



机电工程  
*Journal of Mechanical & Electrical Engineering*  
ISSN 1001-4551, CN 33-1088/TH

## 《机电工程》网络首发论文

题目: 基于 YOLOv8 实例分割和稠密建图的起重机视觉 SLAM 方法研究  
作者: 谢孟杞, 周颖, 王佰伟, 丁海洋, 惠延波, 周强, 周荷花, 王瞧  
收稿日期: 2025-05-16  
网络首发日期: 2025-09-29  
引用格式: 谢孟杞, 周颖, 王佰伟, 丁海洋, 惠延波, 周强, 周荷花, 王瞧. 基于 YOLOv8 实例分割和稠密建图的起重机视觉 SLAM 方法研究[J/OL]. 机电工程.  
<https://link.cnki.net/urlid/33.1088.TH.20250929.1343.002>



**网络首发:** 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式 (包括网络呈现版式) 排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

**出版确认:** 纸质期刊编辑部通过与《中国学术期刊 (光盘版)》电子杂志社有限公司签约, 在《中国学术期刊 (网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊 (网络版)》是国家新闻出版广电总局批准的网络连续型出版物 (ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

# 基于 YOLOv8 实例分割和稠密建图的起重机视觉 SLAM 方法研究\*

谢孟杞<sup>1</sup>, 周颖<sup>1,2</sup>, 王佰伟<sup>1</sup>, 丁海洋<sup>2</sup>, 惠延波<sup>1\*</sup>, 周强<sup>2</sup>, 周荷花<sup>2</sup>, 王瞧<sup>1</sup>

(1. 河南工业大学 机电工程学院, 河南 郑州 450001; 2. 卫华博士后工作站, 河南 新乡 453400)

**摘要:** 针对起重机智能化、无人化升级中, 传统视觉同步定位与地图构建 (SLAM) 系统易受动态特征点干扰, 造成实时性差和地图精度低的问题, 提出了一种将基于角点检测和描述子特征 (ORB) 提取的 ORB-SLAM3 与 YOLOv8 实例分割网络相结合的视觉 SLAM 系统 (SLAM3-YOLOv8)。首先, 新增了目标检测线程, 采用轻量化 YOLOv8n-seg 模型实时分割了动态目标, 利用非极大值抑制 (NMS) 去除了冗余检测框, 并基于实例分割了掩膜, 精准剔除了动态区域的特征点, 同时, 利用张量实时推理引擎 (TensorRT) 优化模型推理了架构; 然后, 集成了稠密建图模块, 该模块基于筛选后的静态特征点, 通过坐标转换将像素坐标和深度值转换为三维坐标, 经点云拼接构建了高精度的三维稠密点云图; 最后, 在慕尼黑工业大学 (TUM) 动态数据集上进行了精度验证, 进行了实验数据和实时性对比分析, 并在起重机测试平台开展了实验, 验证了该系统的准确性。研究表明: 该系统在动态场景下绝对轨迹误差 (ATE) 和相对位姿误差 (RPE) 较 ORB-SLAM3 均显著降低, 最大降幅分别达到 95.42% 和 54.15%, 并且单帧处理时间平均优化至 18.08 ms, 验证了该系统在动态场景中具有更高的定位精度、更小的漂移误差和更高的实时性; 同时, 成功构建出了无重影的三维稠密点云图。证明其能够有效克服起重机复杂作业环境中的动态干扰, 为工业重载设备智能化升级提供可靠技术方案。

**关键词:** 起重机械; 同步定位与地图构建; YOLOv8n-seg; 动态场景; 目标检测; 稠密点云; 绝对轨迹误差; 相对位姿误差  
**中图分类号:** TH21 **文献标识码:** A

## Crane visual SLAM method based on YOLOv8 instance segmentation and dense mapping

XIE Mengqi<sup>1</sup>, ZHOU Ying<sup>1,2</sup>, WANG Baiwei<sup>1</sup>, DING Haiyang<sup>2</sup>, HUI Yanbo<sup>1</sup>, ZHOU Qiang<sup>2</sup>, ZHOU Hehua<sup>2</sup>,  
WANG Qiao<sup>1</sup>

(1. School of Mechanical and Electrical Engineering, Henan University of Technology, Zhengzhou, Henan 450001, China; 2. Weihua Postdoctoral Workstation, Xinxiang, Henan 453400)

**Abstract:** To address the problems of traditional visual simultaneous localization and mapping (SLAM) systems being susceptible to dynamic feature point interference, poor real-time performance, and low map accuracy during the intelligent and unmanned upgrading of cranes, a visual SLAM system (SLAM3-YOLOv8) combining ORB-SLAM3 based on oriented fast and rotated brief (ORB) feature extraction with the YOLOv8 instance segmentation network was proposed. Firstly, an additional target detection thread was incorporated, a lightweight YOLOv8n-seg model was utilized to segment moving objects in real time, redundant detection boxes were removed via non-maximum suppression (NMS) and feature points in dynamic regions were accurately eliminated based on instance segmentation masks, simultaneously, a tensor real-time (TensorRT) inference engine was used to optimize the model inference architecture. Then, a dense mapping module was integrated, this module was based on the filtered static feature points, pixel coordinates and depth values were transformed into 3D coordinates through coordinate transformation, and high-precision 3D dense point cloud maps were constructed through point cloud stitching. Finally, accuracy validation was conducted on the dynamic dataset of the Technische Universität München (TUM), with experimental data and real-time comparison analysis. Experiments were also carried out on a crane testing platform to verify the accuracy of the system. The research results show that: the system significantly reduces absolute trajectory error (ATE) and relative pose error (RPE) compared to ORB-SLAM3 in dynamic scenes. The maximum reductions respectively reach 95.42% and 54.15%, and the average single-frame processing time is optimized to 18.08 ms. This verifies the system and achieves higher localization accuracy, lower drift error in dynamic environments and higher real-time performance. Meanwhile, it also successfully constructs ghost-free dense 3D point cloud maps. This proves its capability to effectively overcome dynamic disturbances in complex crane working environments, provide reliable technical solutions for the intelligent upgrade of industrial heavy-duty equipment.

**Key words:** hoisting machinery; simultaneous localization and mapping (SLAM); YOLOv8n-seg; dynamic scene; object detection; dense point cloud; absolute trajectory error (ATE); relative pose error (RPE)

收稿日期: 2025-05-16

基金项目: 河南省科技攻关联合基金资助项目 (22103810083); 河南工业大学自然科学基金资助项目 (2022BS032)

作者简介: 谢孟杞 (2000-), 男, 河南安阳人, 硕士研究生, 主要从事数字化设计与制造、智能感知等方面的研究。E-mail: 18567833580@163.com

通信联系人: 惠延波, 男, 博士, 教授。E-mail: huiyb@haut.edu.cn

## 0 引言

起重机作为工业物料搬运的核心设备,其安全运行对生产保障至关重要。

当前,物料搬运过程高度依赖人工操控。该过程中涉及到避障及目标点精准定位等问题,多是依靠操作人员的经验去判断和解决,这一模式效率较低,同时易因视觉盲区、疲劳误判或突发干扰的影响,导致吊具偏移、负载晃动,威胁生产安全。2024年全国共报告特种设备相关事故132起,其中,起重机械事故数量占比27.27%,排名第三;死亡人数35.97%,排名第一<sup>[1]</sup>。这些数据凸显了传统人工操作模式在复杂或高要求场景下存在的安全风险。

随着智能化工厂理论不断发展,在特定应用场景中,起重机智能化与无人化已成为行业发展的重要趋势<sup>[2]</sup>——通过环境感知技术补足人工短板,为传统机械安全防护体系提供决策支撑。

目前,同步定位与地图构建(simultaneous localization and mapping, SLAM)技术获得了较快发展<sup>[3]</sup>,尤其是在环境感知方面。该技术已成功应用于位姿估计、地图构建、路径规划等多个方面,如仓储物流智能搬运、建筑工地自主测绘、自动驾驶等领域<sup>[4]</sup>。将SLAM技术引入到起重机领域,有望提高起重机的环境感知能力,减少安全事故的发生,这是实现起重机无人化、智能化的前提<sup>[5-6]</sup>。然而,起重机作业环境易受动态物体影响,场景中频繁移动的载具、作业人员致使动态物体占比高,传统视觉SLAM易因动态特征点误匹配引发定位漂移与地图失真,并且实时性不足,严重制约其在工业场景中的实用性<sup>[7]</sup>。

因此,提高动态场景下的定位精度和地图一致性,是视觉SLAM领域面临的挑战<sup>[8]</sup>。

为应对该挑战,诸多研究学者开始将深度学习技术与SLAM技术相结合,以有效处理动态环境中的特殊对象<sup>[9]</sup>。

BESCOS B等人<sup>[10]</sup>提出了Dyna-SLAM系统,在基于角点检测和描述子特征(oriented fast and rotated brief, ORB)提取的ORB-SLAM2系统中,融入了掩码区域卷积神经网络(mask region-based convolutional neural network, Mask R-CNN)和多视角几何模块,借助实例分割算法,对图像中的物体进行了分割,进而剔除了动态物体上的特征点;但该算法耗时严重,实时性较差。CHAO Yu等人<sup>[11]</sup>提出了动态环境中的完整语义(dynamic environments

of complete semantics, DS)-SLAM系统,融合了语义分割网络和运动一致性检测方法,基于语义分割结果生成了动态物体掩膜,结合运动一致性检测到的运动信息能够有效降低动态物体的影响;但该系统识别的物体种类少。ZHONG Fang-wei等人<sup>[12]</sup>提出了Detect-SLAM系统,引入了单次多框检测器(single shot multi-box detector, SSD)目标检测算法<sup>[13]</sup>,对当前帧进行了目标检测,生成了动态物体的边界框和掩膜,剔除了动态物体的特征点;但使用的目标检测网络有时无法有效区分同类动态物体,导致动态特征点剔除不准确。

随着YOLO算法在目标检测领域的应用越发广泛,其高精度、高实时性的特点为动态场景下的视觉SLAM系统提供了关键技术支持。

WU Wen-xin等人<sup>[14]</sup>提出了一种语义SLAM框架,融入了YOLOv3目标检测网络,能在动态环境中检测动态物体,并利用深度信息进行了特征点剔除,进而减少了动态物体对SLAM的影响。YU Ya-ting等人<sup>[15]</sup>针对在动态场景下位姿估计精度低、鲁棒性差的问题,提出了一种基于ORB-SLAM3框架的YG-SLAM系统,在动态目标检测线程中引入了YOLOv5算法,进行了目标识别,并根据识别结果将动态特征点消除。CONG Pei-chao等人<sup>[16]</sup>在剔除动态特征的基础上融合了稠密建图模块,将剔除后的结果与深度信息相融合,利用点云库构建了不含动态物体的静态稠密点云图。

现有研究虽推动了深度学习与SLAM技术的融合,但在复杂工业场景中仍存在显著缺陷。早期方法虽能完成目标分割与特征剔除工作,但算法耗时严重,实时性不足,且语义识别类型有局限,难以覆盖起重机、吊具等工业设备。轻量化模型虽能提升推理速度,但检测框定位精度不足,易导致动态特征被错误剔除,且依赖稀疏地图无法满足密集场景下的作业需求,尤其在起重机电大尺度运动、高频振动及金属反光干扰下,系统鲁棒性与场景适应性均显著下降。

针对起重机作业场景的特点,为解决工业场景中动态干扰与稀疏地图实用性不足的共性问题,笔者提出一种将ORB-SLAM3与YOLOv8实例分割网络相结合的视觉SLAM系统(SLAM3-YOLOv8)。

首先,在ORB-SLAM3系统架构基础上新增目标检测线程,采用轻量化YOLOv8n-seg模型进行动态目标的实时分割,并基于生成的像素级掩膜精准剔除动态区域的特征点,利用张量实时推理引擎

(tensor real-time inference engine, TensorRT) 优化模型推理架构; 然后, 集成稠密建图模块, 该模块基于筛选后的静态特征点, 通过坐标转换将二维坐标转换为三维坐标, 经点云拼接和滤波处理, 构建高精度的三维稠密点云图; 最后, 在动态数据集及起重机测试平台上开展系统性能验证实验, 验证该系统的准确性。

## 1 系统框架

SLAM3- YOLOv8 系统是在 ORB-SLAM3 系统原有的跟踪线程、局部建图线程和回环检测线程的基础上, 新增目标检测线程, 并且在局部建图线程中引入稠密建图模块, 整体架构如图 1 所示<sup>[17]</sup>。

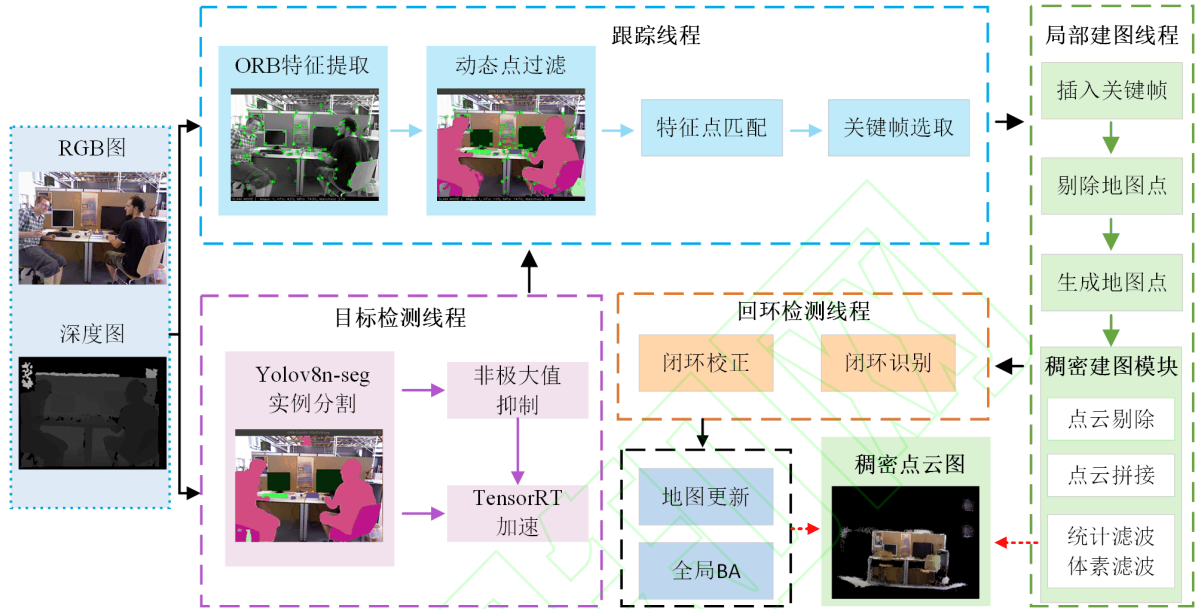


图 1 系统框图

Fig.1 System block diagram

由图 1 可知: 笔者利用深度相机获取起重机场景的深度图和彩色图 (red green blue image, RGB), 将获取到的图像同时传送到跟踪线程和目标检测线程中, 目标检测线程采用 YOLOv8n-Seg 对物体进行分类识别, 对不同的物体生成不同的分割掩膜, 并使用非极大值抑制 (non-maximum suppression, NMS) 功能将重叠的检测框去除, 采用张量实时推理引擎 (TensorRT) 优化模型推理架构; 同时, 跟踪线程对刚采集到的图像进行角点检测和描述子特征 (ORB) 提取, 并将落在动态掩膜内的特征点判定为动态特征后予以剔除; 然后, 依据匹配到的静态特征点开展初步位姿估计与局部地图构建; 最后, 使用稠密建图模块对局部地图生成的点云进行拼接以及统计滤波、体素滤波等降噪处理, 从而生成三维稠密点云图。

## 2 基于 YOLOv8 的目标检测线程

### 2.1 YOLOv8n-seg 实例分割网络

在目标检测线程中, 笔者选用参数数量最少的 YOLOv8n-seg 模型对 RGB 图像进行目标检测和实

例分割。

其网络架构主要由三部分组成: 主干网络、颈部网络和头部网络<sup>[18]</sup>。其中, 各部分作用如下:

1) 主干网络主要负责从输入图像中提取特征, 基于跨阶段部分连接结构 (cross stage partial connections, CSP), 采用更轻量的双卷积跨阶段部分融合模块 (cross stage partial bottleneck with 2 convolutions, C2f), 替代 YOLOv5 中跨阶段部分模块 (cross stage partial module, C3), 增强梯度流并减少参数量;

2) 颈部网络用于融合主干网络输出的多尺度特征图, 增强特征的表达能力;

3) 头部网络负责同时完成目标检测和实例分割任务<sup>[19]</sup>, 采用解耦头的方式将检测和分割任务分离, 从而提升训练效率和模型性能。

### 2.2 动态特征点的剔除

为完成动态特征点剔除工作, 需考虑 YOLO 算法 (使用 Python 语言) 和 SLAM 系统 (使用 C++ 语言) 之间的通信问题。

在 YOLOv8 版本之前, YOLO 算法都需要结合



LibTorch 来加载和执行 PyTorch 训练出来的模型，从而实现两者之间的通信目的。但是 YOLOv8 可以直接导出开放神经网络交换格式（open neural network exchange, ONNX），使用 ONNX Runtime 在 C++ 中加载模型<sup>[20]</sup>。

笔者设计的剔除动态特征点方法如图 2 所示。

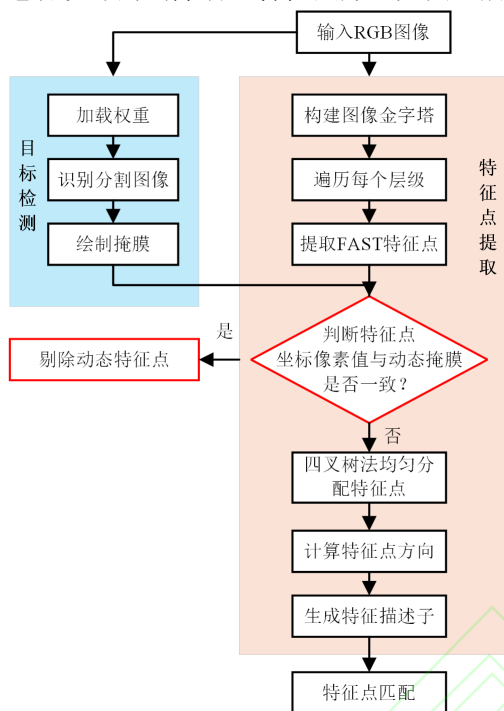


图 2 动态特征点剔除策略

Fig.2 Dynamic feature points removal strategy

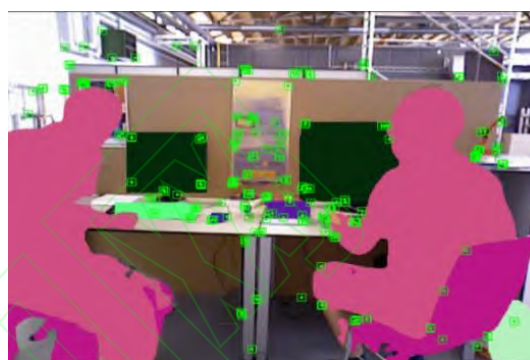
由图 2 可知：当输入 RGB 图像时，使用四叉树法在金字塔的每个层级均匀提取改进的 FAST 角点，结合 YOLOv8n-seg 网络模型，对提取到的 ORB 特征点进行筛选。

具体而言，笔者采用 YOLOv8n-seg 对不同物体实例进行分割并分类，生成不同颜色的掩膜，比较每个特征点的像素和动态掩膜的像素，判断提取到的 FAST 角点是否位于指定的动态掩膜内；然后，将位于该掩膜上的特征点判定为动态特征点并进行剔除。

动态特征点剔除效果如图 3 所示。



(a) 剔除前



(b) 剔除后

图 3 动态点特征剔除

Fig.3 Dynamic points feature removal

由图 3 可知：动态物体的特征点已被有效剔除，其主要目的是确保提取的是静态特征点，降低动态特征点对后续特征匹配及位姿估计的影响。

## 2.3 TensorRT 网络加速

目标检测算法需要较高的实时性来应对起重机复杂多变的作业环境。为提高系统的实时性，笔者在算法中集成 TensorRT 引擎，以加速深度学习模型推理，使用模型压缩、内核自动调优、动态内存优化等技术，显著提升模型在 GPU 上的推理速度<sup>[21]</sup>。

TensorRT 对动态维度的支持有限，尤其是对动态宽高的处理会显著增加引擎构建复杂度。为适配 TensorRT 需做以下修改：

输入、输出仅让 batch 维度变化，宽高保持固定。因为 TensorRT 的后处理代码所期望输入模型的维度是 [batch, num\_classes+4+num\_masks, num\_anchors]，而默认导出模型输出张量的维度为 [batch, num\_anchors, num\_classes+4+num\_masks]，所以需要新增 transpose 节点，交换输出的 2、3 维度。其中，Batch 表示输入模型的样本数量；num\_anchors 表示每个检测头的锚点数量；num\_classes+4+num\_masks 表示分类得分、边界框坐标（4 维）和分割掩码参数。

从而可将 YOLOv8n-seg 模型部署到 TensorRT

环境中，利用 TensorRT 的高性能计算能力完成目标检测任务。

#### 2.4 非极大值抑制

针对目标检测过程中可能产生大量重叠的检测框的问题，笔者使用非极大值抑制去除检测结果中的冗余信息，保留最有可能包含目标物体的检测框<sup>[22]</sup>。交并比（intersection over union, IoU）公式如下：

$$\text{IoU} = \frac{\text{Area}(A \cap B)}{\text{Area}(A \cup B)} \quad (1)$$

式中：\$A, B\$ 为两个边界框；\$\text{Area}(A \cap B)\$ 为它们的交集面积；\$\text{Area}(A \cup B)\$ 为它们的并集面积。

在非极大值抑制的过程中，这个公式用于计算两个边界框之间的交并比。其中，交并比阈值设置为 0.7。当两个检测框的交并比超过这个阈值，置信度较低的框会被抑制，以此抑制掉大量重叠的边界框，仅保留最好的那个边界框，从而获得最终的目标检测结果。

### 3 三维稠密地图重建

#### 3.1 坐标转换

点云地图<sup>[23]</sup>是由大量三维空间坐标点组成的几

何数据集。其构建方法采用的是由深度相机制作的数据集，可直接从深度图像和彩色图像提取像素坐标 \$\mathbf{P}\_u(u, v)\$ 和深度值 \$d\$。

基于针孔相机模型，可利用相机内部的焦距和光心坐标等参数将像素坐标和深度值转换为三维坐标 \$\mathbf{P}\_c(X\_c, Y\_c, Z\_c)\$，转换公式如下：

$$\begin{cases} X_c = \frac{(u - c_x) \cdot d}{f_x} \\ Y_c = \frac{(v - c_y) \cdot d}{f_y} \\ Z_c = d \end{cases} \quad (2)$$

经上式推导可得 \$\mathbf{P}\_u\$ 和 \$\mathbf{P}\_c\$ 之间的转换公式如下：

$$d\mathbf{P}_u = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \mathbf{K}\mathbf{P}_c \quad (3)$$

式中：\$\mathbf{K}\$ 为相机内参矩阵；\$f\_x\$ 和 \$f\_y\$ 为相机焦距；\$c\_x\$ 和 \$c\_y\$ 为光心坐标<sup>[24]</sup>。

#### 3.2 点云拼接

稠密点云图的构建方法如图 4 所示。

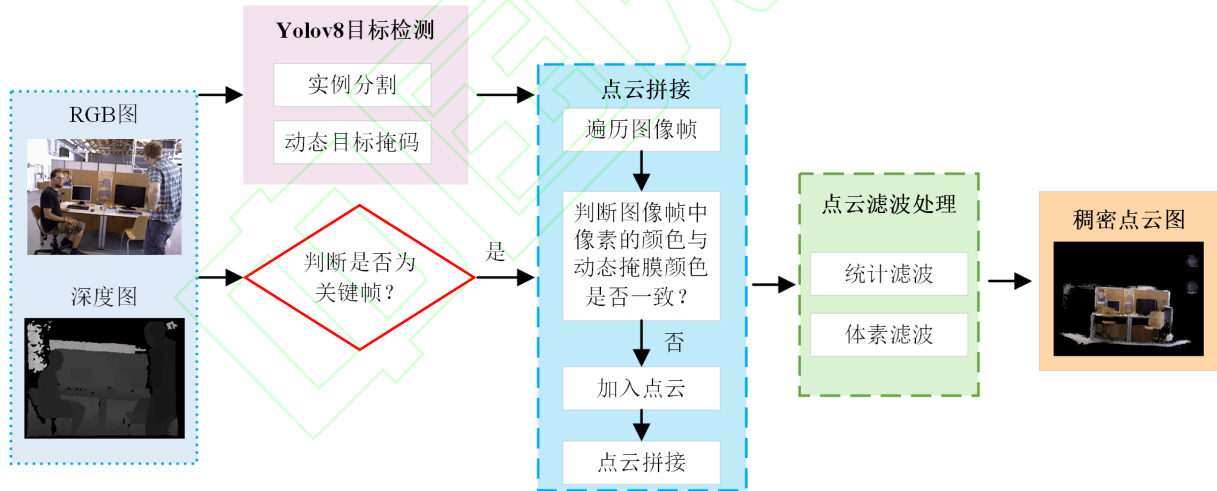


图 4 稠密点云图构建方法

Fig.4 Dense point cloud map construction method

由图 4 可知：系统通过深度相机获取图片的彩色信息和深度信息；图像帧经过跟踪线程判断是否为关键帧，根据 YOLOv8n-seg 目标检测线程对不同物体的分类识别，检查当前坐标下的像素值是否与动态掩膜的像素值匹配，如果匹配，则跳过当前像素值的重建；若不匹配，则将该像素点添加到静态点云。

笔者对输入点云中的离群点进行统计滤波器处

理，去除离群点，得到滤波后的点云，并将其拼接到主点云中；对拼接的整体点云进行体素滤波处理，在保留场景整体结构的同时剔除冗余点，最终生成三维稠密点云图。

### 4 实验及结果分析

#### 4.1 实验环境配置

该实验运行的操作环境为 Ubuntu18.04，

CUDA11.3, Cudnn8.2.1, 硬件配置 CPU 为英特尔 i9-10900, GPU 为英伟达 RTX-2080Ti。该配置能够满足深度学习模型的高效推理需求, 同时支持实时 SLAM 系统的运行。

为评估 SLAM3-YOLOv8 系统性能, 笔者采用德国慕尼黑工业大学 (Technische Universität München, TUM) 发布的 SLAM 基准数据集<sup>[25]</sup>进行实验, 选取涵盖 5 种不同运动状态的数据集进行测试, 分别为 freiburg3 数据序列中的 walking\_xyz、walking\_halfsphere、walking\_rpy、walking\_static、sitting\_static。这 5 个场景均为室内场景。其中, 在含 walking 标注的数据集中, 室内人物处于行走状态, 构成高动态场景; 在含 sitting 标注的数据集中, 室内人物处于小幅度运动状态, 构成低动态场景。

在对视觉 SLAM 系统进行评估时, 通常采用绝对轨迹误差 (absolute trajectory error, ATE) 和相对位

姿误差 (Relative Pose Error, RPE) 衡量系统的定位精度, 并将 ATE 和 RPE 的均方根误差 (root mean squared error, RMSE)、标准差 (standard deviation, SD) 作为评价精度的标准。

在实验过程中, 笔者使用 ORB-SLAM3 系统、Dyna-SLAM 系统、SLAM3-YOLOv5 系统及 SLAM3-YOLOv8 系统进行对比分析, 并在所选 5 组数据集中分别运行 3 次, 以记录实验结果的平均值。

性能提升的计算公式如下:

$$\eta = \frac{n_1 - n_2}{n_1} \cdot 100\% \quad (4)$$

式中:  $\eta$  为提升率;  $n_1$  为 ORB-SLAM3 系统的结果;  $n_2$  为 ORB-SLAM3-YOLOv8 系统的结果。

## 4.2 实验对比分析

ATE 的对比结果如表 1 所示。

表 1 基于 ATE 的实验结果对比

Table 1 Comparison of experimental results based on ATE

数据 序列	ORB-SLAM3		Dyna-SLAM		SLAM3-YOLOv5		SLAM3-YOLOv8		提升率/%	
	RMSE	SD	RMSE	SD	RMSE	SD	RMSE	SD	RMSE	SD
w_xyz	0.283 7	0.118 5	0.013 9	0.099	0.013 4	0.007	0.013	0.006 6	95.42	94.43
w_half	0.211 2	0.075 9	0.021 4	0.011 2	0.024 8	0.016 1	0.018 6	0.008 5	91.19	88.80
w_rpy	0.154 6	0.074 7	0.032 5	0.020 6	0.139 7	0.073 9	0.031 6	0.017 9	79.56	76.04
w_static	0.025 1	0.014 7	0.01	0.006	0.007 7	0.003 7	0.008 5	0.003 8	66.14	74.15
s_static	0.007 1	0.003 2	0.006 7	0.003 4	0.008 2	0.003 9	0.006	0.003	15.49	6.25

由表 1 可知: 不同 SLAM 系统的 ATE 得到了详尽地对比分析, 该指标是衡量全局定位精度的关键;

在高动态场景中, SLAM3-YOLOv8 的 ATE 值显著低于其他方法, 具体表现为在 RMSE 方面, 最高提升了 95.42%, 平均提升了 83.08%; 在 SD 方面, 最高提升了 94.43%, 平均提升了 83.36%。这表明该系统在处理动态物体方面具有显著优势;

在低动态场景中, 提升效果较低, 主要原因在于低动态场景中动态物体较少, ORB-SLAM3 的随机抽样一致性算法 (random sample consensus, RANSAC) 能有效滤除少量动态特征点, 从而保证

系统的鲁棒性和定位精度。在 w\_static 数据中, 该系统效果略低于 SLAM-YOLOv5, 原因可能是 YOLOv8 对动态目标检测更为敏感, 但在动态场景的静态子区域内, 这种高灵敏度易引发过度检测, 致使位姿估计误差略有增加。

总体而言, 除 w\_static 序列外, SLAM3-YOLOv8 系统在其余所有数据序列中均取得最优结果。

SLAM3 结合 YOLOv8 的目标检测能力, 在动态场景中表现出更高的精度和更强的鲁棒性, 而在静态场景中与传统方法差距较小但也略有提升。

RPE 的平移误差对比结果如表 2 所示。

表 2 基于 RPE 的平移误差实验结果对比

Table 2 Comparison of translational error experimental results based on RPE

数据 序列	ORB-SLAM3		Dyna-SLAM		SLAM3-YOLOv5		SLAM3-YOLOv8		提升率/%	
	RMSE	SD	RMSE	SD	RMSE	SD	RMSE	SD	RMSE	SD
w_xyz	0.025 3	0.014 6	0.027 8	0.023 7	0.012 3	0.007 1	0.011 6	0.006 7	54.15	54.11
w_half	0.024 7	0.016 2	0.014 8	0.009 1	0.013 7	0.008 1	0.013 5	0.007 4	45.34	54.32
w_rpy	0.027 2	0.016 8	0.023 6	0.016 6	0.022 5	0.015 1	0.021 4	0.014 2	21.32	15.48

w_static	0.014 3	0.010 6	0.01	0.007 3	0.006 1	0.003 3	0.007	0.004 4	51.05	58.49
s_static	0.005 8	0.003	0.005 5	0.002 8	0.005 9	0.003 1	0.005 4	0.002 8	6.90	6.67

由表 2 可知:不同 SLAM 系统在 RPE 中的平移误差得到了详尽地对比分析,该指标是衡量局部定位中位移距离精度的关键;在高动态场景中,SLAM3-YOLOv8 系统表现出显著优势,具体表现为在 RMSE 方面,最高提升了 54.15%,平均提升了 42.97%;在 SD 方面,最高提升了 58.49%,平均提

升了 45.6%。这充分证明了该系统在动态干扰下运动距离估计的准确性。在低动态场景中,各系统间的平移误差差距相对较小,但该系统仍略优于其他对比系统,保持了稳定的性能。

RPE 的旋转误差对比结果如表 3 所示。

表 3 基于 RPE 的旋转误差实验结果对比

Table 3 Comparison of rotational error experimental results based on RPE

数据 序列	ORB-SLAM3		Dyna-SLAM		SLAM3-YOLOv5		SLAM3-YOLOv8		提升率/%	
	RMSE	SD	RMSE	SD	RMSE	SD	RMSE	SD	RMSE	SD
w_xyz	0.510 9	0.358 8	0.554 2	0.450 2	0.389 6	0.274 9	0.388 4	0.273 5	23.98	23.77
w_half	0.606 5	0.353 1	0.432 3	0.235 2	0.399 7	0.210 7	0.416 2	0.219 5	31.38	37.84
w_rpy	0.632 7	0.356 4	0.532 4	0.347 8	0.540 3	0.354 9	0.519 9	0.328	17.83	7.97
w_static	0.294	0.190 7	0.18	0.090 8	0.170 2	0.087	0.187 6	0.102 8	36.19	46.09
s_static	0.17	0.087	0.162 8	0.084 3	0.170 2	0.087	0.160 7	0.084	5.47	3.45

由表 3 可知:不同 SLAM 系统在 RPE 中的旋转误差得到了详尽地对比分析,该指标是衡量局部定位中姿态方向变化精度的关键;SLAM3-YOLOv8 系统在 RMSE 方面,最高提升了 36.19%,平均提升了 27.35%;在 SD 方面,最高提升了 46.09%,平均提升了 28.92%。在低动态场景中,各系统间的旋转误差差距较小。

结合 RPE 的平移误差和旋转误差分析可知:SLAM3-YOLOv8 系统改进的网络结构与动态目标检测机制,能有效抑制动态干扰对位姿估计的影响。表明其对动态场景下的位姿漂移具有更高的定位精度和更强的鲁棒性。

虽然在 w\_static 数据集中 YOLOv5 的效果优于 YOLOv8,但数值相差很小,并且在绝大多数评估序

列中,该系统综合表现优于其他对比系统。即使在低动态场景中,该系统也展现出轻微优势,进一步验证其改进后的系统在提升局部定位精度方面的普适性与有效性。

根据以上实验表明,无论在高动态还是低动态环境下,该系统均能有效处理各类复杂场景中的动态目标,显著提升 SLAM 系统在复杂场景下的鲁棒性和位姿轨迹估计的准确性。

为更直观地比较估计轨迹与真实轨迹之间的误差,笔者在 w\_xyz、w\_half、w\_rpy、w\_static 四种高动态数据集中,对 4 种系统所预测的相机轨迹进行定性比较。通过直观地对比分析,有助于评估该系统在不同运动情况下的性能表现。

轨迹对比结果如图 5 所示。



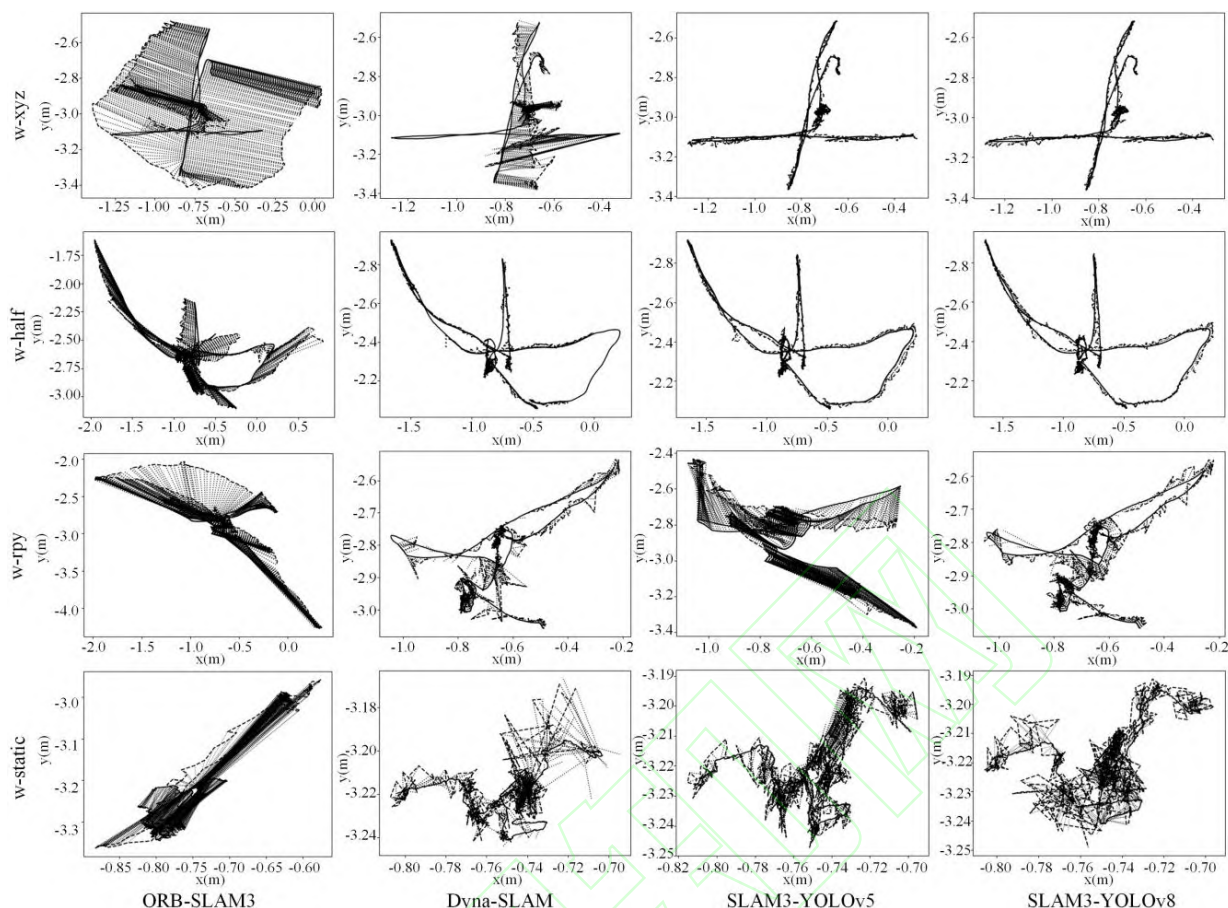


图 5 不同 SLAM 系统预测轨迹对比

Fig.5 Trajectory prediction comparison of SLAM systems

注：实线为真实轨迹；虚线为估计轨迹；点线为真实轨迹和估计轨迹之间的误差。

由图 5 可知：通过定性分析，点线越短越稀疏，估计轨迹与真实轨迹的一致性越高，表明系统的定位精度越高；ORB-SLAM3 系统因无法适应动态环境导致轨迹误差是最大的，另外 3 个动态 SLAM 都有所改进，其中 SLAM3-YOLOv8 系统具有相对较好的处理结果。

#### 4.3 实时性分析

不同 SLAM 系统在五种 TUM 数据集上的跟踪时间如表 4 所示。

表 4 跟踪线程运行时间实验结果对比

Table 4 Runtime comparison of tracking thread (ms)

数据序列	SLAM3	SLAM3-YOLOv5	SLAM3-YOLOv8
w_xyz	19.63	30.07	18.45
w_half	19.91	35.87	18.56
w_rpy	17.31	32.01	16.97
w_static	20.00	24.31	19.11
s_static	14.85	23.05	17.30

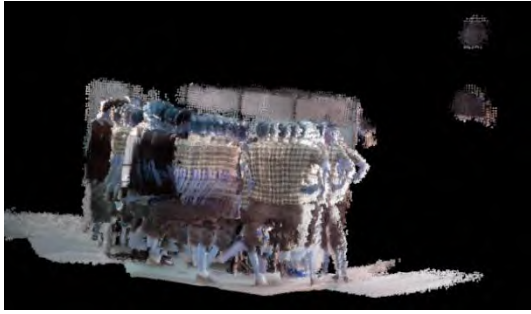
由表 4 可知：尽管在 ORB-SLAM3 的基础上

新增目标检测线程增加了算法复杂度及系统运行时间；但通过引入轻量化的 YOLOv8n-seg 模型并结合 TensorRT 加速网络，显著降低了模型的计算复杂度和推理时间；并且平均每帧的跟踪时间都要优于原系统，单帧处理时间平均优化到 18.08 ms，实时性方面体现出显著优势。

#### 4.4 稠密地图验证

为更好地体现 SLAM3-YOLOv8 系统在高动态环境下的建图效果，笔者选用 walking\_xyz 数据集分别在原始 ORB-SLAM3 系统和该系统进行静态点云地图构建测试。

动态特征点过滤前后构建的稠密点云图效果如图 6 所示。



(a) 动态特征点剔除前



(b) 动态特征点剔除后

图 6 walking\_xyz 数据集稠密点云图

Fig.6 Dense point cloud map of walking\_xyz dataset

由图 6 可知:当未对动态特征点进行滤除时,构建的稠密点云图中会出现人行走时的重影。这是因为动态物体处于不断运动状态,其点云位置信息相应地发生变化。若未剔除这些动态特征点,将影响建图准确性;

相比之下,图 6(b)展示了剔除动态物体后所构建出的稠密点云图。可以看到,该系统能够有效保留环境中的大量静态点云,同时避免重影现象的出现,生成的三维点云图质量较高,具有良好的清晰度和较高的准确性。

#### 4.5 起重机场景下 SLAM 系统验证

为测试 SLAM3-YOLOv8 系统在真实起重机动态场景下的性能,笔者搭建了一个桥式起重机运动实验平台,实验场景如图 7 所示。



图 7 起重机实验平台

Fig.7 Experimental platform for crane system

由图 7 可知:在搭建的桥式起重机实验平台中,小车安装在大车钢轨上,可沿大车钢轨左右水平移动;大车则安装在固定钢轨上,可沿固定钢轨前后水平移动。

由于桥式起重机安装产生的误差会导致大车和小车运行方向不完全平行于地面,因此,只需要考虑水平方向的坐标变化,而忽略垂直方向运动误差带来的影响。

实验采用 Intel RealSense D435i 深度相机作为环境感知的传感器。相机主要参数如表 5 所示。

表 5 D435i 深度相机主要参数

Table 5 Key specifications of D435i depth camera

参数名称/单位	数值
测量范围/m	0.1~10
最小深度距离/cm	-28
深度误差(2 m 内)	<2%
深度图像分辨率	1 280×720
彩色图像分辨率	1 280×720
尺寸/mm	95×25×25

由表 5 可知:该相机可以采集场景中的深度信息,测量范围为 0.1 m~10 m,满足起重机场景下的实验要求。

在实验过程中,笔者将相机倾斜固定于起重机小车上。相机实物图及安装位置如图 8 所示。

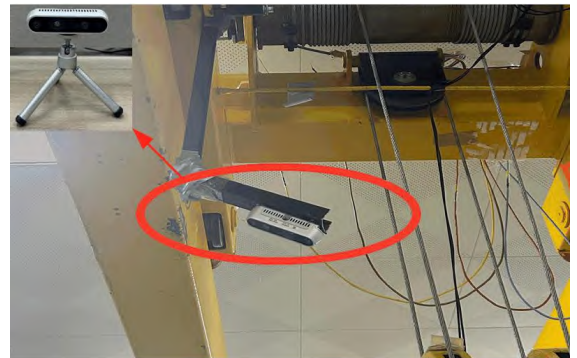


图 8 D435i 相机实物展示与安装位置图

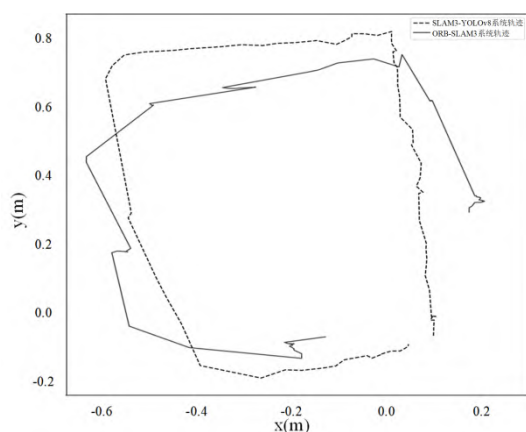
Fig.8 Physical setup and installation location of D435i camera

由图 8 可知:相机固定在小车下方,随小车在大车钢轨上水平移动后静止;小车协同大车在固定钢轨上水平移动,且保持小车与相机的相对位置不变;抵达指定地点后,小车率先返回在大车钢轨上的起始位置,之后大车再返回起始位置,从而形成起重机大车与小车的矩形运动轨迹。

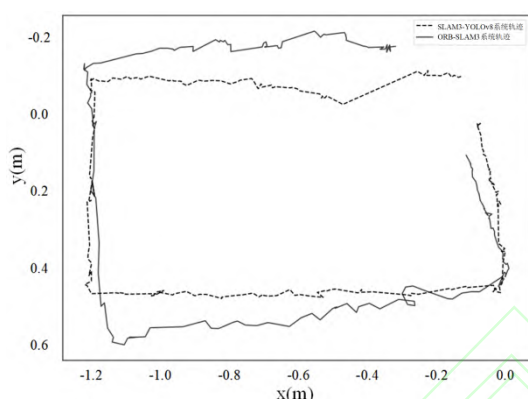
笔者比较了 ORB-SLAM3 系统和 SLAM3-YOLOv8 系统在静态场景、动态场景中获得的轨



迹，对比效果如图 9 所示。



(a) 静态场景位姿轨迹对比



(b) 动态场景位姿轨迹对比

图 9 位姿轨迹对比

Fig.9 Pose trajectory comparison

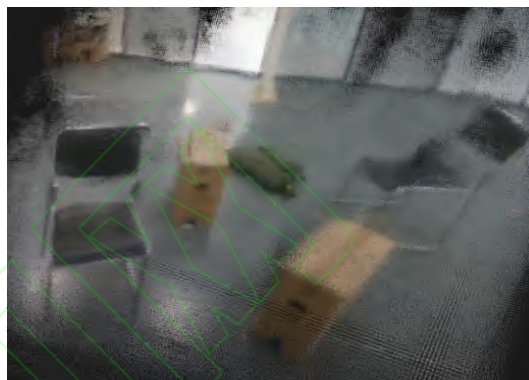
由图 9 可知：无论是在静态场景还是动态场景中，SLAM3-YOLOv8 系统的轨迹形状更接近规则矩形，且轨迹闭合性优于原系统；在运动过程中，电机转动或其他因素可能导致相机振动，进而影响位姿估计，但对整体效果影响不大，尤其在动态场景下，该系统展现出更高的定位精度与更强的鲁棒性。

这证明其定位精度能够满足起重机的应用条件。

构建的三维稠密点云地图如图 10 所示。



(a) 动态特征点剔除前



(b) 动态特征点剔除后

图 10 起重机室内场景下点云地图

Fig.10 Point cloud map in indoor crane scenario

对比图 10，可以明显看出动态特征点对地图质量的影响。

在图 10 (a) 中，动态特征点严重破坏了地图的一致性，导致重影现象的发生，可能对后续的路径规划和导航任务造成干扰；相比之下，SLAM3-YOLOv8 系统在构建稠密点云图方面表现出色。

这一结果表明，该系统能够有效处理场景中动态障碍物的影响，保留场景整体结构的同时剔除冗余点，确保地图的一致性和准确性。

实验结果充分证明了 SLAM3-YOLOv8 系统在起重机动态场景下的定位与地图构建任务中的有效性。该系统不仅能够准确地估计小车的运动轨迹，避免动态障碍物的干扰，而且能够构建出高质量的稠密点云地图，满足起重机动态场景下的定位与地图构建需求。

这一成果对于提高自动化设备在复杂环境中的自主导航能力具有重要意义，可为智能化起重机的精准环境感知与路径规划提供技术支撑。

## 5 结束语

针对起重机动态作业场景下视觉 SLAM 系统易受运动物体干扰，导致定位漂移与地图失真的

问题,笔者提出了一种将 ORB-SLAM3 系统与 YOLOv8 实例分割网络相结合的视觉 SLAM 系统 (SLAM3-YOLOv8);在该系统中进行了 TUM 动态数据集精度验证,并在起重机测试平台开展了实验。

研究结论如下:

1) 提出了一种基于实例分割的动态特征实时剔除机制,新增了目标检测线程,引入了轻量化 YOLOv8n-seg 模型,实时分割了动态目标并生成了掩膜,结合动态特征点剔除策略有效抑制了动态干扰;同时,集成了 TensorRT 加速推理网络,显著提升了系统实时性;

2) 集成了稠密建图模块,通过坐标转换,利用相机内部参数将像素坐标和深度值转换为三维坐标,再对静态点云进行了拼接和滤波处理,生成了静态三维稠密点云图;

3) 基于 TUM 数据集的动态序列实验表明,改进系统在 ATE 与 RPE 指标上较 ORB-SLAM3 均有提升,最大提升分别为 95.42%和 54.15%,单帧处理时间平均优化到 18.08 ms,验证了其在动态干扰下的精确性和鲁棒性;

4) 在起重机真实场景测试中,SLAM3-YOLOv8 系统轨迹闭合误差优于原系统,成功构建了无重影的稠密点云地图。实验证明,该系统能够满足起重机复杂动态环境下的高精度定位与场景建模需求,为工业重载设备的智能化升级提供了可行的技术方案。

笔者后续将针对模型进行轻量化设计,探索多传感器融合策略,以提升系统在极端动态场景下的适应性。

## 参考文献(References):

[1] 市场监管总局.市场监管总局关于 2024 年全国特种设备安全状况的通报[EB/OL].北京:国家市场监督管理总局,2024.  
[https://www.samr.gov.cn/tzsbj/qktb/tb/art/2025/art\\_1b31b22095744942af16cf24cec957ac.html](https://www.samr.gov.cn/tzsbj/qktb/tb/art/2025/art_1b31b22095744942af16cf24cec957ac.html).  
State Administration for Market Regulation. Report on National Special Equipment Safety Status in 2024[EB/OL]. Beijing: SAMR, 2024.  
[https://www.samr.gov.cn/tzsbj/qktb/tb/art/2025/art\\_1b31b22095744942af16cf24cec957ac.html](https://www.samr.gov.cn/tzsbj/qktb/tb/art/2025/art_1b31b22095744942af16cf24cec957ac.html).

[2] 郑宏远,卢宁,宋鹏程,等.智能塔式起重机关键技术

研究[J].机电工程,2023,40(3):435-443.

ZHENG Hong-yuan, LU Ning, SONG Peng-ceng, et al. Research on key technologies of intelligent tower cranes[J]. Journal of Mechanical & Electrical Engineering, 2023, 40(3): 435-443.

[3] 吕品,过佳珺,赖际舟,等.多光线追踪与高度特征约束的激光雷达动态点云检测方法[J].中国惯性技术学报,2024,32(10):1010-1017.

LYU Pin, GUO Jia-jun, LAI Ji-zhou, et al. Dynamic point cloud detection method for LiDAR based on multi-ray tracing and height feature constraints[J]. Journal of Chinese Inertial Technology, 2024, 32(10): 1010-1017.

[4] 柏井良明,王棣棠.机器人工程[M].北京:科学出版社,2001.

KASHIWAI Y, WANG Di-tang. Robotics Engineering[M]. Beijing: Science Press, 2001.

[5] ZHENG Zeng-rui, LIN Shing-feng, YANG Chen-guang. RLD-SLAM: a robust lightweight VI-SLAM for dynamic environments leveraging semantics and motion information[J]. IEEE. Transactions on Industrial Electronics, 2024, 71(11): 14328-14338.

[6] 程董董.智能化起重机同步定位与地图构建方法研究[D].郑州:河南工业大学机电工程学院,2024.

CHENG Dong-dong. Research on Synchronous Localization and Mapping Method for Intelligent Crane[D]. Zhengzhou: College of Mechanical and Electrical Engineering, Henan University of Technology, 2024.

[7] LIU Tian-wei, LUO Jing-wen. Semantic-aware dynamic scene tracking and reconstruction for mobile robots leveraging visual SLAM[J]. The Journal of Supercomputing, 2025, 81(7): 830.

[8] 李泳,刘宏杰,周永录,等.室内动态场景下基于语义关联的视觉 SLAM 方法[J].计算机应用研究,2024,41(8):2528-2532.

LI Yong, LIU Hong-jie, ZHOU Yong-lu, et al. Visual SLAM method based on semantic association in indoor dynamic scenes[J]. Application Research of Computers, 2024, 41(8): 2528-2532.

[9] 符强,曾凡治,纪元法,等.针对室内高低动态环境的视觉 SLAM 算法研究[J].电子测量技术,2024,47(21):159-167.

FU Qiang, ZENG Fan-zhi, JI Yuan-fa, et al. Visual SLAM algorithm for indoor High-Low dynamic environments[J]. Electronic Measurement Technology, 2024, 47(21): 159-167.

[10] BESCOS B, FACIL 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.

[11] CHAO Yu, LIU Zu-xin, LIN Xin-jun, et al. DS-SLAM: A Semantic Visual SLAM Towards Dynamic Environments[C]// 2018 IEEE. /RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE., 2018: 1168-1174.

[12] ZHONG Fang-wei, WANG Sheng, ZHANG Zi-qi, et al. Detect-SLAM: Making Object Detection and SLAM Mutually Beneficial[C]// 2018 IEEE. Winter Conference on Applications of Computer Vision (WACV). IEEE., 2018: 1001-1010.

[13] LIU Wei, ANGUELOV D, ERHAN D, et al. SSD: Single Shot Multi-Box Detector[C]// European Conference on Computer Vision. Cham: Springer International Publishing, 2016: 21-37.

[14] WU Wen-xin, GUO Liang, GAO Hong-li, et al. YOLO-SLAM: a semantic SLAM system towards dynamic environment with geometric constraint[J]. Neural Computing and Applications, 2022, 34(8):1-16.

[15] YU Ya-ting, ZHU Kai, YU Wang-shui. YG-SLAM: GPU-accelerated RGBD-SLAM using YOLOv5 in a dynamic environment[J]. Electronics, 2023, 12(20): 4377.

[16] CONG Pei-chao, LI Jia-xing, LIU Jun-jie, et al. SEG-SLAM: dynamic indoor RGB-D visual SLAM integrating geometric and YOLOv5-based semantic information[J]. Sensors, 2024, 24(7): 2102.

[17] CAMPOS C, ELVIRAR, RODRIGUEZ 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.

[18] LI Yan-ke, SHEN Hua-bao, FU Ya-ping, et al. A method of dense point cloud SLAM based on improved YOLOv8 and fused with ORB-SLAM3 to cope with dynamic environments[J]. Expert Systems with Applications, 2024, 255(12): 4918.

[19] 翟亚红,陈雅玲,徐龙艳,等.改进 YOLOv8s 的轻量级无人机航拍小目标检测算法[J].浙江大学学报:工学版,2025,59(8):1-10.

ZHAI Ya-hong, CHEN Ya-ling, XU Long-yan, et al. Lightweight small target detection algorithm for UAV aerial photography based on improved YOLOv8s [J]. Journal of Zhejiang University : Engineering Science, 2025, 59(8): 1-10.

[20] CHU Guo-ming, YAN Peng, LUO Xu-hong, et al. YG-SLAM: enhancing visual SLAM in dynamic environments with YOLOv8 and geometric constraints[J]. IEEE. Access, 2025, 36(4):1-18.

[21] 周立君,刘宇,白璐,等.使用 TensorRT 进行深度学习推理[J].应用光学, 2020,41 (2):337-341.

ZHOU Li-jun, LIU Yu, BAI Lu, et al. Deep learning inference using TensorRT[J]. Journal of Applied Optics, 2020, 41(2): 337-341.

[22] 李柯泉,陈燕,刘佳晨,等.基于深度学习的目标检测算法综述[J].计算机工程,2022,48(7):1-12.

LI Ke-quan, CHEN Yan, LIU Jia-chen et al. A review of object detection algorithms based on deep learning[J]. Computer Engineering, 2022, 48(7): 1-12.

[23] YANG Nan, WANG Yong, ZHANG Lei, et al. Point cloud semantic segmentation network based on graph convolution and attention mechanism[J]. Engineering Applications of Artificial Intelligence, 2025, 141(2): 109790.

[24] 陈明方,黄良恩,王森,等.移动机器人视觉里程计技术研究综述[J].农业机械学报, 2024,55(3):1-20.

CHEN Ming-fang, HUANG Liang-en, WANG Sen, et al. A review of visual odometry technologies for mobile robots[J]. Journal of Agricultural Machinery, 2024, 55(3): 1-20.

[25] STURM J, ENGELHARD N, ENDRES F, et al. A Benchmark for the Evaluation of RGB-D SLAM Systems[C]// 2012 IEEE. /RSJ International Conference on Intelligent Robots and Systems. IEEE., 2012: 573-580.