

融合实例分割与视觉惯性里程计的丘陵山地果园语义SLAM算法

李云伍^{1,2} 朱明辉¹ 凤柳燕¹ 李明生^{1,2} 曾超然¹ 邓秋萍¹ 徐常塑^{1,2}

(1. 西南大学工程技术学院, 重庆 400715; 2. 西南山地智能农机装备创新支撑中心, 重庆 400715)

摘要: 针对丘陵山地果园地势起伏导致的定位鲁棒性下降, 以及传统 SLAM 因地形语义与高度信息缺失难以精确引导农机在丘陵山地果园下自主导航的问题, 本文提出一种深度融合实例分割与视觉惯性里程计的语义 SLAM 算法。前端采用一种动态 IMU 预积分策略, 在运动剧烈工况下引入龙格库塔 (Runge - Kutta) 法捕捉高频运动特征, 增强位姿估计精度; 同步利用集成 CBAM 与 FPIoU - v2 的改进 YOLO v8 - seg 算法实现像素级语义提取, 并通过语义投影最终构建具备高度、坡度、粗糙度及语义属性等多维地形信息的 2.5D 语义高程地图。试验结果表明, 改进定位算法在不同丘陵果园实地测试环境下绝对位姿误差 (APE) 和均方根误差 (RMSE) 为 0.38/0.24 m, 优于 VINS 与 ORB - SLAM3; 改进 YOLO v8 - seg 算法在自建数据集上 mAP@0.5 达到 73.79%, 推理速度为 39.09 f/s, 满足实时检测要求。本文所提算法可提升系统定位和感知能力, 为果园履带车在非结构化环境下导航与作业提供理论与技术指导。

关键词: 丘陵山地果园; 改进 YOLO v8; 2.5D 语义高程地图; 视觉惯性里程计; 实例分割

中图分类号: S24

文献标识码: A

文章编号: 1000-1298(2026)20-0297-11

OSID:



Semantic SLAM Algorithm for Hilly and Mountainous Orchards Combining Instance Segmentation and Visual Inertial Odometry

Li Yunwu^{1,2} Zhu Minghui¹ Feng Liuyan¹ Li Mingsheng^{1,2} Zeng Chaoran¹ Deng Qiuping¹ Xu Changsu^{1,2}

(1. College of Engineering and Technology, Southwest University, Chongqing 400715, China

2. Southwest Mountain Regions Intelligent Agricultural Machinery Equipment Innovation Center, Chongqing 400715, China)

Abstract: Aiming to address the issues of reduced positioning robustness caused by the steep topography of hilly and mountainous orchards, as well as the difficulty of accurately guiding agricultural machinery for autonomous navigation in such environments due to the lack of terrain semantics and elevation data in traditional SLAM, a semantic SLAM algorithm that deeply integrated instance segmentation with visual inertial odometry was proposed. In the frontend, a dynamic IMU pre-integration strategy was adopted, which introduced the Runge - Kutta method under intense motion conditions to capture high-frequency motion features, thereby enhancing the accuracy of pose estimation. Simultaneously, an improved YOLO v8 - seg algorithm integrated with CBAM and FPIoU - v2 loss function was utilized for pixel-level semantic extraction. A 2.5D semantic elevation map containing multi-dimensional terrain information, including height, slope, roughness, and semantic attributes was finally constructed through semantic projection. Experimental results indicated that the root mean square error (RMSE) of the absolute pose error (APE) for the improved localization algorithm in different hilly orchard field tests was 0.38 m and 0.24 m, respectively, which outperformed VINS and ORB - SLAM3. The improved YOLO v8 - seg algorithm achieved an mAP@0.5 of 73.79% on a self-built dataset with an inference speed of 39.09 f/s, satisfying real-time detection requirements. The proposed algorithm enhanced the localization and

收稿日期: 2026-03-16 修回日期: 2026-05-13

基金项目: 中央高校基本科研业务费专项资金项目 (SWU - KF25009)

作者简介: 李云伍 (1974—), 男, 教授, 博士生导师, 主要从事农业信息感知与自主导航技术研究, E-mail: liyunwu@swu.edu.cn

通信作者: 徐常塑 (1995—), 男, 讲师, 博士, 主要从事智慧农业技术与装备研究, E-mail: xuchangsu@swu.edu.cn

perception capabilities of the system, providing theoretical and technical guidance for the navigation and operation of tracked orchard vehicles in unstructured environments.

Key words: hilly and mountainous orchards; improved YOLO v8; 2.5D semantic elevation map; visual inertial odometry; instance segmentation

0 引言

丘陵山地果园是全球水果生产体系的关键组成部分^[1]。由于 GNSS 信号不稳定、剧烈的地势起伏和多变的地形特征,农机自主导航技术在丘陵山地发展缓慢^[2-4]。SLAM 技术作为自主导航的基础,可解决丘陵山地环境下 GNSS 信号拒止的问题,对环境感知和高精度定位具有更高的要求^[5-7]。

视觉 SLAM 因感知信息丰富、硬件成本低等优势,成为复杂环境下环境感知的首选,但现有视觉 SLAM 算法在定位精度与鲁棒性方面表现不足。对此,众多学者在里程计设计方面开展了大量研究。Campos 等^[8]提出了一种基于特征点法的视觉惯导紧耦合 SLAM 系统 ORB-SLAM3,显著提高了室外大场景下定位精度。陈久朋等^[9]采用神经网络对输入图像进行预处理,并使用地图点补充策略、Sparse BA 和实时增量闭环检测方法,提高了定位与建图精度和鲁棒性。Xu 等^[10]通过引入 YOLO v8 进行动态目标检测,实现了城市动态环境下高精度与鲁棒性,但在重复性特征较多的复杂环境中鲁棒性仍然有限。董蕊芳等^[11]融合环境 3D、2D 占据特征及路标位置等信息,提出基于 ORB-SLAM2 的多图层 SLAM 地图定位,实现了森林公园导航。在地势起伏剧烈且计算资源受限的丘陵山地果园中,传统纯几何 SLAM 面临环境特征高度重复引发的匹配失效,以及位姿估计鲁棒性受地形干扰等局限,引入语义信息的 SLAM 系统已成为农机智能化必然趋势。

近年来,语义信息开始被引入到 SLAM 框架中^[12-13]。Bescos 等^[14]提出 DynaSLAM,通过结合多视图几何与深度学习以进行动态目标检测和背景修复。Lu 等^[15]集成 Mask R-CNN 以剔除动态对象干扰。Wang 等^[16]将 SCTNet-B 嵌入 ORB-SLAM2,生成了辅助导航的语义八叉树地图。Li 等^[17]基于 ORB-SLAM2 框架,提出实现语义立体 V-SLAM 算法,消除环境中运动物体影响,提高运动轨迹计算精确性。现有通用语义 SLAM 方案多聚焦于路标识别与障碍物标注,但在缺乏清晰边界的丘陵山地果园中,通用方案难以辨识树、草地等对农机通过性评估至关重要的地形特征。

针对果园环境下 SLAM 难题,国内外学者开展了大量的算法优化与应用探索^[18-21]。目前针对果

园环境的 SLAM 多利用语义信息提升定位性能与障碍物检测效果,侧重于可行驶区域识别,忽略了对地形特征的深度挖掘。

综上,针对丘陵山地果园 SLAM 方案存在以下问题:纯几何 SLAM 依赖几何特征,在视觉特征重复的果园场景中易发生匹配失效,位姿估计受地势起伏干扰严重,且所建地图缺乏环境特征,难以支撑农机自主导航决策;通用语义 SLAM 多面向结构化场景设计,难以辨识丘陵山地果园中边界模糊的地形特征,制约了农机自主导航执行避障决策性能;果园专用 SLAM 忽略了对坡度、粗糙度等地形特征语义的深度挖掘,且缺乏高度维度信息支撑,致使农机在丘陵山地果园环境中缺乏可靠的导航依据。

本文提出一种深度融合实例分割与视觉惯性里程计的语义 SLAM 算法。针对农机在地势起伏环境中作业时振动剧烈,前端设计动态 IMU 预积分策略,削弱振动对位姿估计干扰。为了解决传统地图缺乏高度与环境信息、三维地图计算资源开销较大的问题,基于改进 YOLO v8-seg 模型提取地形语义信息,并通过语义投影构建融合地形语义、高度、坡度及粗糙度 2.5D 语义高程地图。

1 试验平台

传感器布置在履带式农用机械上组成移动试验平台,如图 1 所示,相关参数如表 1 所示。GNSS 2 根天线分别安装在车辆前端和后端,其连线穿过车身质心,能够可靠地估计航向,并提高基于 RTK 地面实测数据的鲁棒性。IMU 和相机被安置在车身质心顶部附近,以减少车辆振动和动态干扰的影响,并确保在树冠茂密、地形崎岖的果园环境中,信号接收不会受到障碍物阻挡。

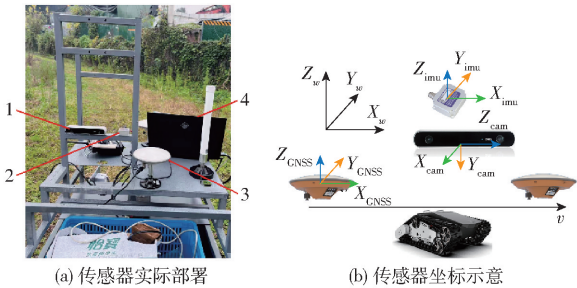


图 1 自主搭建的移动试验平台

Fig.1 Self-built mobile experimental platform
1. 相机 2. IMU 3. GNSS 天线 4. 控制终端

表 1 移动试验平台规格

Tab. 1 Specifications of mobile platform		
设备	参数	数值
LPMS - IG1 - RS232 型 IMU	采样频率/Hz	200
	采样频率/Hz	15
ZED2 型相机	图像尺寸/(像素 × 像素)	1 280 × 720
CGI410 型 GNSS	定位精度/m	0. 01、0. 4、1. 2
Autolabor PC 型控制终端	CPU:AMD Ryzen 7 5800H	/
	GPU:GEFORCE RTX 3060	
Komodo - 02 型履带式底盘	速度/(m·s ⁻¹)	0 ~ 1. 75

2 数据采集

在重庆市 2 处具有代表性的非结构化果园进行数据采集工作。采集场地位于合川区柑橘采摘基地(29°56′02.6″N,106°18′17.6″E)和北碚区果园(29°45′49.1″N,106°22′18.2″E)。构建的原始数据集包含 800 幅高质量图像样本,数据集分布情况如表 2 所示。利用 Roboflow 工具对所有样本进行精细的像素级语义标注,依据果园实际作业需求,将场景划分为 6 个核心语义类别,如图 2 所示。并针对丘陵果园中光照波动问题,对图像执行包括亮度变换、色彩抖动、旋转、裁剪和高斯噪声等数据增强策略,如图 3 所示。预处理策略在有效抑制模型过拟合风险的同时,增强了算法在丘陵山地果园环境下泛化能力与分割精度。最后将数据集按比例 8:1:1 随机划分为训练集、验证集与测试集。

表 2 试验数据集分布情况统计

Tab. 2 Statistics on distribution of experimental dataset			
维度	特征类型	图像数量	图像占比/%
语义类别	草地	659	82. 4
	普通道路	487	60. 9
	泥泞	132	16. 5
	果树	693	86. 6
	树基	516	64. 5
	障碍物	533	66. 6
光照条件	正常光照	400	50. 0
	逆光	400	50. 0
遮挡程度	轻微遮挡	500	62. 5
	严重遮挡	300	37. 5

3 丘陵山地果园定位与语义建图方法

3.1 系统框架

本文设计了一种融合实例分割与视觉惯性里程计的语义 SLAM 算法,如图 4 所示。在前端处理中,通过 IMU 动态预积分策略获得关键帧间位姿变换

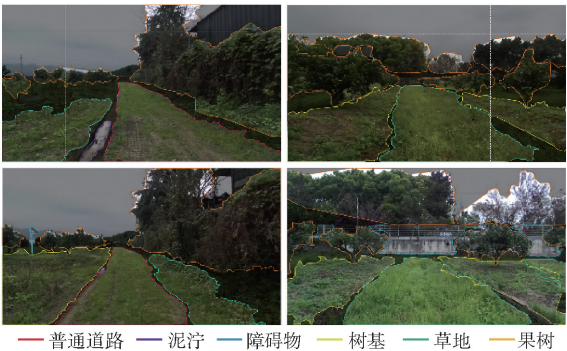


图 2 果园环境实例分割示意图

Fig. 2 Schematic of orchard environment segmentation

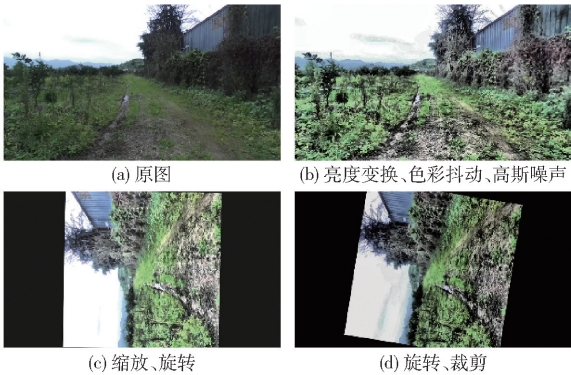


图 3 数据增强示意图

Fig. 3 Schematics of data augmentation

先验,并构造 IMU 残差因子;视觉支路采用 KLT 光流法追踪角点特征,利用对极几何约束估计系统位姿,实现高频的关键帧位姿状态更新。同步应用改进 YOLO v8 - seg 对图像进行实例分割,提取果树、果园常见地形等目标的像素级掩码。在后端融合阶段,通过 2D - 3D 语义投影模块,结合相机深度信息将 2D 掩码映射至三维空间,并利用 EKF 最小化重投影残差以修正状态估计。最终,通过概率递推更新各栅格高度均值与语义标签,并利用主成分分析算法提取坡度与粗糙度信息,构建具备多维地形属性的 2. 5D 语义高程地图。

3.2 动态预积分策略

IMU 预积分是系统中前端中心,为系统提供位姿约束信息。针对丘陵山地果园地面不平的场景,IMU 信号在短时间内会发生剧烈变化。预积分算法精度取决于对载体运动状态的建模精度,常用的中值法假设角速度和加速度在采样时间步内线性变化,仅通过起始 2 点均值进行积分,计算开销小但在剧烈动态下误差较大。相比之下,龙格库塔(Runge - Kutta)法通过在时间步内进行多点采样与加权平均,能够以更高阶精度处理非线性运动特征,降低剧烈运动时积分漂移^[22]。因此,本研究提出一种动态 IMU 预积分策略,即针对 IMU 信号波动较大的动态场景采取四阶 Runge - Kutta 法,而 IMU 波动

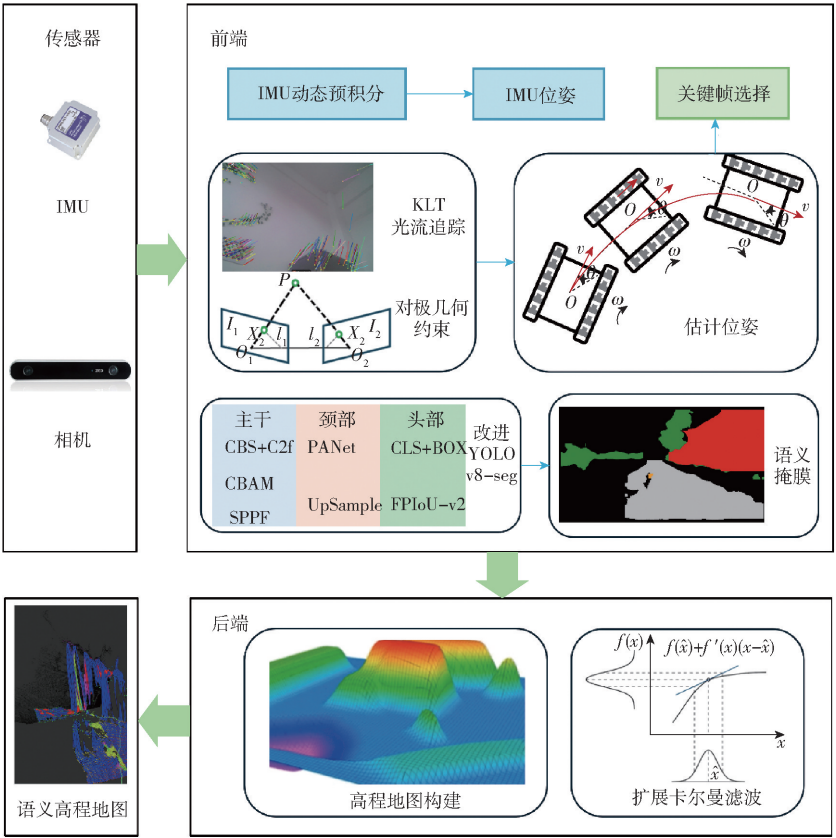


图4 算法框架

Fig.4 Structure of algorithm

较小的场景使用中值法。动态 IMU 预积分框架如图 5 所示。

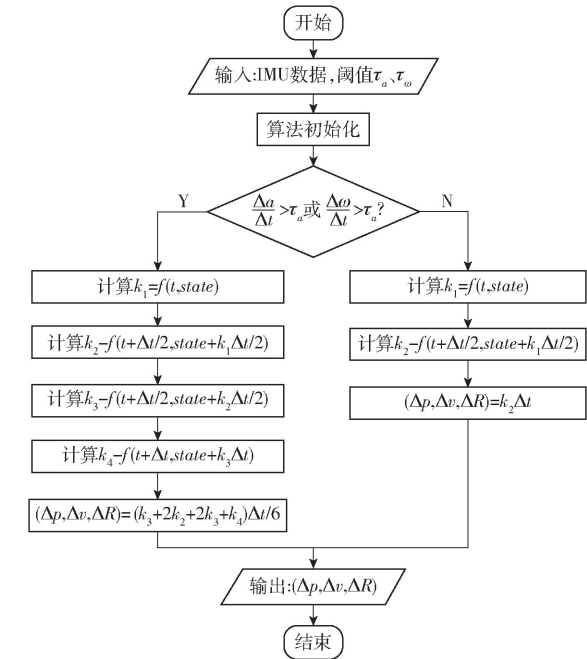


图5 动态 IMU 预积分流程图

Fig.5 Flowchart of dynamic IMU pre-integration

阈值 τ_a 和 τ_ω 的设定与 IMU 测量噪声标准差 σ 成正比,IMU 测量噪声模型为高斯噪声,因此比例系数 k 遵循 3σ 准则取值为 3, 阈值 τ_a 和 τ_ω 取值为 0.1 和 0.05 以确保正常的传感器噪声不会触发动

态状态检测,即

$$\tau = k\sigma \tag{1}$$

对 IMU 载体进行运动模型构建,并使用动态 IMU 预积分与 IMU 运动模型预测值实现残差构造。

针对动态切换在边界处可能引起的数值不稳定性问题,由于中值法与四阶 Runge - Kutta 法均基于相同的 IMU 运动学方程,且在切换瞬间位姿、速度等状态量均严格继承上一时刻估计结果,因此在理论层面保证了状态空间轨迹连续性。

3.3 实例分割算法改进

为赋予局部高程地图语义属性并兼顾前端计算效率,通过在 YOLO v8 - seg 主干网络的 SPPF 模块前集成 CBAM 注意力机制,增强模型对目标特征的聚焦能力,并采用 FP1oU - v2 损失函数强化边界回归过程,算法框架如图 6 所示。

该网络由特征提取主干、多尺度特征融合颈部以及解耦检测分割头构成^[23]。主干网络采用基于跨阶段局部网络思想的 C2f 模块提取多尺度特征,并在 SPPF 模块前引入 CBAM 注意力机制。针对丘陵果园环境下枝叶繁茂、光影剧烈波动导致的背景噪声干扰问题,本算法在主干网络集成 CBAM 注意力机制。通过通道与空间双重注意力的权重再分配,CBAM 能够强化网络对树基、泥泞地形等语义特征的提取能力,从而为不规则地形边缘精确分割奠定基础。

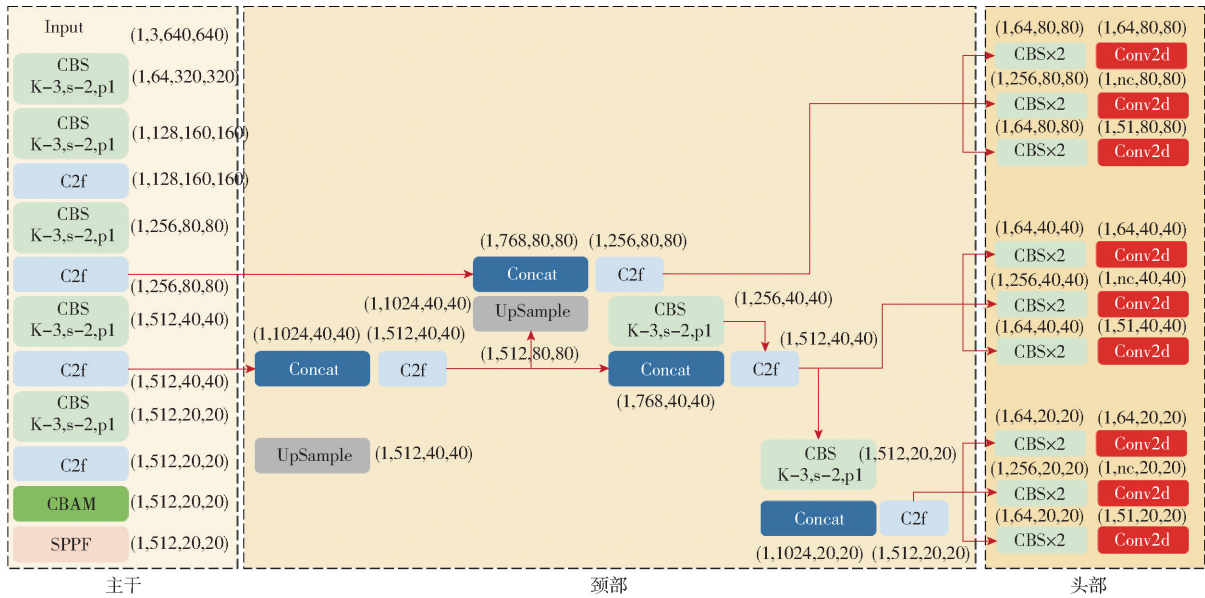


图 6 改进 YOLO v8-seg 算法框架

Fig. 6 Framework of improved YOLO v8-seg algorithm

颈部网络采用路径聚合网络 PANet 结构,利用上采样与特征级联算子实现深层语义信息与浅层空间细节的双向融合,确保掩码预测在复杂地形边界处的精确性。在输出端,解耦头将分类与边界框回归任务分离,并利用掩码分支并行生成全局掩码原型与实例掩码系数。

在边界框回归损失的设计上,针对果园环境下目标尺度差异大、相互遮挡导致回归不稳定的问题,引入 FPIoU-v2 损失函数进行优化。相比于传统的损失函数,FPIoU-v2 通过引入聚焦中心点约束机制与动态尺度惩罚因子,进一步增强回归过程收敛性与精确性。损失函数计算式为

$$\mathcal{L}_{\text{FPIoU-v2}} = 1 - \text{IoU} + \frac{\rho^2(b, b^{\text{gt}})}{c^2} + \alpha v + \gamma(1 - \text{IoU})^\lambda \quad (2)$$

式中 IoU ——预测框 b 与真实框 b^{gt} 交并比
 $\rho^2(b, b^{\text{gt}})$ ——两框中心点欧氏距离
 c ——覆盖两框最小闭合区域对角线长度
 v ——因子动态调整长宽比惩罚项
 α ——权重系数
 γ ——附加惩罚项权重超参数
 λ ——幂次因子

FPIoU-v2 核心改进在于通过聚焦因子动态调整长宽比惩罚项权重系数,特别在预测框与真实框重叠度较低的初始阶段,能强制拉引预测框向目标中心点快速聚集。相比于目前目标检测中主流的 CIoU 与 SIoU 损失函数:与 CIoU 对比,CIoU 虽考虑重叠面积与长宽比,但在预测框与目标重叠度较低的初始阶段,约束性能受限。由于果园内不规则几何特征较多,FPIoU-v2 通过引入聚焦因子,强化预

测框向目标中心回归的趋势,从而提升边界框收敛效率。与 SIoU 对比:SIoU 虽试图通过引入角度代价以降低回归自由度,但在地势剧烈起伏、目标排布杂乱的非结构化果园中,角度约束增益表现出明显的波动性,且易诱发梯度扰动。相比之下,采用 FPIoU-v2 损失函数能有效校正果树边缘像素在复杂地形投影中偏移偏差,提升边界框针对不规则边缘定位精度。

算法通过线性加权合成输出像素级二维语义掩码,并利用位姿初值将其投影至三维空间,从而为后端 2.5D 高程地图赋予语义属性先验与标签,该信息目前仅作用于地图构建层,不参与前端 VIO 重投影误差计算和状态估计反馈。

3.4 高程地图构建与语义高程地图生成

综合利用 ZED2 相机深度信息与 VIO 位姿信息,将 2D 语义掩码投影至局部坐标系下。

为消除物体边缘及动态噪声引起的离群点,采用 SOR 滤波器进行清洗^[24]。计算点云中任意一点 p_i 到其 k 个最近邻点平均欧氏距离 $\bar{d}_i$,计算式为

$$\bar{d}_i = \frac{1}{k} \sum_{j=1}^k \|p_i - p_j\| \quad (3)$$

式中 p_i, p_j ——当前查询点及其邻域内的第 j 个近邻点三维坐标向量

k ——选定的最近邻点数量,取 30

统计整个点云集合中所有点平均距离均值 μ 和标准差 σ 。根据高斯分布 3σ 准则,定义离群点判断阈值 D_{thresh} 为

$$D_{\text{thresh}} = \mu + \alpha_1 \sigma \quad (4)$$

式中 α_1 ——标准差倍数系数

若某点的局部平均距离 $\bar{d}_i$ 大于阈值 D_{thresh} , 则在统计学上判定该点远离了主体点云分布, 属于漂浮噪声并予以剔除。

将局部空间划分为精细的栅格单元 i , 并进行高程更新, 即

$$\hat{h}_i = \frac{\sum_{k \in I} h_k \omega_k}{\sum_{k \in I} \omega_k} \tag{5}$$

式中 $\hat{h}_i$ ——栅格 i 最终高度估值
 h_k ——第 k 次观测高度
 ω_k ——基于深度不确定性的权重

考虑果园地形特征尺度、ZED2 分辨率与计算资源等问题, 取栅格边长 d 为 0.12 m。

为了评估丘陵山地果园地形通过性, 系统应用主成分分析对栅格内的局部点云集 P_{local} 构造协方差矩阵 C , 即

$$C = \frac{1}{n} \sum_{j=1}^n (p_j - \bar{p})(p_j - \bar{p})^T \tag{6}$$

式中 n ——栅格内特征点总数
 $\bar{p}$ ——点集质心

系统通过对 C 进行奇异值分解提取法向量 N 。地形坡度 θ 定义为法向量 N 与世界坐标系垂直轴 z_{world} 的夹角, 计算式为

$$\theta = \arccos \frac{|Nz_{world}|}{\|N\| \|z_{world}\|} \tag{7}$$

此外, 引入地形粗糙度 σ_r 衡量路面平整度, 定义为局部邻域栅格内高度标准差, 计算式为

$$\sigma_r = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (h_i - \bar{h}_{local})^2} \tag{8}$$

式中 N ——局部领域内栅格总数
 h_i ——局部领域内栅格 i 的高度估值
 $\bar{h}_{local}$ ——局部领域内栅格高度平均值

最终, 高度 $\hat{h}_i$ 、坡度 θ 、粗糙度 σ_r 和语义实例信息共同构成了具备多维地形属性的 2.5D 语义高程地图。依据试验平台与场地, 设置坡度阈值为 15°, 粗糙度阈值为 0.15 m。

3.5 评价指标

为了验证所提算法综合性能, 从定位精度、语义分割效果、高程误差选取评价指标。

(1) 绝对位姿误差: 估计位姿与真实位姿直接差值, 可直观反映算法精度和全局轨迹一致性。通过 evo 评估工具得到 RMSE 评估 SLAM 映射准确性。

(2) 语义分割性能: 采用精确率 P 、召回率 R 、模型参数量、平均精度均值 (mAP@0.5) 和帧率等指标评价不同改进模块对原始模型的增益效果。

(3) 高程误差: 为定量评估所提方法在地形重建精度方面的实际性能, 利用高精度 RTK - GNSS 获取的实时海拔作为地面真值。通过提取履带车轨迹线上对应栅格高程估计值与 RTK 观测值, 计算高程 RMSE 及平均误差作为地形重建精度评价指标。

(4) 语义栅格一致率: 鉴于单一试验序列时长接近 400 s, 若对约 12 000 帧图像 2 个试验序列进行逐帧标注, 其成本极高且难以实现, 故引入语义栅格一致率指标。通过统计局部滑动窗口内, 同一栅格单元在连续多次观测中语义标签稳定性, 反映系统在剧烈颠簸下感知可靠性。

4 试验结果与分析

4.1 基于丘陵山地果园的定位精度试验

选定的柑橘园采摘区域场景丰富, 如图 7 所示。A 区包含下坡和路边坡地条件, B 区包含下坡和河边条件, C 区为普通果园行间道路, D 区包含逆光条件。参照文献[25], 采用“环形”参考路径模拟农机在丘陵山地果园中自动作业, 行驶速度为 0.5 m/s。



图 7 试验场景

Fig. 7 Test scenario

目前主流的开源语义 SLAM 多是基于 ORB - SLAM 或 VINS 等经典系统进行模块化扩展, 底层位姿估计机制具有高度同源性。语义地图构建质量与评价结果受限于特定语义类别设定及应用场景, 不同算法由于评价基准不一, 语义层面对比结果参考价值有限。参照文献[19, 26 - 27], 在评估定位精度时, 选取 ORB - SLAM3、ORB-SLAM2 算法作为对比对象。因此, 通过与目前定位应用最广的 VINS^[28] 和 ORB - SLAM3 进行对比, 能客观、准确地反映本算法在非结构化场景下位姿估计性能的实际提升效果。

对动态 IMU 预积分策略使用对比试验进行独

立验证,结果如表 3 所示。A、B 场景果园履带车行驶时长与总路程为 373.4 s 和 141.985 m,C、D 场景果园履带车行驶时长与总路程为 354.43 s 和 181.441 m。本文提出的动态 IMU 预积分策略相比于传统的中值法,在 A、B 场景果园中,绝对轨迹误差均方根误差从 0.44 m 降至 0.38 m;在 C、D 场景果园中则由 0.36 m 降至 0.24 m。相比于 Runge - Kutta 法,本文动态策略在 A、B 场景和 C、D 场景下

RMSE 仅分别增加 0.04、0.03 m,但平均 CPU 负载分别降低 20.68、21.95 个百分点,有效缓解了高阶积分带来的计算压力。验证了动态切换机制在复杂工况下捕捉高频运动特征的同时,能自适应地在平稳工况下降低算法复杂度。此外,试验结果表明,在具有多种场景下果园里,其定位误差大于地形相对简单的常规果园,符合算法在不同复杂度环境下的表现预期。

表 3 不同预积分方法在果园场景下试验结果
Tab.3 Experimental results of different pre-integration methods in an orchard setting

方法	A、B 场景果园			C、D 场景果园		
	平均误差/m	均方根误差/m	平均 CPU 负载/%	平均误差/m	均方根误差/m	平均 CPU 负载/%
原始中值法	0.41	0.44	243.98	0.30	0.36	233.89
Runge - Kutta 法	0.28	0.34	290.44	0.14	0.21	271.74
动态 IMU 预积分策略	0.32	0.38	269.76	0.17	0.24	249.79

丘陵山地果园试验结果如表 4 与图 8 所示。本文算法相比于 VINS 和 ORB - SLAM3 算法,在 A、B 场景中,平均误差和均方根误差分别降至 0.32、0.38 m,在 C、D 场景中进一步降至 0.17、0.24 m。相比于 ORB - SLAM3,本文算法在 2 类场景下绝对误差均值分别降低 20.0% 和 26.1%,绝对误差更加收敛,验证了本文在定位算法改进的有效性。此外,试验结果表明,在具有多种场景下果园(A、B 场景果园)里,其平均误差大于常见果园(C、D 场景果园),符合算法在不同复杂度环境下的表现预期。

表 4 各算法在果园场景下绝对轨迹误差
Tab.4 APE of each algorithm in orchard m

参数	A、B 场景果园			C、D 场景果园		
	VINS	ORB - SLAM3	本文算法	VINS	ORB - SLAM3	本文算法
平均误差	0.81	0.40	0.32	0.79	0.23	0.17
均方根误差	0.87	0.47	0.38	0.86	0.29	0.24

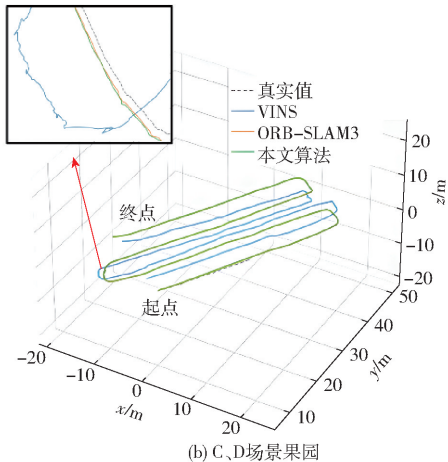
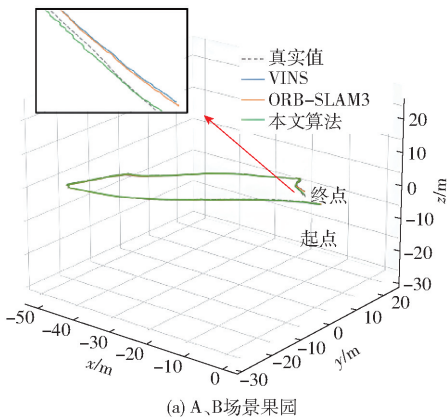


图 8 运行轨迹
Fig.8 Running trajectory

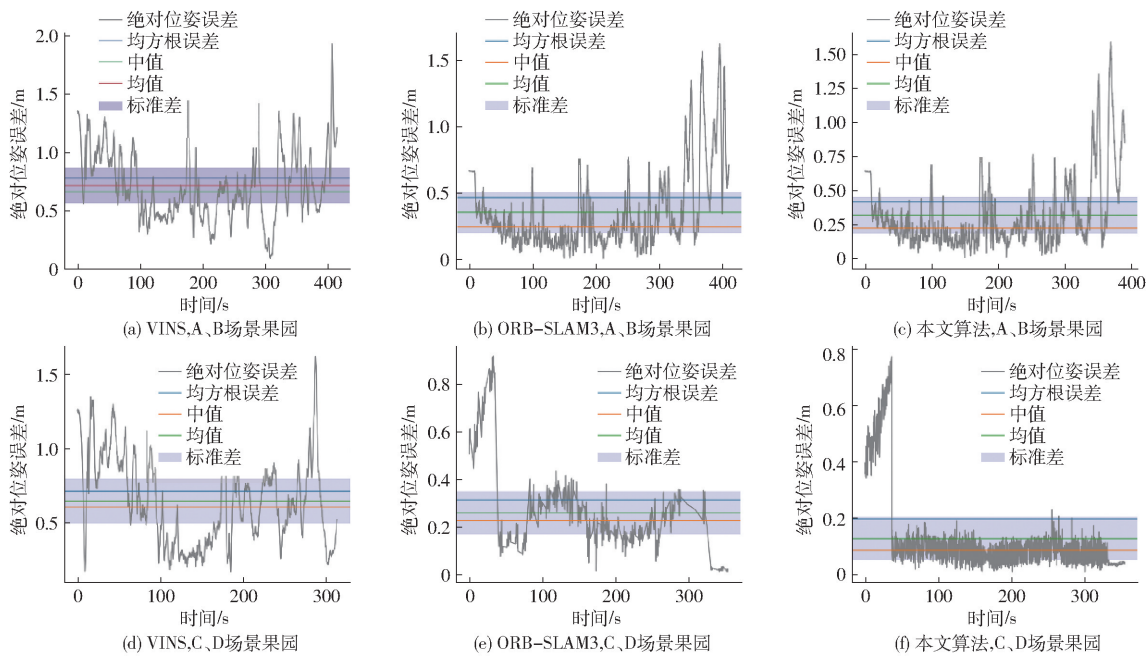


图9 绝对位姿误差

Fig.9 Map of APE

及模型参数量等指标,使用 YOLO v8n 原模型通过 FPIoU - v2、CIoU 和 SIoU 损失函数进行训练,评价其对模型增益效果。结果如表 5 所示。

表 5 不同损失函数试验结果

Tab.5 Experimental results for different loss functions				
损失函数	P/%	R/%	mAP@0.5/%	参数量
FPIoU - v2	79.22	70.10	70.74	3.26×10^6
CIoU	75.05	71.79	72.59	3.26×10^6
SIoU	77.46	67.72	70.31	3.26×10^6

由表 5 可知,在模型参数量均保持在 3.26×10^6 的轻量化水平下,不同损失函数对 YOLO v8 - seg 模型检测性能产生了不同程度的影响。采用 FPIoU - v2 损失函数的模型精确率表现最出色,达到 79.22%,相较于 CIoU 和 SIoU 分别提升 4.17、1.76 个百分点,证明了该算法在丘陵山地果园环境下具有更强的目标定位能力和误检抑制能力。尽管其 mAP@0.5 略低于 CIoU,但 FPIoU - v2 在提升预测精确性的同时,较好地兼顾了召回率与综合精度平衡。

为验证改进 YOLO v8 - seg 算法有效性,以 YOLO v8n 为基线模型,采用 R、P、模型参数量以及 mAP@0.5 等指标评价不同改进模块对原始模型的增益效果。在相同试验环境下依次添加 CBAM 和 FPIoU - v2 损失函数进行消融试验。消融试验结果如表 6 所示。

由表 6 可知,相比于基线模型 (YOLO v8n - seg),单独引入 FPIoU - v2 损失函数后,模型 P、R 及 mAP@0.5 分别提升 4.33、2.63、0.85 个百分点。表

表 6 消融试验结果

Tab.6 Ablation test results

CBAM	FPIoU - v2	P/%	R/%	mAP@0.5/%	参数量
×	×	74.89	67.47	69.89	3.26×10^6
×	✓	79.22	70.10	70.74	3.26×10^6
✓	×	76.76	69.13	72.03	6.52×10^6
✓	✓	81.68	71.09	73.79	6.52×10^6

注:✓表示采用此模块,×表示未采用此模块。

明 FPIoU - v2 通过优化边界框回归策略,有效提升了模型对目标边缘及形状的拟合能力,缓解了复杂果园背景下的定位偏差问题,验证了该损失函数在增强模型定位精度方面的有效性。单独引入 CBAM 注意力机制模块后,模型 P、R 及 mAP@0.5 分别提升 1.87、1.66、2.14 个百分点。表明 CBAM 模块通过通道与空间双重注意力的权重再分配,强化了网络在特征提取阶段对目标关键信息的聚焦能力,能够有效抑制果园中枝叶、光照等复杂背景噪声干扰,提升了模型对重叠及遮挡目标的特征识别与建模水平。当同时引入上述 2 种改进策略后,模型性能达到最优,其 P、R 及 mAP@0.5 较基线模型分别提升 6.79、3.62、3.90 个百分点。试验结果表明,本文改进算法参数量由 3.26×10^6 增至 6.52×10^6 ,但在检测精度与召回率上取得了增益效果。验证了 CBAM 的特征增强策略与 FPIoU - v2 的定位优化策略之间存在良好的协同效应,共同提升了模型在复杂丘陵果园场景下的实例分割效能。

为进一步评估改进算法在同类任务中的综合竞争力,引入 Mask R - CNN^[29] 与 YOLACT^[30] 算法作

为对比基准,结果如表 7 所示。

表 7 不同分割算法在测试集上试验结果
Tab. 7 Different segmentation algorithms on test set

算法	<i>P</i> /%	<i>R</i> /%	mAP@0.5/ %	参数量	帧率/ (f·s ⁻¹)
YOLO v8	74.89	67.47	69.89	3.26×10^6	51.02
改进 YOLO v8-seg	81.68	71.09	73.79	6.52×10^6	39.09
Mask R-CNN	71.61	51.90	71.61	4.40×10^7	4.06
YOLACT	67.62	48.84	67.62	3.65×10^7	10.10

在自建果园数据集下,改进 YOLO v8-seg 算法在检测精度、模型规模及运行效率之间取得最佳平衡效果。在精度指标上,改进算法 mAP@0.5 达到 73.79%,相比于基准模型 YOLO v8n 提升 3.90 个百分点,且优于经典的二阶段算法 Mask R-CNN (71.61%) 及一阶段算法 YOLACT (67.62%)。同时,改进算法精确率 *P* 和召回率 *R* 分别达 81.68%

和 71.09%,表明通过引入 CBAM 注意力机制与 FPIoU-v2 损失函数,模型在复杂背景干扰及目标重叠的丘陵果园场景中展现出更强的特征捕捉与边界拟合能力。在模型效能方面,虽然改进算法因结构优化导致参数量增至 6.52×10^6 ,但体量仍远小于 Mask R-CNN (4.40×10^7) 与 YOLACT (3.65×10^7),仅为后者的 14.8% 左右,表现出明显的轻量化优势。在推理速度上,改进算法帧率为 39.09 f/s,较基准模型有所下降,主要是由于主干网络集成了 CBAM 注意力机制,在通过通道与空间双重注意力的权重再分配来增强特征提取能力的同时,也增加了网络计算深度与计算开销,从而在一定程度上限制了推理速度。但该速度仍远高于 Mask R-CNN 和 YOLACT,满足实时检测要求。

为直观体现算法分割效果,对不同算法分割结果进行可视化,如图 10 所示。

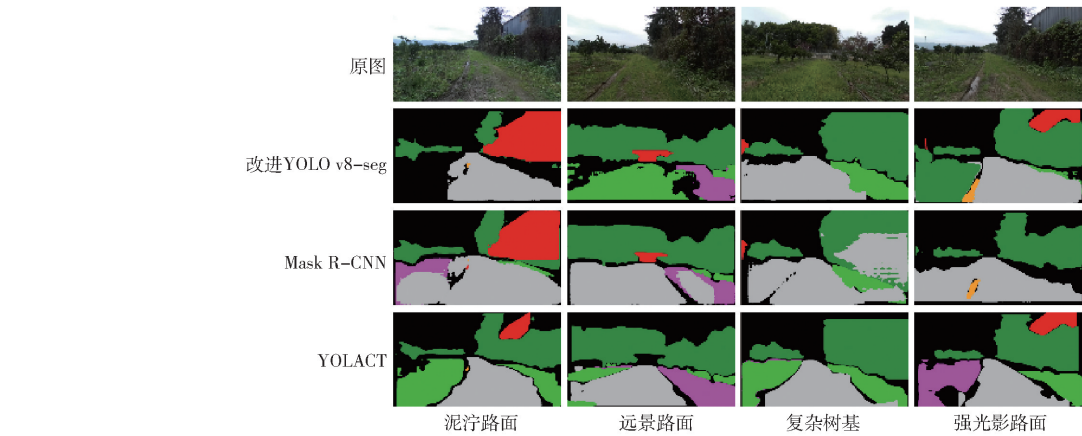


图 10 不同分割算法分割效果

Fig. 10 Segmentation effect of different segmentation algorithms

由图 10 可知,泥泞路面场景中,Mask R-CNN 对地面细微特征的感知能力受限,未能准确识别路面中的泥泞地形,出现了明显的漏检现象;远景路面场景中,对比算法在处理远端非结构化边界时鲁棒性不足,YOLACT 在画面边缘存在误检(粉色斑块),而本文改进算法提取的路面轮廓类别一致性更高;在复杂树基场景中,Mask R-CNN 对近景树干底部的分割精度较低,产生了明显的误检;而强光影路面场景中,光影变化剧烈且伴有积水,YOLACT 在路中间细小的泥泞区域出现漏检,相比之下,本文改进算法提取的泥泞区域最为完整。

由试验结果可知,分割性能受环境因素的制约。丘陵果园剧烈的光影波动易引发过曝或欠曝,掩盖地形关键特征,导致误分类或漏检。消融试验结果表明,引入 CBAM 模块后精确率 *P* 提升 1.87 个百分点,证明注意力机制能有效强化模型对关键信息的聚焦能力,但精确率与召回率落差为 10.59 个百

分点,以及图 10 中光影场景下漏检及误检的结果表明环境光照变化仍是主要的误差来源。其次,地势起伏引起的高频振动,导致成像瞬间产生运动模糊。尽管采用 FPIoU-v2 损失函数将目标定位精度提升约 4.17 个百分点,有效缓解了边界框偏移,但分割结果显示,在地势起伏时,掩码边缘仍存在锯齿化现象。此外,深度相机在面对积水区域时的固有探测噪声,以及远端目标有效特征信息的退化,共同限制了模型对非结构化边界的精确拟合能力。综上,改进算法在处理复杂地形边缘及光影变化方面鲁棒性更强,分割结果最接近真实场景,为语义地图构建提供了参考。

4.3 高程地图构建与语义地图融合

采用履带车轨迹线的高程误差值作为高程地图量化评估指标。在地形起伏剧烈且伴有路边坡地 A、B 场景中,高程重建 RMSE 为 0.28 m,平均误差为 0.23 m;而在地势相对平缓常规果园 C、D 场景中 RMSE 降至 0.15 m,平均误差为 0.11 m。图 11 为典

型作业路段下轨迹高程与真实值曲线,可以看出重建轨迹在时域上能够紧密贴合地面的真实起伏走势,在地形切换点表现出较好的动态跟踪能力。试验结果表明,尽管果园地面存在杂草覆盖及细微颠簸,但本文算法通过概率递推更新机制有效抑制了深度图噪声,能够准确复现地形的高度变化特征。定量指标验证了所构建的 2.5D 高程地图在地形还原上具有较高的可信度,能够为果园履带车通过性评估提供可靠的数据支撑。

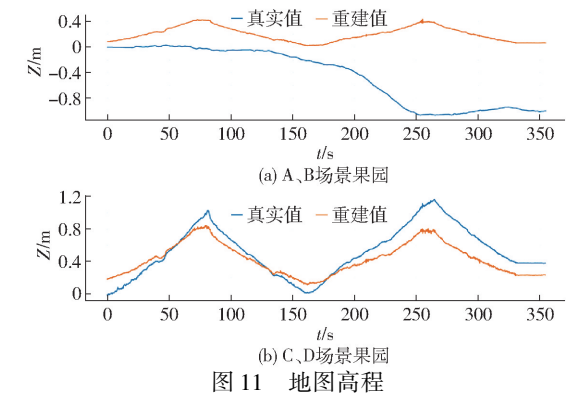


Fig. 11 Map elevation values

为验证算法对丘陵山地果园复杂地形建模能力,在丘陵果园环境下开展了高程地图构建试验,如图 12 所示。图 12a 为试验现场采集的原始 RGB 图像,图 12b 为算法生成的对应场景深度地图。通过将深度信息转换至空间坐标系并进行网格化处理,得到高程地图如图 12c 所示。算法能够准确复现果园地表起伏特征,为障碍物评价与路径规划提供了基础几何拓扑信息。

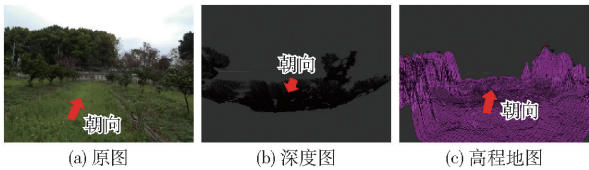


图 12 高程地图构建效果

Fig. 12 Results of elevation mapping

将语义信息与 2.5D 高程地图深度融合,构建了具备地形感知能力的语义高程地图。鉴于逐帧真实值标注成本极高且难以实现,故采用语义栅格一致率作为评估语义投影质量的替代指标。试验结果表明,在地形起伏剧烈且伴有剧烈运动扰动的 A、B 场景中,语义栅格一致率为 56.8%,反映了在非结构化坡地环境下,履带平台频繁颠簸对空间重投影一致性带来的严峻挑战,亦证明了原始语义信息在复杂动态条件下的不稳定性。相比之下,在地势相对平缓的常规果园 C、D 场景中,语义栅格一致率提升至 73.8%,显示出系统在稳定作业状态下具有良好的时空鲁棒性。

由图 13 可知,红框内地形特征被准确辨识,泥泞区域、草地及普通道路分别经投影模块映射为蓝色、浅绿色与黄褐色语义层,初步验证了 2.5D 语义高程地图对复杂地形的表征能力。如图 13a 所示,深度相机在面对泥泞区域时,由于物体表面反射特性的剧烈差异产生非线性深度噪声,致使局部高程地图出现畸变,因此场景 C、D 果园下语义栅格一致率达到 73.8%,而在场景 A、B 果园降至 56.8%。此外,果园中果树枝叶等遮挡地形特征,导致部分栅格出现识别误判。

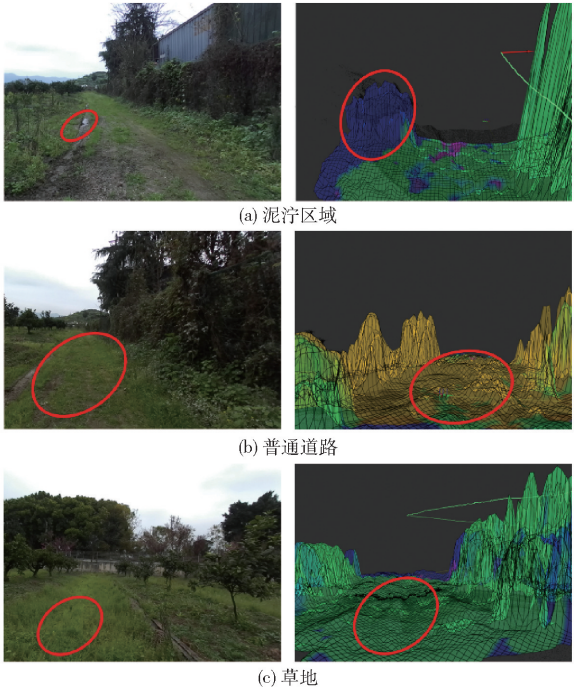


图 13 语义高程地图构建效果

Fig. 13 Results of semantic elevation mapping

综上所述,本文算法能够实现果园复杂环境下高程信息与语义信息的深度融合。所构建的语义高程地图不仅为丘陵山地果园自动化作业提供高度信息,更通过对泥泞、草地等特殊地形的精准标注,为果园履带车在非结构化环境下的路径规划、决策控制提供了语义信息。

5 结论

(1) 提出了一种融合实例分割的定位建图方法。采用 IMU 动态预积分策略,设计了一种适用于丘陵山地果园非结构化环境的前端里程计。试验结果表明,相比于 VINS 与 ORB-SLAM3 视觉里程计,所提算法精度在丘陵山地果园环境下有所提升。

(2) 提出了一种面向丘陵果园复杂环境的改进 YOLO v8-seg 实例分割算法。引入 CBAM 注意力机制,采用 FPIoU-v2 损失函数,并在自建数据集上进行训练优化得到最优模型。结果表明,改进算法

在检测精度与实时性之间取得了最佳平衡,优于 Mask R – CNN 与 YOLACT 等典型算法。

(3)设计并实现了融合语义感知与几何特征的语义高程地图构建方案。该方法解决了传统 SLAM 在丘陵山地果园下环境表征匮乏的问题,为农机通过性评估与路径规划提供直接的地形特征依据,对保障农机在丘陵山地果园下的避障安全性具有重要的支撑作用。

参 考 文 献

[1] Torres-Torriti M. SLAM in agriculture [M] // Nazate-Burgos P. Encyclopedia of smart agriculture technologies. Berlin: Springer, 2023: 1 – 22.

[2] 汪小昆, 谢慎亮, 黄薛凯, 等. 基于 3D 激光雷达 SLAM 的温室芦笋采收机器人自主导航方法[J]. 农业机械学报, 2026, 57(4): 138 – 150.

Wang Xiaochan, Xie Shenliang, Huang Xuekai, et al. Autonomous navigation methods of greenhouse asparagus harvesting robot based on 3D LiDAR SLAM[J]. Transactions of the Chinese Society for Agricultural Machinery, 2026, 57(4): 138 – 150. (in Chinese)

[3] Ding H, Zhang B, Zhou J, et al. Recent developments and applications of simultaneous localization and mapping in agriculture [J]. Journal of Field Robotics, 2022, 39: 956 – 983.

[4] Peng K, Ma W, Lu J, et al. Application of machine vision technology in citrus production[J]. Applied Sciences, 2023, 13: 9334.

[5] Xiong J, Liang J, Zhuang Y, et al. Real-time localization and 3D semantic map reconstruction for unstructured citrus orchards [J]. Computers and Electronics in Agriculture, 2023, 213: 108217.

[6] Jiang S, Qi P, Han L, et al. Navigation system for orchard spraying robot based on 3D LiDAR SLAM with NDT_ICP point cloud registration[J]. Computers and Electronics in Agriculture, 2024, 220: 108870.

[7] Bai Y, Zhang B, Xu N, et al. Vision-based navigation and guidance for agricultural autonomous vehicles and robots: a review [J]. Computers and Electronics in Agriculture, 2023, 205: 107584.

[8] Campos C, Elvira R, Gómez-rodríguez J J, et al. ORB – SLAM3: an accurate open-source library for visual, visual-inertial and multi-map SLAM[J]. IEEE Transactions on Robotics, 2021, 37(6): 1874 – 1890.

[9] 陈久朋, 陈治帆, 伞红军, 等. 基于卷积神经网络的移动机器人自适应光照增强单目视觉 SLAM 算法[J]. 农业机械学报, 2024, 55(12): 383 – 391, 403.

Chen Jiupeng, Chen Zhifan, San Hongjun, et al. Adaptive illumination enhanced monocular vision SLAM algorithm for mobile robots based on convolutional neural networks[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(12): 383 – 391, 403. (in Chinese)

[10] Xu X, Hu J, Zhang L, et al. Detection-first tightly-coupled LiDAR-visual-inertial SLAM in dynamic environments[J]. Measurement, 2025, 239: 115506.

[11] 董蕊芳, 王宇鹏, 阚江明. 基于改进 ORB_SLAM2 的机器人视觉导航方法[J]. 农业机械学报, 2022, 53(10): 306 – 317.

Dong Ruifang, Wang Yupeng, Kan Jiangming. Visual navigation method for robot based on improved ORB_SLAM2[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(10): 306 – 317. (in Chinese)

[12] Xiao Y, Xu W, Li B, et al. Research on visual-inertial measurement unit fusion simultaneous localization and mapping algorithm for complex terrain in open-pit mines[J]. Sensors, 2024, 24(22): 7360.

[13] 苗中华, 祝如江, 张伟, 等. 基于深度强化学习和动态融合感知的农业机器人 SLAM 算法[J]. 农业工程学报, 2026, 42(6): 1 – 11.

Miao Zhonghua, Zhu Rujiang, Zhang Wei, et al. SLAM algorithm for agricultural robots based on deep reinforcement learning and dynamic fusion perception[J]. Transactions of the CSAE, 2026, 42(6): 1 – 11. (in Chinese)

[14] 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.

[15] Lu J, Yao B, Guo H, et al. MOLO – SLAM: a semantic SLAM for accurate removal of dynamic objects in agricultural environments[J]. Agriculture, 2024, 14(6): 7 – 10.

[16] Wang P, Luo W, Liu J, et al. Real-time semantic SLAM-based 3D reconstruction robot for greenhouse vegetables [J]. Computers and Electronics in Agriculture, 2025, 237(7): 110582.

[17] Li Y, Song G, Hao S, et al. Semantic stereo visual SLAM toward outdoor dynamic environments based on ORB – SLAM2[J]. Industrial Robot, 2023, 50(3): 542 – 554.

[18] Yuan Q, Wang P, Luo W, et al. Simultaneous localization and mapping system for agricultural yield estimation based on improved VINS – RGBD: a case study of a strawberry field[J]. Agriculture, 2024, 14(5): 34 – 38.

[19] 冯桑, 张禧龙, 杨润彬, 等. 融合语义分割的葡萄果园机器人稠密地图构建方法[J]. 农业机械学报, 2026, 57(6): 36 – 44.

Feng Sang, Zhang Xilong, Yang Runbin, et al. Dense mapping method for grape orchard robots integrating semantic segmentation[J]. Transactions of the Chinese Society for Agricultural Machinery, 2026, 57(6): 36 – 44. (in Chinese)

[J]. Forestry Machinery & Woodworking Equipment, 2023, 51(4):24–32. (in Chinese)

[31] 曾小松, 宦克为, 曹献文, 等. 基于竞争性自适应重加权采样和二维卷积神经网络的小麦水分含量的定量测定[J]. 长春理工大学学报(自然科学版), 2024, 47(6):36–44.
Zeng Xiaosong, Huan Kewei, Cao Xianwen, et al. Determination of moisture content in wheat using competitive adaptive reweighted sampling and two-dimensional convolutional neural networks[J]. Journal of Changchun University of Science and Technology (Natural Science Edition), 2024, 47(6):36–44. (in Chinese)

[32] 陈立旦, 赵艳茹. 可见-近红外光谱联合随机蛙跳算法检测生物柴油含水量[J]. 农业工程学报, 2014, 30(8):168–173.
Chen Lidan, Zhao Yanru. Measurement of water content in biodiesel using visible and near infrared spectroscopy combined with random-frog algorithm[J]. Transactions of the CSAE, 2014, 30(8):168–173. (in Chinese)

[33] 龙燕, 连雅茹, 马敏娟, 等. 基于高光谱技术和改进型区间随机蛙跳算法的番茄硬度检测[J]. 农业工程学报, 2019, 35(13):270–276.
Long Yan, Lian Yaru, Ma Minjuan, et al. Detection of tomato hardness based on hyperspectral technology and modified interval random frog algorithm[J]. Transactions of the CSAE, 2019, 35(13):270–276. (in Chinese)

[34] 焦李成, 杨淑媛, 刘芳, 等. 神经网络七十年: 回顾与展望[J]. 计算机学报, 2016, 39(8):1697–1716.
Jiao Licheng, Yang Shuyuan, Liu Fang, et al. Seventy years of neural networks: review and prospect[J]. Chinese Journal of Computers, 2016, 39(8):1697–1716. (in Chinese)

[35] 高凯, 杨海波, 尹航, 等. 基于波段优化光谱指数的马铃薯叶水势估测[J]. 灌溉排水学报, 2025, 44(3):1–11.
Gao Kai, Yang Haibo, Yin Hang, et al. Estimating leaf water potential of potato using band-optimized spectral indices[J]. Journal of Irrigation and Drainage, 2025, 44(3):1–11. (in Chinese)

[36] 蒋鹏艳, 高晓东, 何娜娜, 等. 基于多光谱遥感的刺槐正午叶水势反演[J]. 西南林业大学学报(自然科学), 2025, 45(4):142–149.
Jiang Pengyan, Gao Xiaodong, He Nana, et al. Inversion of *Robinia pseudoacacia* midday leaf water potential based on UAV multispectral images[J]. Journal of Southwest Forestry University (Natural Sciences), 2025, 45(4):142–149. (in Chinese)

[37] 李虎, 钟韵, 冯雅婷, 等. 无人机遥感的多植被指数土壤水分反演模型[J]. 光谱学与光谱分析, 2024, 44(1):207–214.
Li Hu, Zhong Yun, Feng Yating, et al. Multi-vegetation index soil moisture inversion model based on UAV remote sensing [J]. Spectroscopy and Spectral Analysis, 2024, 44(1):207–214. (in Chinese)

(上接第 307 页)

[20] Huo P, Ma S, Su C, et al. Emergency obstacle avoidance system of sugarcane basecutter based on improved YOLO v5s[J]. Computers and Electronics in Agriculture, 2024, 216: 108468.

[21] Xu W, Liu L. PCNN orchard heterologous image fusion with semantic segmentation of significance regions[J]. Computers and Electronics in Agriculture, 2024, 216: 108454.

[22] Cui J S, Zhang F R, Feng D Z, et al. An improved SLAM based on RK-VIF: vision and inertial information fusion via Runge-Kutta method[J]. Defence Technology, 2023, 21: 133–146.

[23] Rochefort-Beaudoin T, Vadean A, Achiche S, et al. From density to geometry: YOLO v8 instance segmentation for reverse engineering of optimized structures[PP/OL]. V1. arXiv (2024-04-29) [2026-03-15]. <https://doi.org/10.48550/arXiv.2404.18763>.

[24] Miki T, Wellhausen L, Grandia R, et al. Elevation mapping for locomotion and navigation using GPU[J]. IEEE Robotics and Automation Letters, 2022, 7(3): 7412–7419.

[25] GB/T 37164—2018 自走式农业机械导航系统作业性能要求及评价方法[S].

[26] Zheng C, Zhang P, Li Y. Semantic SLAM system for mobile robots based on large visual model in complex environments[J]. Scientific Reports, 2025, 15: 8450.

[27] Wang H, Jiang B, Xu H. Semantic SLAM based on ORB-SLAM3 combined with YOLO v5 in dynamic environment[C]//Proceedings of the IEEE 7th International Conference on Communications, Information System and Computer Engineering. IEEE, 2025: 535–539.

[28] Qin T, Li P, Shen S. VINS-Mono: a robust and versatile monocular visual-inertial state estimator[J]. IEEE Transactions on Robotics, 2018, 34(4): 1004–1020.

[29] He K. Mask R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision. IEEE, 2017: 2961–2969.

[30] Bolya D. YOLACT: real-time instance segmentation[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. IEEE, 2019: 9157–9166.