，通过注意机制感知全局语义信息，精确标记潜在动态物体，结合自适应几何阈值判断策略准确、清晰地识别环境中的动态物体并将其剔除，同时最优近邻像素匹配方法有效地补充了剔除动态物体区域的静态背景。因此，本章所提出的算法精度高，轨迹误差小于其他三种算法，构造的轨迹图更加接近真实轨迹，视觉 SLAM 系统能够在动态环境中稳定的运行。



(a) ORB-SLAM2 half (b) DS-SLAM half (c) DynaSLAM half (d) DSB-SLAM half

(e) ORB-SLAM2 rpy (f) DS-SLAM rpy (g) DynaSLAM rpy (h) DSB-SLAM rpy

(i) ORB-SLAM static (j) DS-SLAM static (k) DynaSLAM static (l) DSB-SLAM static

(m) ORB-SLAM2 xyz (n) DS-SLAM xyz (o) DynaSLAM xyz (p) DSB-SLAM xyz

图 3-13 4 种算法在不同序列场景下的轨迹图

4 种动态场景序列的均方根误差(Root Mean Square Error，RMSE)如表 3-1 所示。均方根误差数值越低代表系统鲁棒性越高，定位精度越高。从表 3-1 中可以看出在这 4 种序列中本章算法都取得了较低的误差数值，其中 half 和 static 序列均方根误差小于 ORB-SLAM2、DS-SLAM 且与 DynaSLAM 算法的误差相持平，rpy 序列相较于 ORB-SLAM2、DS-SLAM、DynaSLAM 分别减少 95.6%、93.5%、17.1%，xyz 序列相较于 ORB-SLAM2、DS-SLAM、DynaSLAM 分别减少 97.2%、48.0%、13.3%，这表现了本章所提的算法具有良好的定位精度。

表 3-1 half、rpy、static、xyz 序列 RMSE（单位：米）

测试序列	ORB-SLAM2	DS-SLAM	DynaSLAM	本章算法
half	0.351	0.025	0.025	0.025
rpy	0.662	0.444	0.035	0.029
static	0.090	0.008	0.006	0.006
xyz	0.459	0.025	0.015	0.013

3.4 硬件平台设计及真实场景实验

3.4.1 电力巡检机器人硬件平台设计

本章基于 Husky 机器人平台构建电力巡检机器人进行潜在动态物体语义分割以及动态物体剔除区域背景修复测试实验，Husky 机器人平台参数如表 3-2 所示。本章的实验是通过视觉传感器来采集外界数据，本章采用 Intel RealSenceD435i 深度摄像机作为视觉传感器，D435i 深度摄像机如图 3-14 所示。实验中需要处理动态物体以及语义分割等相关方面的数据，故使用分段主从机的模式来进行测试实验，其中主机为个人电脑，从机为加在 Husky 机器人平台上的外接可充电显示器，外接可充电显示器配置参数如表 3-3 所示，改装的 Husky 电力巡检机器人平台如图 3-15 所示。

表 3-2 Husky 机器人平台参数

参数名称	参数信息
重量	50 千克（110 磅）
最大有效负载	75 千克（165 磅）
全地形有效负载	20 千克（44 磅）
最大速度	1 米/秒 （3.3 英尺/秒）
离地间隙	130 毫米（5 英寸）
爬坡	45°
横越	30°
工作环境温度	-10~30℃
工作时间	正常运行 3 小时，待机 8 小时
电池	24V，20Ah（密封铅酸电池）
充电器	短路、过流、过压保护
充电时间	10 小时
用户电源	5V / 12V / 24V
通讯	RS-232 115200 波特
内部感应	电池状态、车轮里程计、电机电流

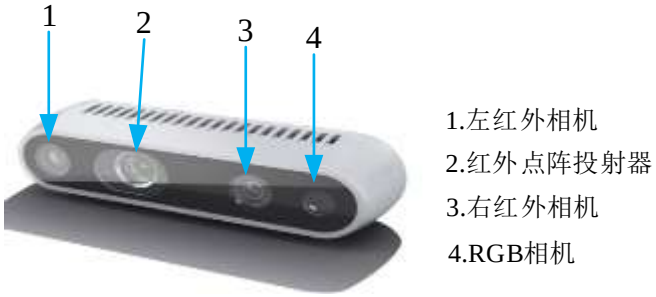


图 3-14 D435i 深度摄像机

表 3-3 可充电显示器配置

硬件名称	参数信息
CPU	Intel Core i5-630OHQ
GPU	NVDIA GeForce GTX 960
RAM	32G
操作系统	Ubuntu 18.o4

图 3-15 Husky 电力巡检机器人平台

3.4.2 系统软件设计

根据图 3-2 中所展示的本章技术路线图，并根据各部分实现的功能不同制定如图 3-16 的软件系统框架。

首先通过 D435i 相机采集彩色和深度图像帧，并通过 ROS 节点的形式将其发送至对应的订阅节点进行分析与处理。将经过处理的彩色图像帧输入语义分割线程中，对其进行各类物体类别的标记，并将形成的语义分割掩码图像帧储存在个人主机中。同时对彩色图像帧进行特征点的提取，调用个人主机中储存的掩码图，并利用多视图几何联合掩码图判断剔除动态物体。在修复线程中利用本章提出的最优像素算法进行修复，最终完成电力巡检机器人视觉 SLAM 算法的所有任务。

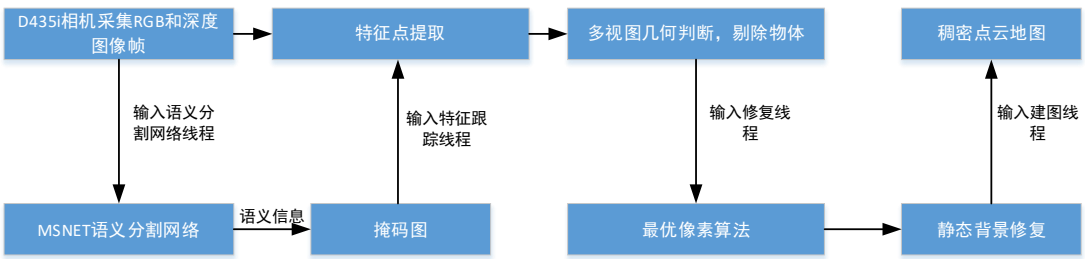
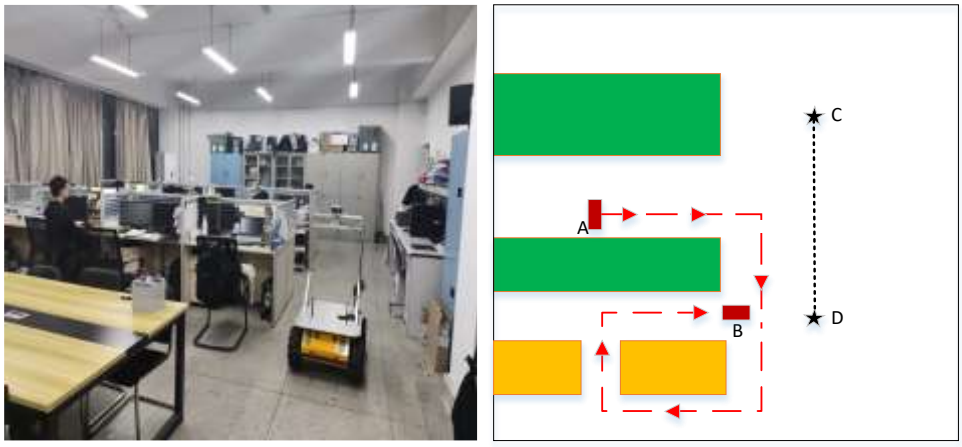


图 3-16 软件系统框架

3.4.3 真实场景实验

利用图 3-16 所设计的系统进行从一个配电房到另一个配电房巡检任务的验证，由于真实环境中配电房所处位置比较偏僻，且两个配电房之间的距离比较远，从一个配电房巡检到另一个配电房的过程中会出现动态障碍物，真实巡检过程中存在复杂环境，因实验室的设备条件有限，故在本章中以实验室环境作为模拟场景来进行实验验证。真实实验室场景环境如图 3-17（a）所示，大小为 10m×6m。真实场景几何平面布局图如图 3-17（b）所示，其中绿色和黄色平面长方形图为实验室工作台，AB 段为电力巡检机器人运行路线，CD 为巡检路上遇到的障碍行人路线。



（a）真实环境　　（b）真实场景布局

图 3-17 真实实验环境场景

电力巡检机器人运行过程中某一帧图像如图 3-18 所示。其中图 3-18（a）为原始图像帧，图 3-18（b）为语义分割效果图。从图中可以看出，本章所提算法 MSNET 在动态场景中可对潜在动态物体准确分割。 DSB-SLAM 算法背景修复图如 3-18（c）所示，其中红色框为辅助标记，框内为修复的动态物体区域。本章算法与 ORB-SLAM2、DynaSLAM 算法与真实轨迹的对比如图 3-19（d）所示。由于 ORB-SLAM2 算法无法在具有动态物体的场景下运行，其轨迹与真实轨迹相比误差较大。DynaSLAM 算法加入了 MASK-RCNN 算法对动态物体进行识别，因此其轨迹误差相对于 ORB-SLAM2 算法有所降低，但仍比本章所提出的算法轨迹误差大。由于本章算法增加了语义分割网络，识别剔除动态物体并进行静态背景修复，使得 DSB-SLAM 算法的轨迹误差最小，非常接近真实轨迹，这表明了本章所提算法在真实场景中有极大可靠性和鲁棒性，电力巡检机器人视觉

SLAM 算法能够在真实场景中稳定运行。

（a）原始图像帧　　（b）语义分割效果图

（c）背景修复图　　（d）轨迹图

图 3-18 电力巡检机器人运行过程中某一帧图像

ORB-SLAM2、DynaSLAM 和 DSB-SLAM 三种算法在 TUM 序列中和真实场景中的运行时间对比如图 3-19 所示。从图中可以看出，与 DynaSLAM 算法相比，本章算法的运行时间减少了约 20%。由于增加了语义分割网络和动态目标的消除，该算法的运行时间较 ORB-SLAM2 有所增加，但仍能有效地完成动态物体的识别和静态背景的修复。

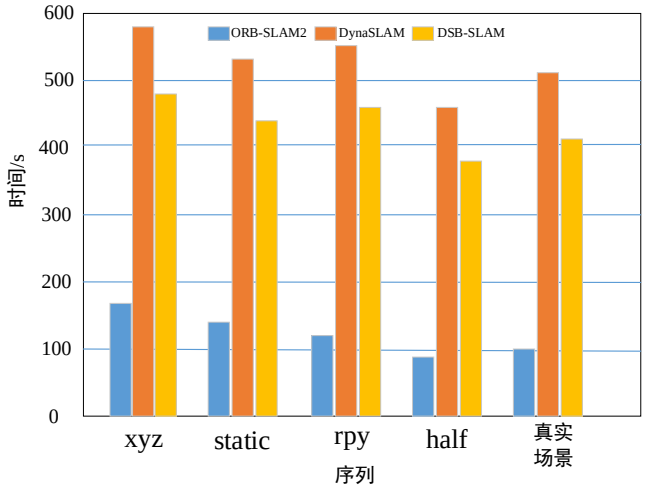


图 3-19 三种算法在 TUM 序列中和真实场景中的运行时间对比

3.5 本章小结

针对传统视觉 SLAM 算法在动态场景下难以对动态物体进行准确识别以及剔除动态物体后无法有效修复剔除区域等问题，本章提出了一种 DSB-SLAM 算法，该算法具有以下优点。针对现有语义分割网络对动态物体进行分割时由于物体的移动造成分割模糊或分割错误等问题，提出一种多类特征增强和多类特征引导的语义分割网络 MSNET，提高动态物体的语义分割能力。采用最优近邻像素匹配修复补全剔除区域的图像，为系统定位提供更多特征点信息。在公开数据集 TUM 和真实场景中进行验证，实验结果表明本文算法均方根误差与 ORB-SLAM2 和 DynaSLAM 算法相比分别减少了 95.6%、17.1%，表现出较好的定位能力，同时使得电力巡检机器人能够稳定的在动态环境中运行。

第 4 章 基于区域映射深度优化的动态多特征 RGBD-SLAM 算法

由于第 3 章中剔除动态物体后视觉 SLAM 系统只进行单一特征点的提取，使得系统无法快速高效的获取丰富特征信息进行下一线程任务，进而影响电力巡检机器人视觉 SLAM 算法的高效工作。因此，本章在第 3 章的基础上，提出一种基于区域映射深度优化的动态多特征 RGBD-SLAM 算法(Improved Dynamic Multi-Feature RGBD-SLAM Algorithm Based on Region Mapping Depth Optimization，DR-SLAM)。针对 RGBD-SLAM 在特征提取环节，深度图易出现深度不连续、单一特征点提取特征信息不足以及突然出现动态物体造成轨迹误差增大的问题，利用轻量级语义分割网络获取 RGB 帧中语义信息，将得到的语义掩码映射到其对应的深度图像中，并在掩码映射后的区域内进行近邻修复以完成深度图的优化。为减少动态物体对 SLAM 系统轨迹精度的影响，通过迭代最近点求解相机位姿，结合多视图几何得到物体位姿、运动估计矩阵以及动态视觉误差，进而估计物体运动状态并剔除动态物体。根据双向映射背景修复模型补全剔除区域的静态信息，最后进行点线面特征的提取以完成系统稳定高效运行的任务。本章技术路线如图 4-1 所示。

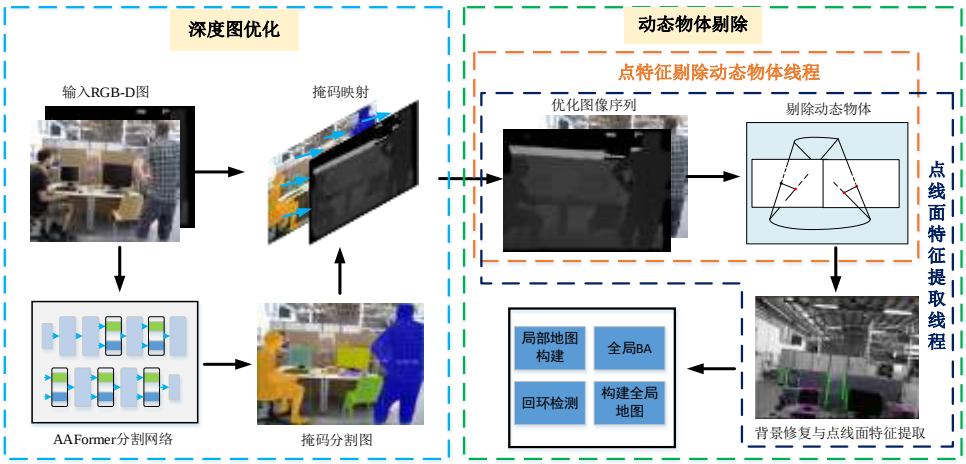


图 4-1 本章技术路线

4.1 深度优化

良好的深度图像是特征信息准确跟踪的前提。在 RGBD-SLAM 系统中，使

用 D435i 相机，用红外结构光进行深度测量时，容易受到环境和设备的影响。将特征根据深度值和相机内参投影至相机坐标系时，会出现由于深度不连续导致投影不准确的情况，进而降低帧与帧之间的特征跟踪精度。因此，高质量的深度图像是提高 SLAM 系统性能的重要因素。为了提高深度图像的准确度，本章算法采用 AFFormer 语义分割网络。AFFormer 语义分割网络是使用 Transformer 架构的新型语义分割网络，具有全视野、高效准确的优点。通过实时轻量级的网络对 RGB 图像进行检测，得到图像中物体的类别集合，如公式（4-1）所示：

$$C=\left\{Clx_1,Clx_2\cdots Clx_n\right\} \tag{4-1}$$

其中，Clx_i 表示的第 i 个类别，n 表示算法中检测到当前帧中的物体类别总数，C 是类别集合。

对于每个 $Clx_i\in C$，网络对其进行特征纹理的提取，得到各种特征纹理的集合，如公式（4-2）所示：

$$Clx_i=\left\{R_{depth},ClV\cdots\right\} \tag{4-2}$$

其中，R_{depth} 表示的是 RGB 图像深度特征纹理信息，ClV 表示的是颜色值信息，以及一些其他的特征信息。

AFFormer 语义分割网络对 RGB 图像中的类别进行标定，使得 Clx_i 在 RGB 图中生成对应的掩码 Mx_i。本章将掩码映射到深度图上对应区域，此时 RGB 图像中所有类别的深度特征纹理信息通过掩码映射输送到对应的深度图像区域，得到单个类别对应的优化区域，数学表达式如公式（4-3）和公式（4-4）所示：

$$\sum_{i=1}^{n}\left\{Mx_i\mid R_{depth}\right\}\rightarrow\sum_{i=1}^{n}D_{depth} \tag{4-3}$$

$$R_i=Mx_i\cdot D_{depth} \tag{4-4}$$

其中，D_{depth} 表示深度图像，R_i 表示单个类别对应的优化区域。

通常认为由深度相机生成的深度图像同一类别的深度值相差不大。为了更好地解决物体边界处出现深度不连续现象，本章采用物体内修复方法。根据 RANSAC 算法和 R_i 检测出正常深度值点以及异常深度值点并剔除掉异常深度值，最后通过近邻值替代法完成深度图的优化，具体做法如下。

1、通过深度相机得到在 R_i 的深度值 Dpv，其中 $Dpv=\{Dpv_1,Dpv_2,Dpv_3\cdots\}$，

其中包含正常深度值点和异常深度值点。

2、运用 RANSAC 算法剔除异常值。通过步骤 1）得到单个类别的所有深度值，为了得到最准确的深度值，本章选择优化区域的所有深度值作为模型的数据点。若任意两个数据点满足式（4-5），则可以认为是正常深度值点，反之为异常的深度值点，并将异常值剔除。

$$\left|Dpv_i - Dpv_j\right|_{i\neq j} \leqslant X \tag{4-5}$$

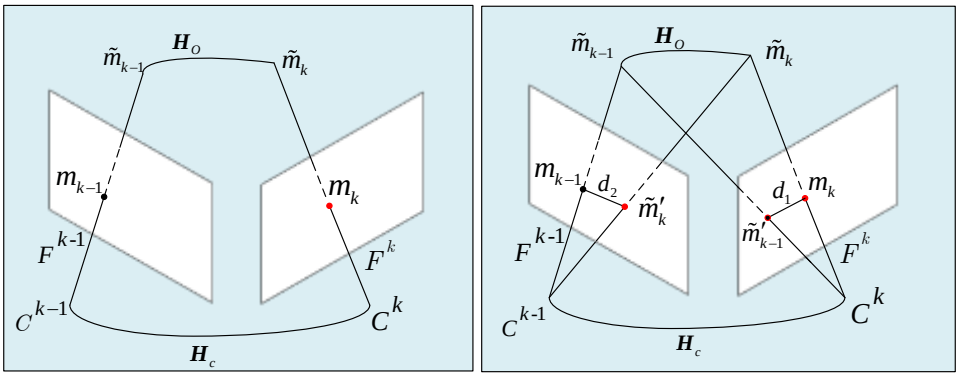
3、近邻值替代法完成深度图的优化。通过步骤 2）得到异常值点，以此点为原点四周索引，直至索引到深度正常值，用其替代异常值点，即完成深度图优化。

深度相机在不同实验场景下得到的深度值一般都是不相同的，本章经实验得出在 TUM 数据集上的阈值 X 通常取 0.8。

4.2 动态物体剔除

4.2.1 联合深度图与多视图几何剔除动态物体

动态物体对 RGBD-SLAM 算法的定位精度有着重要影响，当前大多数点线面 SLAM 系统都在静态环境中实现了高精度的定位，但当环境中闯入动态物体时极易造成轨迹误差增大问题，使得整体算法缺乏克服机器人在动态环境下稳定运行的能力，从而影响整个 SLAM 系统的定位。传统基于点到直线去除动态物体的方法难以高效去除动态物体影响，因此，本章在第三章的基础上提出一种联合深度图与多视图几何剔除动态物体的方法，剔除动态物体示意图如图 4-2 所示，其中 C^{k-1} 和 C^{k} 表示前后两帧的相机中心。



(a) 物体位姿　　(b) 物体运动估计

图 4-2 剔除动态物体示意图

SLAM 系统在运行过程中，在已知相机内参和深度图像深度值的前提下，将经过深度优化的深度图像转化为点云，计算公式见式（4-6）：

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = Dpv_i \begin{bmatrix} \frac{1}{f_x} & 0 & 0 \\ 0 & \frac{1}{f_y} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \tag{4-6}$$

其中，(x, y, z) 是点云坐标；(u, v) 是图像坐标；f_x、f_y 为 RGB-D 相机的内参。

通过式（4-6）获取当前图像帧与后一帧的点云地图点 $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2 \cdots \boldsymbol{x}_n\}$ 和 $\boldsymbol{Y} = \{\boldsymbol{y}_1, \boldsymbol{y}_2 \cdots \boldsymbol{y}_n\}$，并通过 ICP 算法得到相机前后的位姿差，从而求解出相机位姿。得到相机前后的位姿差，采用最小二乘法求解旋转矩阵 $\boldsymbol{R} \in \mathbb{R}^{3\times3}$ 和平移向量 $\boldsymbol{t} \in \mathbb{R}^3$，最终得到 $\boldsymbol{H}_c$，计算公式见式（4-7）：

$$\begin{cases} M_r = \arg\min\limits_{\boldsymbol{H}_c} \dfrac{1}{N_y} \sum\limits_{i=1}^{N_y} \left\| \boldsymbol{x}_i - \boldsymbol{R}\boldsymbol{y}_i - \boldsymbol{t} \right\|^2 \\ \boldsymbol{H}_c = \begin{bmatrix} \boldsymbol{R} & \boldsymbol{t} \\ \boldsymbol{0}_{1\times3} & 1 \end{bmatrix} = \exp(\boldsymbol{h}_c) \end{cases} \tag{4-7}$$

其中，$\boldsymbol{x}_i \in \mathbb{R}^3$ 为前一帧点云；$\boldsymbol{y}_i \in \mathbb{R}^3$ 为后一帧点云；M_r 为最小残差；$\boldsymbol{H}_c \in \mathrm{SE(3)}$ 为相机姿态相对变换矩阵；$\boldsymbol{h}_c \in \mathrm{SE(3)}$ 为相机位姿相对变换向量；N_y 为匹配的点数。

$\tilde{\boldsymbol{h}}_c \in \mathbb{R}^6$ 为从 SE(3)映射到 $\mathbb{R}^6$ 的运算，通过高斯-牛顿算法得到准确相机位姿，计算公式见式（4-8）：

$$\boldsymbol{h}_c^* = \arg\min_{\tilde{\boldsymbol{h}}_c} \sum_{i=1}^{j} \rho_h \left(\boldsymbol{e}\left(\boldsymbol{h}_c\right)^{\mathrm{T}} \boldsymbol{\Sigma}_P^{-1} \boldsymbol{e}\left(\boldsymbol{h}_c\right) \right) \tag{4-8}$$

其中，$\boldsymbol{h}_c^* \in \mathbb{R}^6$ 为最小二乘解；ρ_h 为惩罚因子；$\boldsymbol{\Sigma}_P \in \mathbb{R}^{4\times4}$ 为 $\boldsymbol{R}$、$\boldsymbol{t}$ 协方差矩阵；j 为残差运算所需 3D 点投影至 2D 点数量；$\boldsymbol{e}(\boldsymbol{h}_c) \in \mathbb{R}^{4\times4}$ 为 $\boldsymbol{h}_c$ 的分解矩阵。

由相机位姿和多视图几何确定物体位姿，将所有潜在动态物体建模为位姿变换矩阵 $\boldsymbol{H}_o \in \mathbb{R}^{4\times4}$。通过对应的 RGB 相邻帧间几何关系将空间中的点 $\tilde{m}$ 从参考帧 F^{k-1} 关联到后一帧 F^k，计算公式见式（4-9）：

$$\begin{cases} m_k = \varepsilon\left(\boldsymbol{H}_c \boldsymbol{H}_o \varepsilon^{-1}\left(I_{k-1}\right)\right) \\ I_{k-1} = \left(m_{k-1}, m_{z-1}\right) \end{cases} \tag{4-9}$$

其中，m_k 是 F^k 中的平面坐标；ε 和 ε^{-1} 分别为投影函数和反向投影函数；I_{k-1} 是空间点投影到 F^{k-1} 中的三维点；m_{k-1} 为 I_{k-1} 在 F^{k-1} 中的平面坐标；m_{z-1} 为 I_{k-1} 在 F^{k-1} 中的深度值。

多视图几何得到的物体变换矩阵存在误差，本章通过重投影误差与最小二乘法对物体姿态变换矩阵 $\boldsymbol{H}_o$ 进行优化处理。同理，通过 $\tilde{\boldsymbol{h}}_o \in \mathbb{R}^6$ 从 SE(3)到 $\mathbb{R}^6$ 的映射运算，得到准确的物体运动估计矩阵 $\boldsymbol{h}_o^*$。

$$\begin{cases} e(\boldsymbol{H}_o) = \tilde{m}'_{k-1} - \varepsilon\left(\boldsymbol{H}_c \boldsymbol{H}_o \varepsilon^{-1}\left(I_{k-1}\right)\right) \\ \boldsymbol{H}_o = \exp(\boldsymbol{h}_o) \\ \boldsymbol{h}_o^* = \arg\min\limits_{\tilde{\boldsymbol{h}}_o} \sum\limits_{i=1}^{j} \rho_h\left(\boldsymbol{e}\left(\boldsymbol{h}_o\right)^{\mathrm{T}} \boldsymbol{\Sigma}_P^{-1} \boldsymbol{e}\left(\boldsymbol{h}_o\right)\right) \end{cases} \tag{4-10}$$

其中，$e(\boldsymbol{H}_o)$ 为位姿变换矩阵优化解（最小化重投影误差）；$\boldsymbol{h}_o$ 为位姿变换向量，$\boldsymbol{e}(\boldsymbol{h}_o)$ 为 $\boldsymbol{h}_o$ 的分解矩阵；$\tilde{m}'_{k-1}$ 为 F^{k-1} 中 m_{k-1} 的坐标投影到 F^k 中的坐标。

采用平面几何映射判断物体状态，空间中的点 $\tilde{m}_{k-1}$ 投影到 F^{k-1} 帧中得到 2D 像素坐标 m_{k-1}，此时假设 m_{k-1} 通过几何关系投影到 F^k 中的点为 $\tilde{m}'_{k-1}$，与 F^k 中真实投影点 m_k 的像素距离 d_1 为动态视觉误差权重。为了减少误差，同时将 F^k 中的 2D 像素坐标 m_k 也投影到 F^{k-1} 中，与 F^{k-1} 中真实投影点 m_{k-1} 形成动态视觉误差权重 d_2。物体的动态视觉误差由像素动态视觉误差代替，为了得到准确的物体动态视觉误差，本章采用深度值类比权重法，计算公式见式（4-11）：

$$\begin{cases} d_1 = \left\{\left\|\tilde{m}'_{k-1}, m_k\right\|\right\} \\ d_2 = \left\{\left\|\tilde{m}'_k, m_{k-1}\right\|\right\} \\ \theta = \left(d_1 m_z + d_2 m_{z-1}\right) / \left(d_1 + d_2\right) \end{cases} \tag{4-11}$$

其中，m_z 为空间中的点 $\tilde{m}_k$ 到 F^k 中 m_k 的深度值；θ 为物体动态视觉误差。

姿态不确定性是判断物体是否运动的重要指标，姿态不确定性可以通过运动微分熵值导出，因此本章采用运动微分熵作为判断物体运动的依据，计算公式见式（4-12）：

$$H(\theta)=\exp(\sqrt{\frac{1}{2}\log|\boldsymbol{P}_{\theta}|+\frac{\boldsymbol{\omega}}{2}\log(\boldsymbol{\omega}\pi e)}) \tag{4-12}$$

其中，$H(\theta)$ 为运动熵；e 为自然指数；$\boldsymbol{P}_{\theta}$ 为不确定误差，服从 $\boldsymbol{\omega}$ 维高斯分布，通常 $\boldsymbol{P}_{\theta}=\left(\boldsymbol{J}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{J}\right)^{-1}$，$\boldsymbol{\Sigma}^{-1}$ 为对角矩阵，$\boldsymbol{J}$ 为雅各比矩阵。

基于此，将物体动态视觉误差与一个由运动微分熵 $f(H(\theta))$ 引导的动态阈值 $\Delta\theta=f(H(\theta))$ 进行对比，其中 $f(\cdot)$ 为求微分。若 $\theta>\Delta\theta$，则判断为动态物体并将其剔除。动态物体剔除方法如算法 1 所示。

算法 1 动态物体剔除方法

输入：连续帧图像 F_k，优化图
输出：图像中动态物体与静态物体
参数：$\Delta\theta$
FOR new F_k available THEN
  图像物体类别集合
  获取图像特征点
    FOR 图像中的地图点 THEN
      物体位姿与运动估计，计算物体动态视觉误差
      $\theta=\left(d_1 m_z+d_2 m_{z-1}\right)/\left(d_1+d_2\right)$
      计算姿态不确定性
      $\Delta\theta=H(x_o)=\exp(\sqrt{\frac{1}{2}\log|\Sigma x_o|+\frac{\omega}{2}\log(\omega\pi e)})$
      IF $\theta>\Delta\theta$
        特征点为动态物体
      ELSE
        特征点为静态物体
      END IF
    END FOR
END FOR

4.2.2 背景修复与点线面特征提取

本章提出一种基于双向映射模型的剔除动态物体区域的修复方法，借助前后对应帧中的静态图像特征信息修复剔除区域彩色图像和深度图背景。完成静态背景修复后，再进行点线面特征的提取以完成电力巡检机器人视觉 SLAM 算法的定位任务。

基于双向映射的背景修复示意图如图 4-3 所示。以待修复关键帧 F_t 为起始点，向着褐绿色箭头指向方向移动 8 帧的关键图像帧窗口。将窗口内的关键帧图像依次与待修复关键帧图像对齐，沿着黑色虚线箭头方向索引所对应的像素填充

到待修复关键帧对应区域中。

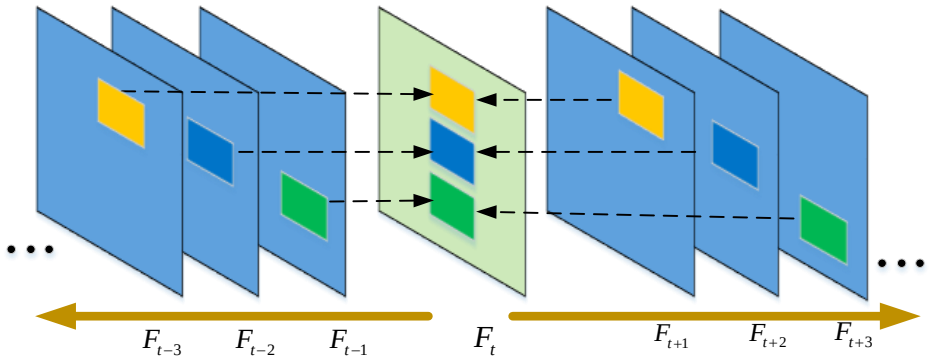


图 4-3 基于双向映射的背景修复示意图

4.3 公开数据集实验及分析

本章实验所用电脑设备为 3.3 节中的设备。为了验证本章算法的有效性，使用 TUM 公开数据集进行验证。

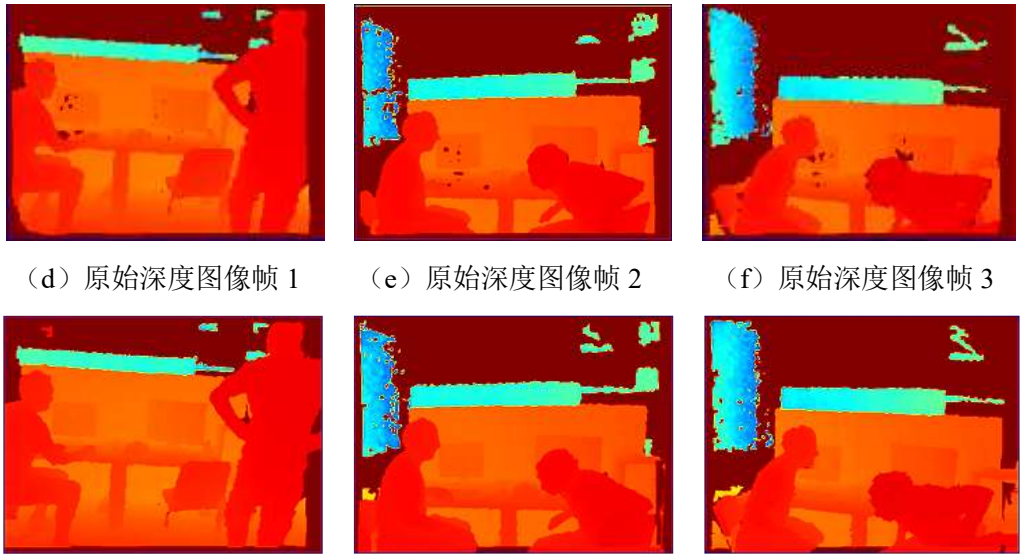
4.3.1 深度优化实验及分析

深度优化效果如图 4-4 所示。其中图 4-4（a）~（c）为 TUM 数据集中三帧原始彩色图像帧，图 4-4（d）~（f）为三帧经过图像对比度处理的原始深度图像帧，图 4-4（g）~（i）为三帧采用本章算法优化后得到的深度图像帧。由于深度相机在运行过程中易受环境中噪音影响，使产生的深度图像存在深度缺失，尤其是在动态场景及两个物体的连接处。从图 4-4（g）~（i）可以看出，采用本章算法优化得到的深度图像中存在的黑洞值大多数都被修复，而且从图 4-4（d）和图 4-4（g）中能够看出本章算法能够粗略重现深度图像中一部分深度完全丢失的地方，如图 4-4（a）中的手臂和后背就被重现出来，深度优化算法提高了特征跟踪的精度，为后续动态物体剔除提供了良好的深度依据。

（a）原始彩色图像帧 1

（b）原始彩色图像帧 2

（c）原始彩色图像帧 3



（d）原始深度图像帧 1　（e）原始深度图像帧 2　（f）原始深度图像帧 3

（g） 优化深度图像帧 1　（h）优化深度图像帧 2　（i）优化深度图像帧 3

图 4-4 深度优化效果

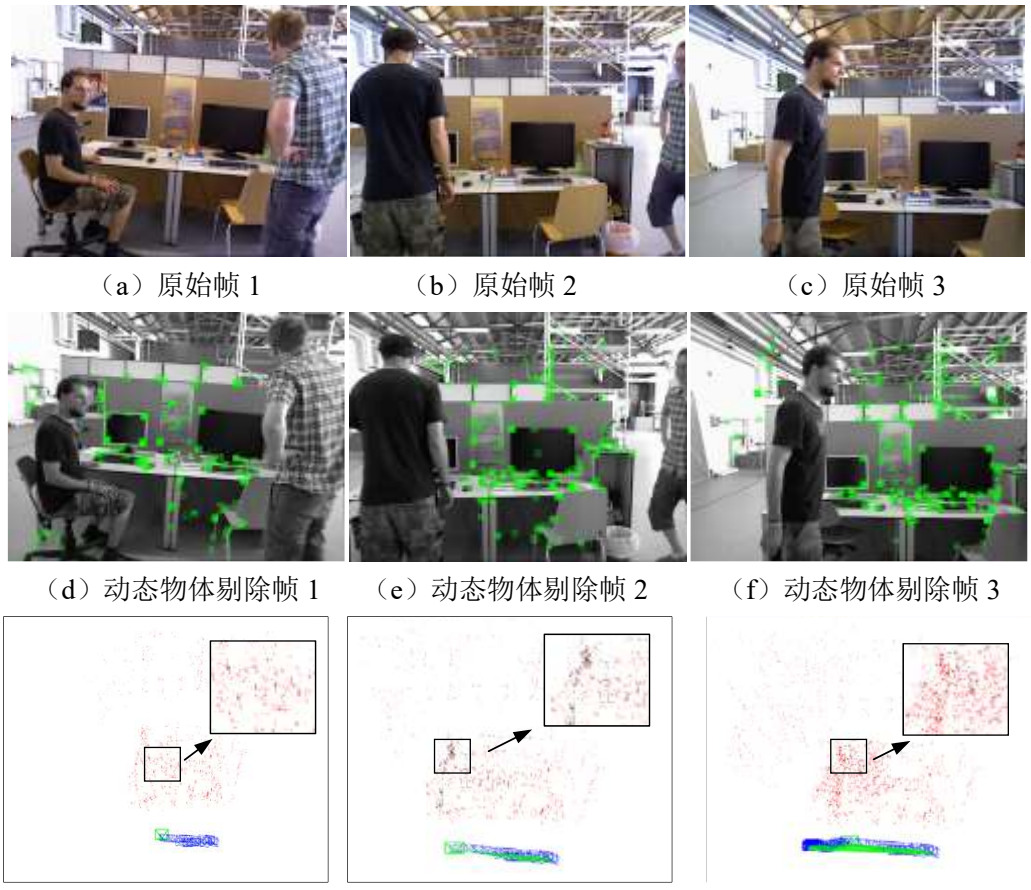
由于 RGB-D 相机自身的原因，得到的深度图会出现深度不连续的问题，影响特征跟踪精度。当特征点根据深度值和相机内参投影至相机坐标系时，出现了由于深度不连续影响投影精度的情况，进而降低帧与帧之间的特征跟踪的准确率。为了体现深度优化算法与 SLAM 系统的联系以及在 SLAM 线程中的优势，本章采用特征跟踪准确率（Feature Tracking Accuracy）作为深度优化算法的评价指标。特征跟踪准确率指的是当前帧中提取的特征点投影到下一帧的成功率，特征跟踪的准确率越高表示跟踪精度越高。本章在 TUM 数据集上选取了四个高动态序列 fr3_walking_xyz(xyz)、fr3_walking_static(static)、fr3_walking_rpy(rpy)、fr3_walking_halfsphere(half)以及两个无动态序列 fr2_xyz 和 fr2_rpy，并将每个序列图像帧数据的中位数（Median）与平均值（Mean）作为实验数据指标。ORB-SLAM3、RGBD-SLAM、不加优化的本章算法（DR-SLAM without Optimization, DR-SLAM-WO）和本章 DR-SLAM 算法在 TUM 数据集下的特征跟踪准确率如表 4-1 所示，表中加粗黑色字体表示最优值。从表 4-1 可以看出，在四个高动态序列中，DR-SLAM 的特征跟踪准确率是最高的，DR-SLAM-WO 的特征跟踪准确率是次高的；在两个无动态序列中，DR-SLAM 的特征跟踪准确率与 ORB-SLAM3 几乎相同且都高于 DR-SLAM-WO 和 RGBD-SLAM，验证了本章深度优化算法能够提高动态 SLAM 系统特征跟踪精度，从而提升动态 SLAM 系统的位姿估计准确率，为减少轨迹误差提供良好基础。

表 4-1 四种算法在 TUM 数据集下的特征跟踪准确率（单位：%）

TUM 序列	ORB-SLAM3		RGBD-SLAM		DR-SLAM-WO		DR-SLAM	
	Median	Mean	Median	Mean	Median	Mean	Median	Mean
xyz	79.7	80.1	78.8	79.2	82.3	82.7	**85.1**	**84.9**
static	80.5	81.3	79.6	79.9	81.5	82.0	**83.9**	**84.3**
rpy	81.3	81.4	80.1	79.7	84.8	84.5	**90.8**	**91.1**
half	81.3	81.9	80.8	81.0	82.3	82.9	**87.1**	**87.3**
fr2_xyz	92.6	**93.5**	90.6	91.3	90.1	90.6	**92.6**	93.4
fr2_rpy	**91.8**	91.9	91.3	91.5	91.2	91.3	91.7	**92.1**

4.3.2 点特征剔除动态物体效果图

点特征动态物体剔除效果如图 4-5 所示，其中图4-5（a）~（c）为在 TUM 数据集上任意选取的三个动态原始帧；图 4-5（d）~（f）为三个动态物体剔除图像帧，从图中可以看出，DR-SLAM 在动态物体场景下能有效剔除动态物体；图 4-5（g）~（i）为与动态物体剔除帧对应的稀疏地图点帧。如图 4-5（h）所示，其中红色点为静态特征点，黑色点为本章标记的动态特征点。随着动态物体的减少，黑色点的数量也逐渐减少，验证了 DR-SLAM 能够有效区分动态特征点，为后续的多特征信息提取减少了动态物体的影响。



（a）原始帧 1　（b）原始帧 2　（c）原始帧 3

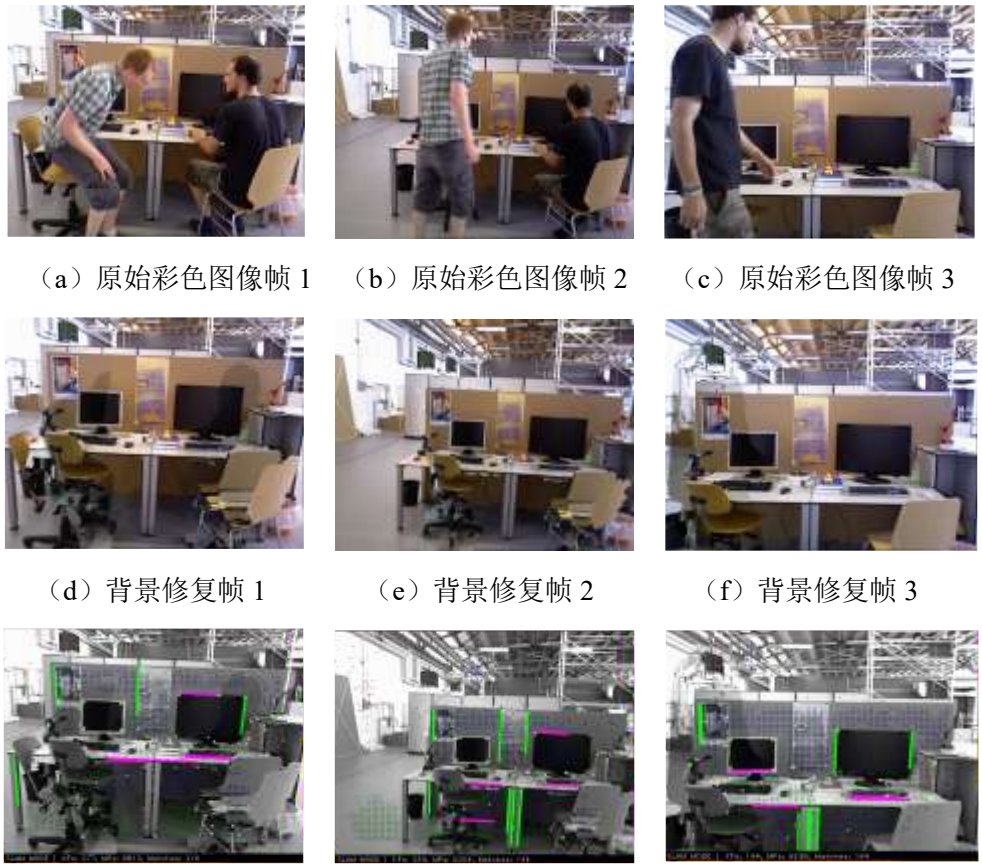
（d）动态物体剔除帧 1　（e）动态物体剔除帧 2　（f）动态物体剔除帧 3

（g）稀疏地图点帧 1　（h）稀疏地图点帧 2　（i）稀疏地图点帧 3

图 4-5 点特征动态物体剔除效果

4.3.3 背景修复与点线面特征提取

背景修复与点线面特征提取效果如图 4-6 所示，其中图 4-6（a）~（c）为在 TUM 数据集上任意选取的三个动态原始彩色图像帧；图 4-6（d）~（f）为三个原始图像帧背景修复帧，可以看出双向映射模型能有效地进行剔除区域静态背景的修复；图 4-6（g）~（i）为静态背景修复后的点线面特征提取帧。本章通过剔除动态物体后进行背景修复，并进行点线面特征的提取，能够有效降低动态物体对 SLAM 系统的的影响，提供了良好的位姿估计精度，从而减少轨迹误差，进而提高电力巡检机器人视觉 SLAM 算法的定位精度。



（a）原始彩色图像帧 1　（b）原始彩色图像帧 2　（c）原始彩色图像帧 3

（d）背景修复帧 1　（e）背景修复帧 2　（f）背景修复帧 3

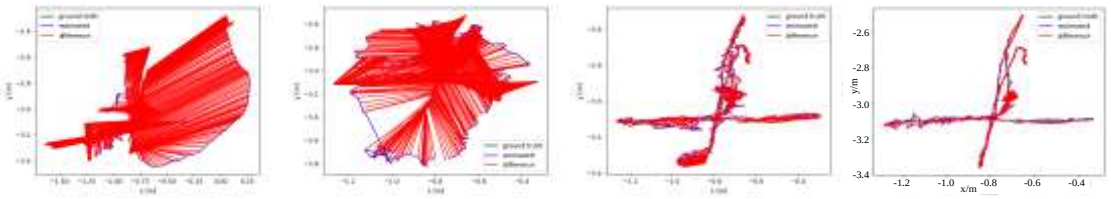
（g）点线面特征提取帧 1　（h）点线面特征提取帧 2　（i）点线面特征提取帧 3

图 4-6 背景修复与点线面特征提取效果

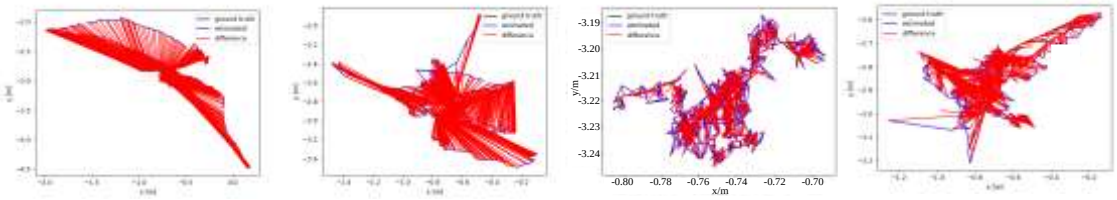
4.3.4 轨迹图对比实验和定位精度

为了验证 DR-SLAM 剔除动态物体的能力，本章选取 TUM 数据集的三个动态序列 fr3_walking_xyz（xyz：相机原地左右上下转动拍摄的数据集）、fr3_walking_rpy（rpy：相机上下左右移动拍摄的数据集）、fr3_walking_halfsphere

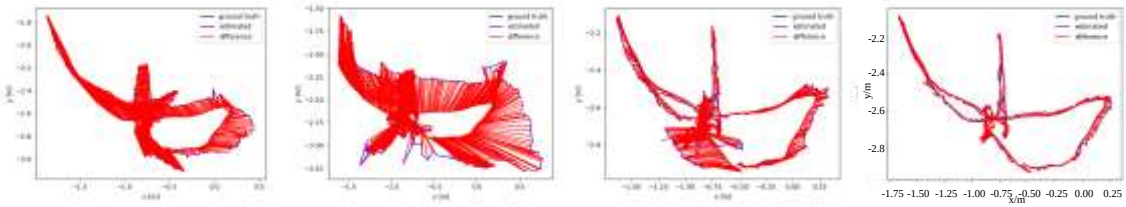
（half：相机左右上下以及翻转拍摄的数据集）进行验证评估。ORB-SLAM3、RGBD-SLAM、DS-SLAM 和 DR-SLAM 四种算法在这三个序列上运行的轨迹对比如图 4-7 所示。ORB-SLAM3 和 RGBD-SLAM 为静态环境中的 SLAM 算法，DS-SLAM 和 DR-SLAM 为动态环境中的 SLAM 算法。从图 4-7（a）和图 4-7（b）可以看出，由于ORB-SLAM3 和RGBD-SLAM 无法识别环境中存在的动态物体，因此所生成的轨迹与真实轨迹相比存在较大的轨迹误差；从图 4-7（d）可以看出，由于 DR-SLAM 减小了深度不连续对动态物体的影响，所以 DR-SLAM 生成的轨迹误差小于其他三种算法。图 4-7（e）~（l）为四种算法在另两组动态序列的轨迹对比，从图中可以看出 ORB-SLAM3、RGBD-SLAM 和 DS-SLAM 算法所生成的轨迹误差依旧较大，而 DR-SLAM 所生成的轨迹误差还是小于其他三种算法，验证了 DR-SLAM 对于环境中的动态物体具有良好地剔除能力，能够有效减少轨迹误差，提高 SLAM 系统的定位精度，使得电力巡检机器人能够在动态环境中快速高效的稳定运行。



（a）ORB-SLAM3 xyz（b）RGBD-SLAM xyz（c）DS-SLAM xyz（d）DR-SLAM xyz



（e）ORB-SLAM3 rpy （f）RGBD-SLAM rpy （g）DS-SLAM rpy （h）DR-SLAM rpy



（i）ORB-SLAM3 half （j）RGBD-SLAM half （k）DS-SLAM half （l）DR-SLAM half

图 4-7 三个序列上运行的轨迹对比

为验证算法的性能，实验采用绝对位姿误差（Absolute Pose Error，APE）和相对位姿误差（Relative Pose Error，RPE）作为评价指标。其中 APE 是比较估计轨迹和参考轨迹并计算整个轨迹的统计数据，适用于测试轨迹的全局一致性；RPE 是相邻两帧之间的估计值与真实值的差异。

在上述实验的三个序列上增加 fr3_walking_static（static：相机拍摄场景中的人物是静止的，没有动态物体）序列来评测算法的定位与建图能力。为了消除偶然误差，本章在每个序列上进行 5 次实验，并且取均方根误差（Root Mean Square Error, RMSE）、平均值（Mean）、标准差（Standard Deviation, SD）作为衡量单位。ORB-SLAM3、RGBD-SLAM 和 DS-SLAM 和本章 DR-SLAM 四种算法的绝对轨迹误差（Absolute Trajectory Error, ATE）如表 4-2 所示，表中加粗黑体表示最优数值。

表 4-2 四种算法在 TUM 数据集下的 ATE 对比（单位：米）

TUM 序列	ORB-SLAM3			RGBD-SLAM			DS-SLAM			DR-SLAM		
	RMSE	Mean	SD	RMSE	Mean	SD	RMSE	Mean	SD	RMSE	Mean	SD
xyz	0.753	0.586	0.469	0.448	0.421	0.152	0.078	0.057	0.048	**0.056**	**0.045**	**0.033**
static	0.410	0.310	0.268	0.425	0.363	0.220	0.056	0.045	0.037	**0.035**	**0.032**	**0.015**
rpy	0.986	0.837	0.522	0.672	0.616	0.268	0.241	0.332	0.115	**0.233**	**0.205**	**0.111**
half	0.287	0.278	0.073	0.717	0.668	0.261	0.139	0.113	0.107	**0.137**	**0.098**	**0.095**
平均值	0.609	0.503	0.333	0.566	0.517	0.225	0.129	0.137	0.077	**0.115**	**0.095**	**0.064**

由于 ORB-SLAM3、RGBD-SLAM 和 DS-SLAM 三种算法都是直接通过深度相机获得深度图，存在深度值丢失的问题，易出现跟踪精度下降，导致无法准确剔除动态点，从而无法有效完整地剔除动态物体。而 DR-SLAM 通过深度图优化以及多视图几何双向投影进行动态物体剔除并进行背景修复，既避免了深度黑洞值的影响又保留了大量的特征信息，该算法四个序列的所有衡量指标都达到最好。通常 SLAM 系统中 RMSE 指标最能反映轨迹精度优劣，本章选取 RMSE 指标做定量分析。在 xyz 序列上，DR-SLAM 相较于 ORB-SLAM3、RGBD-SLAM 以及 DS-SLAM 的绝对轨迹误差 RMSE 分别减少 92.6%、87.5%和 28.2%；在 static 序列上，DR-SLAM 相较于 ORB-SLAM3、RGBD-SLAM 以及 DS-SLAM 的绝对轨迹误差 RMSE 分别减少 91.5%、91.8%和 37.5%；在 rpy 序列上，DR-SLAM 相较于 ORB-SLAM3、RGBD-SLAM 以及 DS-SLAM 的绝对轨迹误差 RMSE 分别减少 76.4%、65.3%和 3.3%；在 half 序列上，DR-SLAM 相较于 ORB-SLAM3、RGBD-SLAM 以及 DS-SLAM 的绝对轨迹误差 RMSE 分别减少 52.3%、80.9%和

1.4%。总的来说，DR-SLAM 相较于 ORB-SLAM3、RGBD-SLAM 以及 DS-SLAM 的平均绝对轨迹误差 RMSE 分别减少 78.2%、81.4%和 17.6%，验证了本章动态剔除算法能够有效剔除环境中的动态物体，为 SLAM 系统的定位提供了良好基础。

ORB-SLAM3、RGBD-SLAM、DS-SLAM 和 DR-SLAM 四种算法在 TUM 数据集运行时间的对比如图 4-8 所示。从图 4-8 可以看出，四种算法运行时间的平均值分别为 105s、101s、537s 和 458s。由于 DR-SLAM 减少了深度不连续对 SLAM 系统的影响，使 DR-SLAM 运行时间平均值与 DS-SLAM 相比减少了 15%；与 RGBD-SLAM 和 ORB-SLAM3 相比，由于 DR-SLAM 增加了语义分割网络映射对深度图的优化以及动态物体的剔除，使得 DR-SLAM 运行时间较 RGBD-SLAM 有较大的增加，但仍可满足 SLAM 系统正常运行需求。

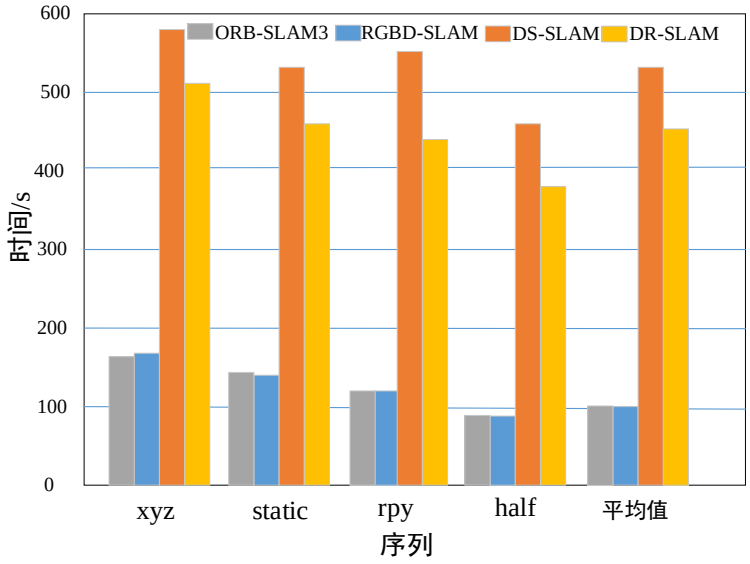


图 4-8 四种算法在 TUM 数据集下运行时间对比

4.4 真实场景实验

4.4.1 系统软件设计

本章节的电力巡检机器人硬件设计采用 3.4.1 节中的设计方案，系统软件设计根据图 4-1 中所展示的本章技术路线图，并根据各部分实现的功能不同制定如图 4-9 的软件系统框架。

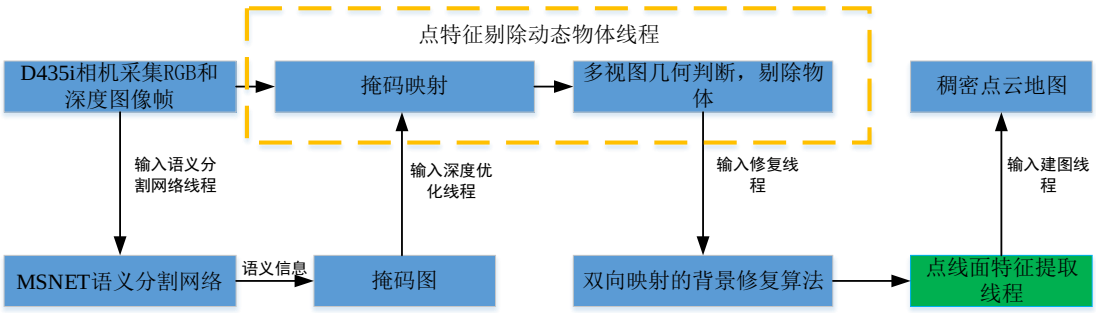
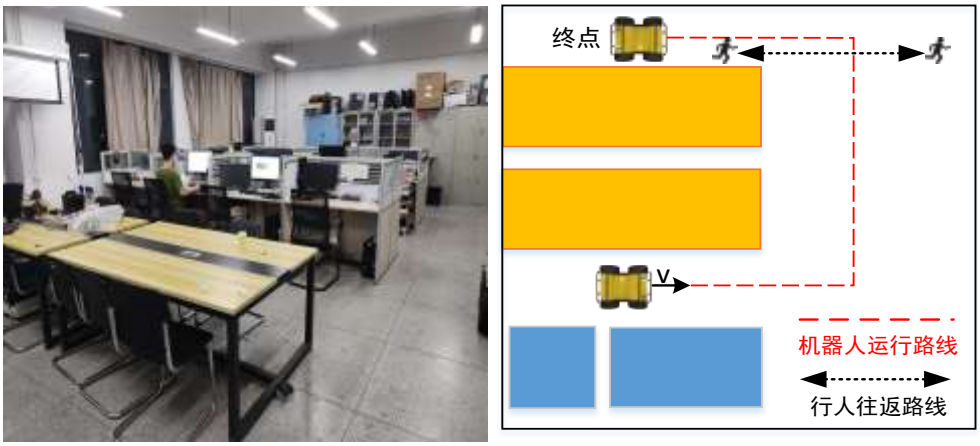


图 4-9 的软件系统框架

首先通过 D435i 相机采集彩色和深度图像帧，并通过 ROS 节点的形式将其发送至对应的订阅节点进行分析与处理。将经过处理的彩色图像帧输入语义分割线程中，对其进行各类物体类别的标记，并将形成的语义分割掩码图像帧储存在个人主机中。点特征剔除动态物体线程包含深度优化线程，其中通过本章提出的掩码映射法来进行深度优化，并联合多视图几何剔除动态物体。在修复线程中利用本章提出的双向映射的背景修复算法进行修复，并进入点线面特征提取线程进行特征信息的提取，最终完成 SLAM 的所有任务。

4.4.2 真实场景实验

模拟的巡检环境如图 4-10 所示，真实环境如图10（a）所示，场地大小为10m×6m。真实场景布局如图 10（b）所示，黄色和蓝色部分为实验室工作台。



（a）真实环境　　（b）真实场景布局

图 4-10 模拟的巡检环境

真实场景下 RGBD-SLAM 算法与 DR-SLAM 算法的定位轨迹对比如图 4-11 所示。本章算法是针对 RGBD-SLAM 的问题进行研究的，而 ORB-SLAM3 是在静态场景中运行，无法有效地在动态场景中工作，又因为 DS-SLAM 的分割算法

只能分割简单的物体，无法在实验室复杂的环境中运行，故本章只与 RGBD-SLAM 作对比。DR-SLAM 与 RGBD-SLAM 的绝对轨迹误差分别为 0.078m 和 0.635m，由于 RGBD-SLAM 算法无法有效剔除动态物体，导致算法运行轨迹误差较大，而本章算法加入深度优化修复深度不连续以及联合多视图几何剔除动态物体，提高了特征匹配精度，减小了轨迹误差，因此本章算法在动态场景下能稳定运行，具有较高的鲁棒性，使得电力巡检机器人能够稳定高效的在动态环境中稳定运行。

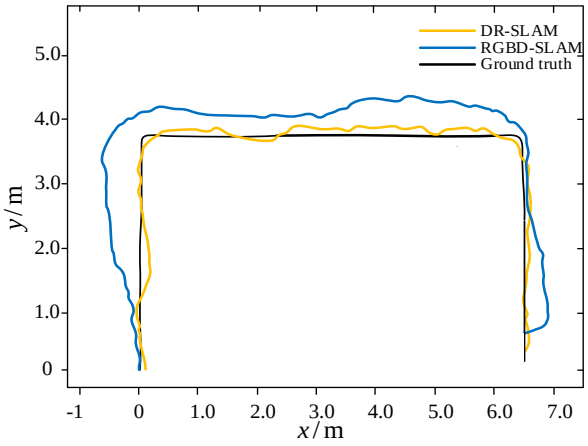


图 4-11 轨迹对比

4.5 本章小结

本章提出了一种 DR-SLAM 算法，该算法具有以下创新点：提出一种基于语义信息的深度图优化，改善了 RGBD-SLAM 在特征提取环节由于深度相机得到的深度图易出现深度不连续影响特征跟踪的问题，提高了 SLAM 系统的位姿估计精度。针对突然出现动态物体造成轨迹误差增大的问题，提出一种联合深度信息与多视图几何完成动态物体剔除的方法，利用双向映射模型进行静态背景的修复，并进行点线面特征提取，提高了定位精度减少轨迹误差。本章算法在公开数据集 TUM 上进行测试，结果表明，由于 DR-SLAM 采用区域映射深度优化获取了良好的深度信息，以及联合多视图几何剔除动态物体和背景修复，使得本章算法与 RGBD-SLAM 和 DS-SLAM 算法相比，在位姿和轨迹精度方面具有良好的优势，体现了良好的抗干扰能力与动态物体剔除能力，使得电力巡检机器人视觉 SLAM 算法能够稳定高效的处理巡检过程中遭遇的动态物体，完成巡检任务。

第 5 章 全文总结展望

5.1 全文总结

首先介绍了本文的背景及意义，引出巡检机器人的作用，然后介绍电力巡检机器人的分类，介绍电力巡检机器人的核心技术视觉 SLAM 研究现状，特别是视觉 SLAM 和 RGBD-SLAM 技术的研究热点内容，最后重点介绍了动态场景下视觉 SLAM 和 RGBD-SLAM 算法的定位与构图技术。在第二章中首先分别对视觉 SLAM 系统的视觉里程计、后端优化、回环检测和建图以及传统深度优化几个部分相关基础知识叙述，然后对 Vision Transformer 的图像块的处理以及工作原理进行分析。在第三章和第四章中，针对传统视觉 SLAM 算法难以对动态物体进行准确识别以及剔除动态物体后无法有效修复剔除区域等问题，和针对 RGBD-SLAM 在特征提取环节，深度图易出现深度不连续、单一特征点提取特征信息不足以及突然出现动态物体造成轨迹误差增大的问题。提出了基于语义分割和背景修复的动态视觉 SLAM 算法和基于区域映射深度优化的动态多特征 RGBD-SLAM 算法，最后在公开数据集 TUM 上进行实验，并搭建硬件平台验证本文所提算法在真实场景中的可行性。

论文主要研究内容与创新点概括如下：

（1）本文针对传统视觉 SLAM 算法在动态场景下难以对动态物体进行准确识别以及剔除动态物体后无法快速高效的修复等问题，改进 Vision Transformer 语义分割网络结合自适应几何阈值和背景修复提出 DSB-SLAM 算法。该算法包含潜在动态物体分割和背景修复两个环节。潜在动态物体分割环节通过第三章所提的 MSNET 语义分割网络对图像帧中存在的物体进行语义分割，并提取特征点。在背景修复环节，通过自适应几何阈值剔除动态物体，利用最优近邻像素匹配进行背景修复，并进行特征点的提取，为 SLAM 系统的定位提供丰富信息，使得电力巡检机器人视觉 SLAM 算法可以在动态环境中稳定运行。

（2）针对 RGBD-SLAM 在特征提取环节，深度图易出现深度不连续、单一特征点提取特征信息不足以及突然出现动态物体造成轨迹误差增大的问题，本文提出基于区域映射深度优化的动态多特征 RGBD-SLAM 算法。该算法包含深度

图优化和动态物体剔除两个环节。深度图优化环节通过轻量级的 AFFormer 网络提取语义信息，并将提取的语义信息映射到其对应的深度图上进行深度修复。动态物体剔除环节包含点特征剔除动态物体线程和点线面特征提取线程。首先点特征联合多视图几何完成动态物体的剔除，并利用双向映射背景修复模型补全剔除区域的静态信息，最后进行点线面特征的提取，为电力巡检机器人视觉 SLAM 算法稳定高效运行提高了良好的保障，从而完成巡检任务。

（3）验证本文所提算法的可行性，搭建电力巡检机器人硬件实验平台，并通过实验室模拟场景进行实验，模拟电力巡检机器人从一个配电房巡检到另一个配电房的过程，实验结果表明本文所提的两种算法与其它主流的动态 SLAM 算法相比具有良好的轨迹精度，系统的定位精度高，且抗干扰能力强，使得电力巡检机器人视觉 SLAM 算法能稳定高效的在动态环境中运行。

5.2 展望

本文通过对电力巡检机器人的核心技术视觉 SLAM 算法进行分析研究，根据已有的视觉 SLAM 和 RGBD-SLAM 技术进行改进并提出基于语义分割和背景修复的动态视觉 SLAM 算法和基于区域映射深度优化的动态多特征 RGBD-SLAM 算法两种算法。为验证所提两种算法的真实性与有效性，本文通过公开数据集以及搭建硬件平台在模拟场景中进行验证，实验验证了本文所提两种算法的可行性，并为电力巡检机器人视觉 SLAM 算法下一步发展提供了思路。但在研究的过程中出现了一些新的亟待解决的问题。

（1）相似结构化场景中的定位问题。在相似结构化场景中运行时本文算法难以进行大量有效特征信息的提取，帧与帧之间的位置关系不清晰，这样易造成定位失败。因此，下一步需要在本文构建的算法基础上提高相似结构化场景下定位能力。

（2）深度学习框架下语义 SLAM 的运行效率。本文所提两种算法均采用基于 Transformer 架构的语义分割网络对图像数据进行处理，资源占用较大，电力巡检机器人视觉 SLAM 算法在环境中响应时间较长。因此，如何提高深度学习框架下语义 SLAM 的运行效率是下一步必要解决的问题。

参考文献

[1] 苏晖,刘鲁京.电厂智能巡检机器人导航技术研究及应用[J].电子测试,2016(23):40-41.

[2] 李刚,智宏鑫.电力巡检机器人路径规划方法综述[J].电力科学与工程:1-10.

[3] 马晓君,保守福,齐鹏等.水电站巡逻机器人避障轨迹自动化控制技术[J].计算机测量与控制:1-13.

[4] 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.

[5] 高兴波,史旭华,葛群峰,陈奎烨.面向动态物体场景的视觉 SLAM 综述[J].机器人,2021,43(06):733-750.

[6] 陈孟元,韩朋朋,刘金辉等.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.

[7] 曹剑飞,余金城,潘尚杰,高枫,于超,胥智林,黄正峰,汪玉.采用双视觉里程计的 SLAM 位姿图优化方法[J].计算机辅助设计与图形学学报,2021,33(08):1264-1272.

[8] 徐陈,周怡君,罗晨.动态场景下基于光流和实例分割的视觉SLAM方法[J].光学学报,2022,42(14):147-159.

[9] 陈孟元,钱润邦,郭行荣等.动态场景下基于实例分割与运动一致性约束的 VSLAM 算法[J].中国惯性技术学报,2023,31(10):986-995.

[10] 陈孟元,郭行荣,钱润邦等.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.

[11] 梁佳宁,龙瀛.城市机器人的应用与空间应对研究综述[J].城市与区域规划研究,2023,15(01):47-71.

[12] 段昊宇翔,张宇,陈昊然等.工业机器人轨迹规划及轨迹优化研究综述[J].农业装备与车辆工程,2023,61(02):54-57.

[13] 张贵峰,张志强,沈锋.变电站巡检机器人现状与发展综述[J].云南电力技术,2022,50(06):2-8.

[14] 黄伦,朱玉琴,舒畅等.智能巡检机器人在自然环境试验中的应用探讨[J].装

备环境工程,2024,21(01):121-126.

[15]罗骏杰,董悦,潘亦辰.变电站智能巡检机器人多目标识别方法研究[J].电气技术与经济,2024(01):49-52.

[16] 黄松涛.基于机器视觉的电力巡检机器人自动化系统设计[J].自动化技术与应用,2024,43(01):35-38+43.

[17]王光璞,丁伟,刘庆达等.变电站智能巡检机器人数字孪生系统研究与应用[J].电气开关,2024,62(01):52-55.

[18]孙新柱,龚光强,陈孟元等.室内场景下基于曼哈顿约束的多重特征视觉 SLAM 方法[J].中国惯性技术学报,2023,31(09):890-899.

[19]曾庆化,罗怡雪,孙克诚等.视觉及其融合惯性的 SLAM 技术发展综述[J].南京航空航天大学学报,2022,54(06):1007-1020.

[20]王朋,郝伟龙,倪翠等.视觉 SLAM 方法综述[J].北京航空航天大学学报:1-9.

[21]Mur-Artal R, Tardós J D. Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras[J]. IEEE Transactions on Robotics, 2017, 33(5): 1255-1262.

[22]Campos C, Elvira R, Rodríguez J J G, et al. Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam[J]. IEEE Transactions on Robotics, 2021, 37(6): 1874-1890.

[23]Von Gioi R G, Jakubowicz J, Morel J M, et al. LSD: A fast line segment detector with a false detection control[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 32(4): 722-732.

[24]Gomez-Ojeda R, Moreno F A, Zuniga-Noël D, et al. PL-SLAM: A stereo SLAM system through the combination of points and line segments[J]. IEEE Transactions on Robotics, 2019, 35(3): 734-746.

[25]Li Y, Yunus R, Brasch N, et al. RGB-D SLAM with structural regularities[C]. 2021 IEEE International Conference on Robotics and Automation (ICRA). Xian, China, May 31 - June 4, 2021: 11581-11587.

[26]Moss M L, Qing C. The dynamic population of Manhattan[J]. NYU Wagner School of Public Service Working Paper, 2012.

[27]C, 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(3): 163.

[28] Likas A, Vlassis N, Verbeek J J. The global k-means clustering algorithm[J]. Pattern Recognition, 2003, 36(2): 451-461.

[29] Sun Y, Liu M, Meng M Q H. Motion removal for reliable RGB-D SLAM in dynamic environments[J]. Robotics and Autonomous Systems, 2018, 108: 115-128.

[30] Kaneko M, Iwami K, Ogawa T, et al. Mask-slam: Robust feature-based monocular slam by masking using semantic segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Utah, USA, June 18-23, 2018: 258-266.

[31] Zhong F, Wang S, Zhang Z, et al. Detect-SLAM: Making object detection and SLAM mutually beneficial[C]. 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). Lake Tahoe, NV, USA, March 12 -15, 2018: 1001-1010.

[32] 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(4): 4076-4083.

[33] Yu C, Liu Z, Liu X J, et al. DS-SLAM: A semantic visual SLAM towards dynamic environments[C]. 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Madrid, Spain, October 1-5, 2018: 1168-1174.

[34] Liu Y, Miura J. RDS-SLAM: real-time dynamic SLAM using semantic segmentation methods[J]. IEEE Access, 2021, 9: 23772-23785.

[35] 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 and Autonomous Systems, 2019, 117(1): 1-16.

[36] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017, 30.

[37] Han K, Wang Y, Chen H, et al. A survey on vision transformer[J]. IEEE

Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(1): 87-110.

[38] Feng X, Pei W, Jia Z, et al. Deep-masking generative network: A unified framework for background restoration from superimposed images[J]. IEEE Transactions on Image Processing, 2021, 30: 4867-4882.

[39] Barnes C, Shechtman E, Finkelstein A, et al. PatchMatch: A randomized correspondence algorithm for structural image editing[J]. ACM Trans. Graph., 2009, 28(3): 24.

[40] Dong B, Wang P, Wang F. Head-free lightweight semantic segmentation with linear transformer[J]. arXiv preprint arXiv:2301.04648, 2023.

[41] An L, Pan X, Li T, et al. A visual dynamic-SLAM method based semantic segmentation and multi-view geometry[C]. International Conference on High Performance Computing and Communication (HPCCE 2021). Xiamen, China, December 3-5, 2022, 12162: 255-263.

[42] Guan L, Franco J S, Pollefeys M. Multi-view occlusion reasoning for probabilistic silhouette-based dynamic scene reconstruction[J]. International Journal of Computer Vision, 2010, 90(3): 283-303.

[43] Caillette F, Howard T. Real-time markerless human body tracking with multi-view 3-D voxel reconstruction[C]. Proc. BMVC. 2004, 2: 597-606.

[44] Schubert D, Goll T, Demmel N, et al. The TUM VI benchmark for evaluating visual-inertial odometry[C]. 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Madrid, Spain, October 1-5, 2018: 1680-1687.

[45] Dai J, Li Y, He K, et al. R-fcn: Object detection via region-based fully convolutional networks[J]. Advances in Neural Information Processing Systems, 2016, 29.

[46] Chen L C, Zhu Y, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]. Proceedings of the European Conference on Computer Vision (ECCV). Munich, Germany, September 8-14, 2018: 801-818.

攻读学位期间发表的学术论文和科研成果

一、发表学术论文

【1】 陈孟元*，郭行荣，钱润邦，龚光强，程浩，基于区域映射深度优化的动态 RGBD-SLAM，中国惯性技术学报，2024，32（01）：42-51.

【2】 Chen* Mengyuan，Guo Hangrong，Qian Runbang，VSLAM algorithm based on improved Vision Transformer semantic segmentation in dynamic scenes，Mechanical Sciences，2024，15（1）：1-16.

【3】 H. Guo, R. Qian, M. Chen, G. Gong and H. Cheng, "RGBD-SLAM Algorithm Based on Semantic Segmentation and Background Restoration in Dynamic Scenes," *2023 38th Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, Hefei, China, 2023, pp. 1329-1333.

二、参与的科技项目

【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）

致谢

在毕业论文完成之际，我的硕士研究生学习生活即将结束，在三年的时间中我学习了许多知识，收获了许多的技术能力，在此，我要特别感谢在三年里给予我生活和学习巨大帮助的老师，同学，家人。

我要感谢我的导师陈孟元教授，感谢陈老师在三年来对我的帮助。在科研方面，陈老师以身作则与我们一起探讨论文创新点，同时指导我们实验验证创新点的可行性，在论文撰写过程中陈老师对我们有极大的耐心，教导我们论文写作的注意事项，更改论文中的错误，为我们解决了一个又一个难题。老师对待工作与学术有着严谨的态度，对待学习有着刻苦钻研的精神，这些为我们树立了榜样。

我要感谢实验室已经毕业的师兄在学习生活中给了我很大的帮助，在刚开始对 SLAM 不了解的时候师兄们鼓励我们多探索，在生活中给了我巨大帮助。也感谢实验室师弟师妹们你们为实验室带来了活力，愿你们在今后工作学习中一帆顺风顺。同时，我也要感谢我同门钱润邦、程浩、龚光强、姚满宇和我的室友丁峰、胡鹏飞，李熬，我们在一起生活学习三年，在硕士研究生期间相互帮助，即使毕业后大家会走向祖国各地，但是我仍然相信这是我人生中的一段不可忘却的宝贵经历。

我要感谢我的父母，是你们对我无私的奉献与帮助才让我能够全身心投入到研究生的学习中，感谢你们默默无闻的付出。

最后非常感谢在珍贵的科研和教学时间中抽出时间来评审论文的各位老师。

尚德教学 唯实惟新