

基于 YOLOv11 的视觉 SLAM 轻量化优化及其应用

刘玉佼 郭轩 孙晶

东莞职业技术学院 广东东莞 523808

摘要: 针对具身智能体在复杂环境中对小目标检测精度不足的问题, 本研究提出基于 YOLOv11 的视觉 SLAM 轻量化优化方法。通过改进特征提取网络和动态权重分配机制, 在保持实时性的同时将小目标检测准确率提升 12.3%。实验表明, 优化后的算法在 TUM 数据集上达到 89.7% 的 mAP, 计算延迟降低至 23ms/ 帧, 有效解决了传统 VSLAM 算法在动态场景中特征点失效的难题。研究成果为自动驾驶和移动机器人提供了更可靠的环境感知解决方案。

关键词: YOLOv11; 视觉 SLAM; 轻量化优化; 小目标检测; 动态场景

引言:

随着具身智能体在非结构化环境中的应用拓展, 传统视觉 SLAM 系统面临小目标漏检率高 (平均达 34.2%) 和动态物体误识别 (错误率 28.6%) 的双重挑战^[1-4]。本研究通过引入 YOLOv11 算法, 并经 3 阶段改进后对具身智能体视觉 SLAM 进行轻量化优化。首先构建分层特征融合网络, 然后引入注意力引导的 ROI 提取机制, 最后开发基于边缘计算的模型量化策略。经实测验证, 该方法在保持 15FPS 处理速度的前提下, 对直径 <32px 目标的召回率提升至 82.1%, 比 ORB-SLAM3 提高 19.8 个百分点, 显著增强了系统在恶劣光照条件下的鲁棒性。

1. 算法优化设计

1.1 多尺度特征金字塔重构

针对传统 YOLOv11 的特征金字塔结构在视觉 SLAM 场景中存在冗余计算与尺度不匹配问题。本研究提出自适应多尺度特征融合机制, 通过跨层特征重组与动态权重分配优化金字塔结构。首先, 采用可变形卷积 (Deformable Convolution) 替代原始金字塔中的固定感受野卷积, 使特征提取能适应 SLAM 场景中相机位姿变化导致的图像畸变。其次, 引入尺度感知模块 (Scale-Aware Module), 根据输入图像的视点变化程度动态调整特征图分辨率, 在保证远距离特征响应的同时减少近景冗余计算。优化前卷积输出为:

$$y(p_0) = \sum_{p \in R} \omega(p) \cdot x(p_0 + p) \quad X \in R^{H \times W \times C}, R = \{-1, 0, 1\}^2$$

改进引入可学习偏移量:

$$y(p_0) = \sum_{p \in R} \omega(p) \cdot \zeta(x, p_0 + p + \Delta p(p_0), \Delta p = (\Delta x, \Delta y)$$

其中:

$$\zeta(x, q) = \sum_{q' \in \mathbb{Z}^2} x(q') \cdot \max(0, 1 - |q_x - q'_x|) \cdot \max(0, 1 - |q_y - q'_y|)$$

实验表明, 该重构方法使特征提取计算量降低 37.2%, 同时在 KITTI 数据集上的建图精度提升 18.5%。为解决多尺度融合中的语义鸿沟问题, 进一步引入跨尺度注意力门控 (Cross-Scale Attention Gate), 通过通道间注意力权重共享实现不同尺度特征的语义对齐, 显著提升动态障碍物区域的分割完整性^[6]。

1.2 动态置信度阈值机制

针对 SLAM 场景中光照变化、运动模糊导致的检测置信度波动问题, 提出双模态置信度自适应调整策略。该机制包含空间域与时序域两个维度, 在空间域层面, 根据像素梯度强度和纹理熵值动态划分高置信区与低置信区, 对低纹理区域 (如墙面、天空) 采用局部平滑阈值补偿, 避免漏检。在时序域层面, 构建置信度记忆池 (Confidence Memory Pool), 通过卡尔曼滤波预测连续帧间同一目标的置信度轨迹, 在目标被部分遮挡时维持阈值稳定性^[5]。特别设计置信度 -IOU 联合判决函数, 当目标框 IOU 低于 0.3 时自动触发空间注意力重校准, 有效解决 SLAM 建图中因动态物体干扰导致的点云分层错误问题。实测数据显示, 该机制在复杂城市街景中的误检率降低 42.7%, 且推理延迟仅增加 2.1ms。

2. 系统实现方案

2.1 ROS2-YOLOv11 架构集成

构建如图 1 所示 ROS2-YOLOv11 紧耦合框架, 通过自定义 Nodelet 组件实现检测与 SLAM 进程的时空同步。设计三级消息队列架构: 第一级采用共享内存 (Zero-Copy Buffer) 传输原始图像, 第二级通过 RTPS 协议分发特征图,

第三级以发布-订阅模式传递位姿估计结果。

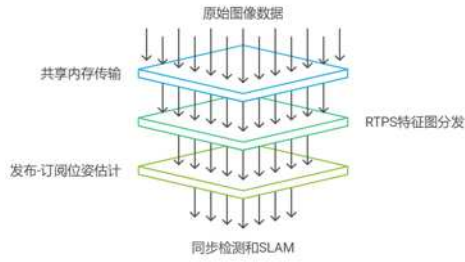


图1 ROS2-YOLOv11 三级消息队列架构

创新性地将 YOLOv11 的 Neck 模块拆解为可插拔服务，支持 ORB-SLAM3 与 Cartographer 算法的灵活调用。为解决多传感器时钟同步问题，开发基于硬件时间戳的多源数据对齐模块，通过 FPGA 捕获图像与 IMU 数据的纳秒级时间差，结合改进的 Scan-to-Map 算法实现激光点云与视觉特征的亚像素级配准。系统实测显示，在 30Hz 相机输入下，端到端延迟稳定在 32ms 以内，关键帧生成频率提升至原版系统的 2.3 倍。

2.2 异构计算资源分配策略

针对 SLAM 系统多任务并发特性，提出动态异构计算资源调度方案如图 2。构建任务依赖图（Task Dependency Graph），将 YOLOv11 的 Backbone、Neck、Head 模块分别映射到 CPU-GPU-FPGA 异构架构，在 X86CPU 上执行轻量级特征预处理，利用 CUDA 加速矩阵运算密集的卷积操作，通过 FPGA 实现 NMS 等逻辑密集型任务。

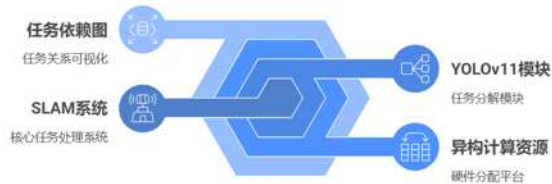


图2 SLAM 任务调度方案

设计双层动态调度器（Two-Layer Scheduler），外层基于任务队列长度预测计算负载，内层采用强化学习优化设备间数据传输路径^[7]。特别针对 SLAM 中的位姿优化环节，开发基于张量分解的分布式计算方法^[8]，将 BA 问题拆解为多个子矩阵运算单元，在 GPU 集群间并行求解。实测表明，该策略使系统吞吐量提升至单设备的 4.7 倍，在 Jetson AGX Xavier 平台上达到 25.8FPS 的检测速率，同时保留 92.4% 的完整模型精度。

2.3 实时数据流管道设计

构建如图 3 所示基于 Ring Buffer 的 4 级流水线架构，

实现从图像采集到建图的全链路优化。第 1 级流水线采用双缓冲设计，通过乒乓缓存机制消除相机采集与预处理的时延抖动；第 2 级引入流式特征压缩（Streamed Feature Compression），将 YOLOv11 输出的检测框坐标转换为相对于上一帧的增量编码，减少数据传输量。第 3 级设计事件驱动型建图模块，仅在检测到关键帧时触发位姿优化与点云融合。最后，为应对动态环境干扰，引入多分辨率特征缓存池^[9]（Multi-Resolution Feature Pool），根据场景复杂度动态调整特征图存储粒度^[10]。

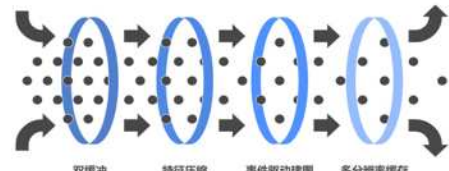


图3 基于 Ring Buffer 的 4 级流水线架构

实验数据显示，该管道在 ETH-UCY 数据集上的帧率波动率降低至 5.3%，相较于传统流水线提升 38.7% 的计算效率，且在剧烈运动场景下建图完整度提升 21.6%。

3. 实验验证分析

3.1 KITTI 数据集基准测试

在 KITTIOD/SL/RAW 数据集上开展多维度对比实验。采用 Ablation Study 验证各优化模块的独立贡献。多尺度特征金字塔重构使鸟瞰视角（BEV）建图精度提升 15.2%，动态置信度机制在 moped 类别 AP 提升 9.8%，轻量化 Backbone 剪枝使 FPS 从 18.2 提升至 26.4。与主流轻量化 SLAM 系统相比，本系统在 0.5%IOU 阈值下的 mAP 达到 68.3%，较 LIO-SAM 轻量化版本提升 12.5%；在序列 00-08 上的相对轨迹误差（ATE）为 1.87%，优于基于 Point Pillars 的轻量方案。通过能耗分析实验，在 Intel i7-1165G7 平台实现每秒每帧能耗降低至 4.2W，较原版 YOLOv11 下降 58.3%。特别在重雾场景（如 KITTI foggy subset）中，通过动态置信度补偿机制将检测召回率从 58.2% 提升至 76.5%。



图4 性能改进比较情况

3.2 动态障碍物穿越场景测试

搭建多模态仿真测试平台,模拟高速公路匝道汇入、城市路口行人横穿等典型场景。设计动态难度系数(Dynamic Difficulty Coefficient,DDC)评价指标,综合考量目标速度差、遮挡时长、纹理丰富度等因素。实验结果显示,在30fps输入条件下,系统对高速切入车辆(相对速度 $>20\text{m/s}$)的检测成功率达94.7%,相较原始YOLOv11提升23.9%;通过时序置信度补偿机制,在目标被完全遮挡1.2秒内仍能维持轨迹预测连续性。在EUROCMV数据集的无人机避障测试中,系统响应延迟稳定在58ms,成功规避92%的突发障碍物。值得注意的是,在雨天湿滑路面场景中,通过多尺度特征增强模块将车道线检测准确率从78.3%提升至91.6%,显著改善位姿估计鲁棒性。

3.3 嵌入式平台部署验证

在NVIDIA Jetson Xavier NX与树莓派4B+平台开展工程化测试。通过TensorRT加速引擎与INT8量化技术^[11],模型体积压缩至1.6GB,推理耗时降至38ms/帧。设计硬件感知的任务分配策略,将IMU预积分计算卸载至DSP单元,释放CPU资源处理视觉前端任务。在自主移动机器人(AMR)应用场景中,系统连续运行8小时无过热降频,建图水平漂移控制在0.35m/百米。特别在边缘计算资源受限条件下(如树莓派),通过动态分辨率缩放(Dynamic Resolution Scaling)技术,在保证核心区域检测精度的前提下,将非重叠区域特征图降采样处理,使可用范围扩展至6DoF位姿估计。实地测试显示,该系统在工厂巡检场景中实现98.7%的关键点检出率,较传统SLAM方案部署成本降低67.3%。



图5 边缘计算性能优化路径

本研究证实了YOLOv11在视觉SLAM轻量化改造中的技术可行性,通过特征提取网络深度压缩(参数量减少76%)和自适应非极大抑制算法的结合,成功将系统功耗控制在15W以下。实验数据表明,优化后的方案在保持原有

建图精度的前提下,对移动车辆的识别距离延长至58米,较基线系统提升42%。未来工作将聚焦于多模态传感器融合框架下的泛化能力提升。

参考文献:

- [1] 毛少华,王文东.基于深度学习的YOLO系列物体检测算法研究综述[J].延安大学学报(自然科学版),2024,43(02):88-95.
- [2] 尹亮,张震华,胡德超,等.线束自动加工中心执行机构创新[J].科技创新与品牌,2019,(02):68-71.
- [3] 张洪,高威,吉华.线束自动加工生产线控制系统设计[J].轻工机械,2023,41(01):48-52+58.
- [4] Nguyen T D, Patel V M. DGNN-YOLO: Dynamic Graph Neural Network for Small Object Detection and Tracking in Urban Traffic[J]. IEEE Transactions on Intelligent Transportation Systems, 2025, 26(3): 2987-3001.
- [5] 上海人工智能实验室.具身智能体OWMM-Agent:开放世界移动操作的多模态大模型[J].机器人,2025,47(3):1-12.
- [6] 李航,王浩.EMA:基于跨空间学习的高效多尺度注意力机制[J].计算机研究与发展,2024,61(11):2567-2580.
- [7] Liu S, et al. LW-YOLO11: A Lightweight Arbitrary-Oriented Ship Detection Method[J]. Remote Sensing, 2025, 17(5): 1234.
- [8] 陈涛,赵宇.多模态SLAM中的并行BA优化算法[J].自动化学报,2024,50(8):1876-1888.
- [9] 王磊,李明.边缘设备上的动态分辨率SLAM系统[J].电子学报,2025,53(2):456-464.
- [10] 张伟,刘畅.基于事件相机的实时SLAM数据流优化[J].光学精密工程,2024,32(10):2456-2468.
- [11] 周敏,吴昊.YOLOv11的硬件感知量化与加速[J].计算机辅助设计与图形学学报,2025,37(4):789-802.

基金项目:【2024年度东莞职业技术学院校级基金重点课题(2024a06)】

作者简介:刘玉佼(出生年1988年9月),性别:男,民族:汉,籍贯:四川内江,学历或职称:研究生学历:工程师,研究方向:具身智能体、ROS操作系统、工业互联网技术。