

基于 YOLO11n 的芦笋采收机器人多目标分割与识别定位

梁明旭^{1,2,4}, 夏先飞^{1,4*}, 袁昌荣², 吕俊潼^{1,4}, 官庆硕^{1,4}, 贾磊^{1,3,4}

(1. 农业农村部南京农业机械化研究所, 南京 210014; 2. 南京工程学院机械工程学院, 南京 211167; 3. 西北农林科技大学机械与电子工程学院, 杨凌 712100; 4. 苏州中农院华东农业科技中心, 苏州 215300)

摘要: 针对芦笋采收机器人作业过程中的芦笋嫩茎细长、遮挡重叠和母茎干扰导致的多目标分割与识别难题, 该研究以 YOLO11n-seg 为基线模型, 提出改进模型 YOLO11n-SAL。引入多尺度边缘增强模块 (multi-scale edge enhancement module, MEEM), 通过对图像边缘特征的分解和增强提升模型边界分割的精度, 增强对细长目标的检测能力; 同时, 结合分离与增强注意力机制 (separated and enhancement attention module, SEAM), 通过感知并融合遮挡芦笋的多尺度特征增强遮挡环境下对芦笋的检测识别效果。试验结果表明, 改进后的 YOLO11n-SAL 模型在各项评估指标上均取得显著提升, 边框下的检测精度、召回率、mAP_{0.5} 和 mAP_{0.5-0.95} 分别为 94.2%、83.1%、91.2% 和 76.2%; 掩码下的分割精度、召回率、mAP_{0.5} 和 mAP_{0.5-0.95} 分别为 93.4%、77.9%、90.7% 和 62.7%。热力图分析与检测分割效果对比表明, YOLO11n-SAL 模型在不同场景下对芦笋边缘特征的感知能力以及遮挡条件下的多目标分割识别效果均有明显改善, 与基线模型相比, 检测分割效果更好, 且能够有效应对复杂情况的干扰, 显著提升了不同场景下芦笋的分割识别精度。为验证模型实际部署的识别及定位效果, 基于深度相机和机械臂进行芦笋识别、定位及采收抓取试验, 定位成功率达 90% 以上, 采收抓取效果良好, 为农业机器人精准采收作业提供了可靠的技术支撑。

关键词: 多目标检测; 分割; 绿芦笋; YOLO11n 改进模型; 精准采收; 视觉定位

doi: 10.11975/j.issn.1002-6819.202509299

中图分类号: S126

文献标志码: A

文章编号: 1002-6819(2026)-12-0106-10

梁明旭, 夏先飞, 袁昌荣, 等. 基于 YOLO11n 的芦笋采收机器人多目标分割与识别定位[J]. 农业工程学报, 2026, 42(12): 106-115. doi: 10.11975/j.issn.1002-6819.202509299 <http://www.tcsae.org>
LIANG Mingxu, XIA Xianfei, YUAN Changrong, et al. Multi-target segmentation, recognition, and localization for an asparagus-harvesting robot based on YOLO11n[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2026, 42(12): 106-115. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.202509299 <http://www.tcsae.org>

0 引言

芦笋是一种高附加值蔬菜, 富含多种维生素和抗氧化物质, 能够有效增强人体免疫力。近年来, 随着全球健康饮食风潮的兴起以及消费者对绿色食品需求的不断增加, 芦笋的市场需求迅速增长。作为全球最大的芦笋生产国之一, 中国在芦笋种植、加工和出口方面均占据重要地位。截至 2025 年, 全国芦笋种植面积已超过 80 000 hm², 出口额高达 8 亿美元。芦笋主要分为白芦笋和绿芦笋两类。白芦笋为覆膜土下生长, 产量低且种植规模相对较小; 而绿芦笋土上生长, 营养价值高, 种植规模大。目前, 绿芦笋的采收仍以人工为主, 主要有“留母茎采”和“光头采”两种采收模式。“光头采”又称“剃光头”采收, 是指在采收期内将所有破土而出的嫩茎全部采尽, 不留任何地上植株。而“留母茎采”则是在采收过程中, 在每株上保留 2~4 支健壮嫩茎, 让其自然生

长成为进行光合作用的母茎, 同时采收后续萌发的其他嫩茎^[1-3]。由于芦笋生长速度快, 收获周期短, 需要每日巡查和采收, 占用大量劳动力资源, 实现绿芦笋的机械化采收已成为当务之急。然而, 因其复杂的生长环境、枝叶遮挡以及嫩茎细长且颜色与母茎相近等因素, 给采收机器人的目标检测带来了严峻挑战。

近年来, 随着深度学习的发展, 基于深度学习的 YOLO 算法在农业领域得到了广泛应用^[4], 在农作物识别与定位^[5-6]、病虫害监测^[7-9]和成熟度判断^[10-11]等环节助力智能农业发展。在农作物的目标检测研究方面, 任锐等^[12]针对不同栽培模式下玉露香梨果实检测存在精度差、漏检和误检率高等问题, 提出了一种基于改进 YOLOv8n 的轻量化检测模型 YOLO-iBPD, 该模型通过引入 C2fi 模块增强模型的特征提取和表达能力, 采用优化双向特征金字塔网络融合多尺度特征以及使用损失函数提升回归精度, 并结合知识蒸馏方法压缩模型, 最终模型大小仅 3.34 MB, 平均精度达 93.32%。针对荔枝品种间纹理特征不同导致的识别难题, 黄铭杰等^[13]提出一种 SCL-YOLO11 模型, 改进后的模型准确率达到 99.61%, 相较于 YOLO11, 提高了 2.89 个百分点, 召回率提高 2.96 个百分点, 参数量和计算量明显降低。为实现西红柿的快速准确检测, 戈刚等^[14]设计了一款 TOMO-YOLO 模型, 通过引入权重分配的特征融合策略以及自适应检测头提高了复杂场景下的鲁棒性, 试验表明, 识

收稿日期: 2025-09-30 修订日期: 2025-12-02

基金项目: 南京市重大科技专项 (前沿技术) 项目 (202512115); 中国农业科学院科技创新工程基础科学研究中心项目 (CAAS-BRC-SAE-2025-02); 中国农业科学院科技创新工程

作者简介: 梁明旭, 研究方向为机器视觉及农业机器人技术。

Email: 1175063865@qq.com

*通信作者: 夏先飞, 博士, 研究员, 研究方向为收获机械化及农业机器人技术。Email: xiaxianfei@caas.cn

别精度达 88.6%，平均召回率达 85%。针对樱桃果实重叠、遮光等环境因素造成识别不准确的问题，GAI 等^[15]提出了一种改进的 YOLO-V4 深度学习算法，采用 DenseNet 替换主干网络以增强特征复用与传递能力，将 YOLO-V4 模型中的先验框改为了适合樱桃果实形状的圆形标记框，改进后模型平均精度提升 0.15，能有效检测不同成熟度樱桃。针对草莓果实检测定位难和识别效果差等问题，HU 等^[16]设计了一款 YOLOv3 与 Mask R-CNN 相结合的识别定位系统，通过结合具有高分类准确率的双阶段 Mask R-CNN 网络以及具有多尺度融合结构特征的单阶段 YOLOv3 网络使检测精度达 94.5% 以上，可以适应复杂的果园环境。在作物的精细分割和定位方面，针对传统环境下点云缺失难以计算芦笋的实际长度等问题，CHEN 等^[17]设计了一种使用 YOLO-V9 架构结合目标检测和实例分割的模型，并利用深度相机获取的点云数据进行长度计算，使用动态比例系数计算芦笋的实际长度和采摘点，试验结果表明，长度测量的平均绝对误差为 0.9 cm，平均相对误差为 3.5%。为了解决现有轻量级模型在复杂背景、光照条件变化以及葡萄卷须尺度不同的葡萄园中进行分割时容易出现的误检、漏检和分割错误问题，SHEN 等^[18]提出了一种基于改进 YOLOv8 的轻量级葡萄卷须实例分割模型，通过结合多尺度注意力混合头、多通道增强颈部和星形融合模块，使检测与分割在交并比为 0.5 时平均精度分别达 99.3% 与 99.5%，模型参数量仅 2.3M，适用于边缘设备部署。针对番茄采收机器人采摘时，背景复杂、颜色及形状各异、果梗姿态多样、叶子遮挡等问题，张勤等^[19]提出了一种基于 YOLACT 实例分割的视觉定位方法。该方法通过果梗粗分割与精细分割结合深度信息填补，实现采摘点精确定位，并利用果梗几何特征预测采摘姿态。试验结果表明，采摘点识别成功率平均达 98.07%，处理速率 21 帧/s，采摘成功率达 98.15%，显著提升了非结构化环境下的采摘效率与可靠性。针对自然环境下苹果采摘点定位受枝叶遮挡和果实重叠影响的挑战，王文波等^[20]提出一种基于

目标区域语义分割的高效方法，采用 MobileViT-Seg 模型结合轻量级编码器与多尺度池化解码器，在复杂环境下实现精确的苹果轮廓分割，并对掩膜区域进行圆拟合以确定采摘点二维坐标，最终融合 RGB-D 信息获取三维空间位置。试验表明，该模型在轻量化的同时，平均交并比达 89.79%，检测速度 100.06 帧/s，采摘点定位平均准确率为 90.80%，为苹果自动化采摘提供了可靠的技术支持。

针对复杂环境下，机器人采收绿芦笋时，芦笋识别困难，采收精度差等问题，基于现有的研究分析发现，YOLO 系列模型通过其端到端的特征金字塔网络设计，在面对目标重叠、遮挡等复杂场景时，展现出显著的架构优势，而 YOLO11 作为 Ultralytics 团队于 2024 年 9 月发布的全新模型，在特征提取能力与计算效率方面实现了进一步突破。该模型通过引入 C3k2 模块、优化的 SPPF 结构及解耦检测头等创新设计，在保持高推理速度的同时进一步提升了检测精度，且模型采用无锚框机制，支持目标检测、分类、姿态估计及旋转目标检测等多种任务，具备优异的部署适应性，故本研究基于 YOLO11n-seg 分割模型进行改进，使用模型对待采芦笋进行识别与分割并输出边框与分割掩码，提取目标的关键特征与位置信息，为芦笋采收机器人的精准作业提供技术支撑。

1 数据集采集与预处理

1.1 数据采集

数据集由 2025 年 8 月及 9 月于四川省农业科学院简阳芦笋基地采集图像并结合采集的视频进行抽帧并裁剪后作为数据采集来源，分为“光头采”和“留母茎采”两种场景，共 330 张，不同场景下的部分数据集如图 1 所示。芦笋采收环境较为复杂，“光头采”场景下，目标小、多且密集，容易发生重叠；“留母茎”场景下，背景复杂，且芦笋细长，头部特征小，容易被母茎遮挡，且待采芦笋和母茎颜色相同，区分难度大。

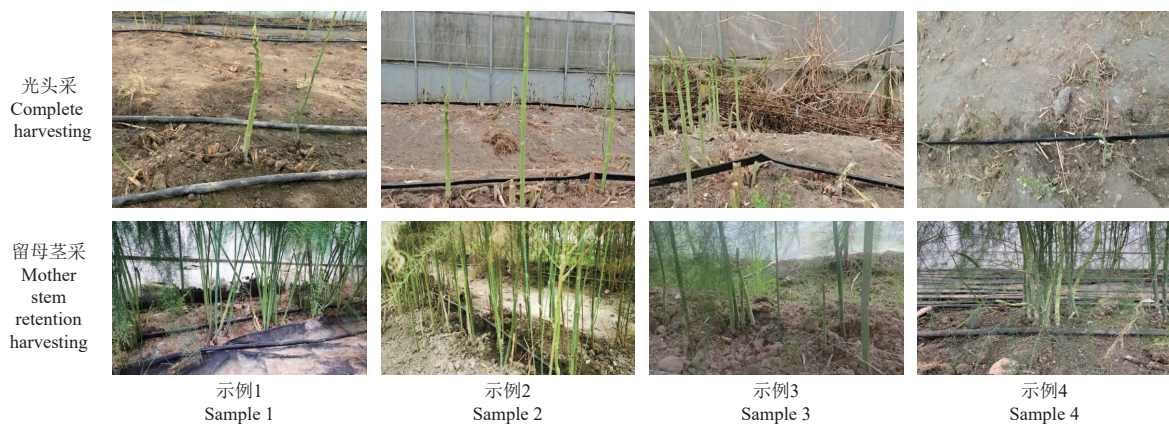


图 1 “光头采”和“留母茎采”芦笋植株图像

Fig.1 Images of asparagus plants under "complete harvesting" and "mother stem retention harvesting"

1.2 数据集的构建

首先使用 Labelimg 软件对图片进行预处理，该工具能够高效地为实例分割任务创建标注数据。标注过程中，对图像中每一棵达到采收标准的芦笋边缘轮廓进行精细

标注，共标注出 915 个待采芦笋，同时为确保标注质量，制定了统一的标注规范。数据集中仅包含一个层级的标签，即“asparagus”（芦笋），标注示意如图 2 所示。标注完成后进行可视化复查，将标注轮廓与原始图像叠

加对比, 确保轮廓精确覆盖目标芦笋且无背景像素混杂。



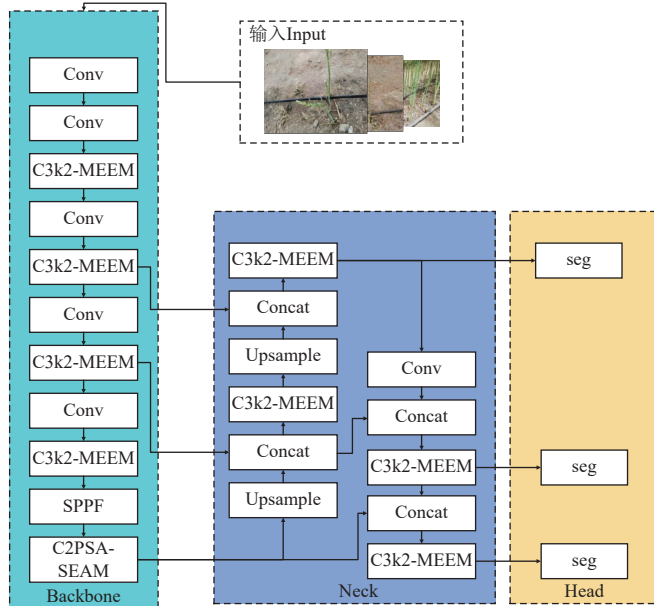
图2 标注示意

Fig.2 Labeling example

为提高模型的鲁棒性与泛化能力, 并考虑采收机器人作业过程中相机视角变化、田间光照波动及目标姿态差异等因素, 完成图像与标注后, 先将原始数据集按 7:2:1 划分为训练集、验证集和测试集, 分别包含 231、66 和 33 张图像。随后, 仅在训练阶段对训练集启用在线数据增强, 分别用于模型验证和泛化性能评价。数据增强基于 Ultralytics YOLO 框架实现, 主要包括色调调整、饱和度调整、平移、缩放和 Copy-Paste 等方法; 同时启用 RandAugment^[21] 自动增强策略, 通过随机组合增强算子并控制增强幅度, 增加训练样本的表现差异, 提高模型在复杂农业场景下的适应能力。由于数据增强采用在线动态方式, 仅在训练阶段对训练集随机启用, 且不生成额外固定图像文件, 因此增强处理后训练集、验证集和测试集的固定样本量不变。

2 YOLO11n 分割模型的改进策略

本研究以 YOLO11n-seg 为基线模型, 提出改进的 YOLO11n-SAL 模型, 整体框架如图 3 所示。



注: C3k2-MEEM 为改进后的重要特征提取模块; C2PSA-SEAM 是改进后的具有注意力机制的特征提取模块; Upsample 为上采样操作; seg 为分割头。Note: C3k2-MEEM is improved significant feature extraction module; C2PSA-SEAM is improved feature extraction module with attention mechanism; Upsample denotes the upsampling operation; seg refers to the segmentation head.

图3 YOLO11n-SAL 的网络结构

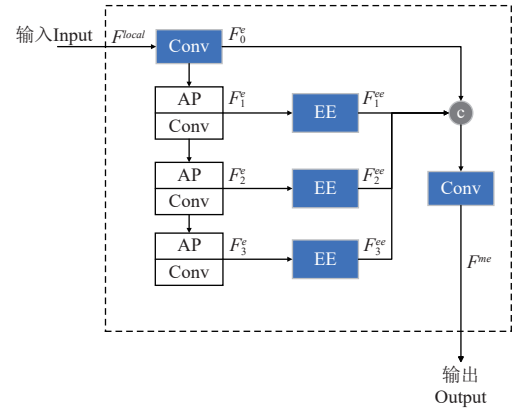
Fig.3 Architecture of the YOLO11n-SAL network

主要改进内容如下: 1) 针对绿芦笋茎干细长、宽高比极端的特性, 传统边界框和分割方法易导致漏检或定

位偏差, 为强化芦笋轮廓和边缘信息, 提升分割掩码的精细度, 在 C3k2 的基础上融入多尺度边缘增强模块 (MEEM)^[22], 并重新设计瓶颈 (Bottleneck) 结构, 提出 C3k2-MEEM 模块, 替换颈部 (Neck) 和主干网络 (Backbone) 的 C3k2 模块, 以改善边界分割质量, 提升对细长目标的检测精度。2) 鉴于芦笋实际采摘过程中头部特征不明显且易遮挡, 目标尺寸较小导致识别困难, 引入分离与增强注意力机制 (SEAM)^[23], 并替换 C2PSA 模块中的注意力层, 提高对多尺度特征的感知, 解决环境遮挡问题。

2.1 融合多尺度边缘增强模块

基于机器学习的目标检测模型, 一般采用固定的卷积头, 限制了对目标边缘细节特征的信息获取, 然而在对芦笋图像分割时, 常存在轮廓模糊、边界不清晰的问题, 极大影响对关键特征的识别, 本研究将标准的 Bottleneck 中的第二个卷积替换为多尺度边缘增强模块 (MEEM), 替换 C3k 中的 Bottleneck, 构建 C3k2-MEEM 模块, 目的是为了从输入图像中提取并增强多尺度的边缘信息。MEEM 的框架如图 4 所示。



注: EE 为边缘增强器, F^{local} 为输入特征; $F_t^e (t=0, 1, 2, 3)$ 为 t 尺度下的边缘特征; $F_l^{ee} (l=1, 2, 3)$ 为经过增强器输出的各尺度边缘特征。 F^{me} 为 MEEM 模块输出的边缘特征; c 表示通道维度上的特征拼接操作; AP 表示平均池化。

Note: EE is edge enhancer, F^{local} is the input feature; F_t^e is the edge features at scale $t (t=0, 1, 2, 3)$; $F_l^{ee} (l=1, 2, 3)$ is the enhanced edge features output by the enhancer at each respective scale; F^{me} is the final edge feature output by the MEEM module; c denotes the feature concatenation operation along the channel dimension; AP represents average pooling.

图4 MEEM 结构

Fig.4 Architecture of multi-scale edge enhancement module (MEEM)

提取原特征 F^{local} 使用平均池化和 1×1 卷积在不同尺度下提取边缘信息 F_0^e , F_0^e 如随后通过 3 级平均池化操作逐步下采样, 得到不同尺寸下的边缘特征 F_{t+1}^e 。每一尺度的特征均经过专门的边缘增强器处理, 增强器通过计算特征与其平均池化结果的差值来突出边缘响应, 再通过 1×1 卷积和残差连接强化边缘信息并输出增强后的特征 F_l^{ee} , 计算如式 (1) ~ (4)。

$$F_0^e = \phi_{1 \times 1}(F^{local}) \quad (1)$$

$$F_{t+1}^e = AP(\phi'_{1 \times 1}(F_t^e)), 0 \leq t \leq 2 \quad (2)$$

$$F_l^{edge} = F_l^e - AP(F_l^e) \quad (3)$$

$$F_l^{ee} = \phi_{1 \times 1}'(F_l^{edge}) + F_l^e \quad (4)$$

$$F^{me} = \phi_{1 \times 1}([F_0^e, F_1^{ee}, F_2^{ee}, F_3^{ee}]) \quad (5)$$

式中 $\phi_{1 \times 1}$ 表示使用 1×1 卷积层，AP 表示使用 3×3 核的平均池化。 $\phi_{1 \times 1}'$ 表示一个包含归一化与 Sigmoid 激活函数的 1×1 卷积层。

随后，将原始边缘特征与多尺度增强特征沿通道维度拼接，并通过 1×1 卷积融合，生成最终的多尺度边缘增强特征 F^{me} ，该特征既保留了不同尺度的边缘结构，又通过增强机制强化了显著性区域的细节。

2.2 引入分离与增强注意力机制

在复杂环境中，芦笋常因母茎遮挡干扰，导致特征消失或边界模糊，同时存在多目标之间的重叠，相互遮挡，极大增加了模型检测分割的难度，而传统的注意力模块未能充分考虑遮挡场景下空间与通道特征的关联性，难以有效解决芦笋关键特征遮挡问题。

分离与增强注意力机制 (separated and enhancement attention module, SEAM) 是一种自集成注意力机制，其通过多视角的特征融合以及一致性正则化来增强模型对关键特征的感知，以此来提高模型的鲁棒性和泛化能力，SEAM 框架如图 5 所示。

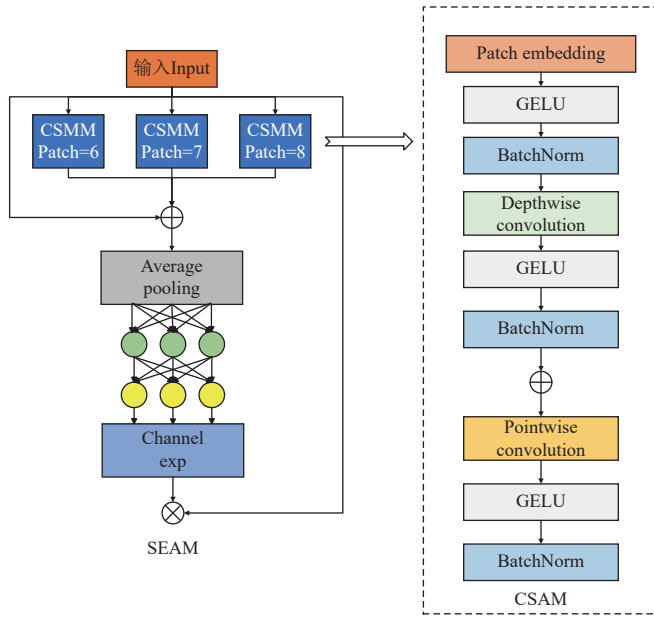


图 5 SEAM 结构

Fig.5 Architecture of separated and enhancement attention module(SEAM)

输入特征图首先通过 CSMM (channel and spatial mixing module) 的特征增强子模块。CSMM 采用深度可分离卷积与残差连接相结合的结构：首先，深度卷积在每个通道内独立提取空间特征，以捕捉局部细节与空间模式；随后，逐点卷积融合跨通道信息，建立通道间的语义关联。

随后对处理后的特征进行全局平均池化以压缩空间信息，并将得到的通道描述符送入一个两层全连接层网络生成初始的通道注意力权重；该权重经由指数函数进行归一化处理，将其值域从 $[0,1]$ 映射至 $[1,e]$ 以增强对定位偏差的鲁棒性。最后，归一化权重与原始输入特征

图进行逐通道相乘，输出经过注意力机制增强的特征，从而有效提升模型在遮挡场景下的多尺度特征响应能力。

3 试验与结果分析

为了进一步探究改进模型性能及其对核心问题的解决效果，设计了相关试验并进行结果分析。

3.1 试验环境

考虑到试验所使用的模型，数据集大小对硬件的要求较高，所以本研究在一台云服务器上进行试验，所使用的服务器的硬件、软件见表 1。

表 1 试验环境配置

硬件 Hardware	软件 Software
CPU:16 核	Ubuntu 22.04
GPU: NVIDIA-GEFORCE-RTX-4090	Pytorch 2.5.1
内存 Memory: 64/GB	Python 3.10.12
显存 Video memory: 24/G	CUDA 12.1.1

此外，试验涉及的图像处理使用 Opencv 库，代码的编写、调试及运行均在 PyCharm2025.2.0.1 上完成。

3.2 评价指标

本研究采用的评价指标为精准率 (precision, P)、召回率 (recall, R)、平均精度均值 (mAP_{0.5} 和 mAP_{0.5-0.95})，由于本研究包含目标检测和实例分割两个任务，所以评价指标分为边框 (Box) 指标和分割掩码 (Mask) 指标。

边框精度和边框召回率分别衡量模型预测的边界框中有多少是正确有效的以及模型能够找出多少真实存在的目标，反映检测的准确性和完整性。边框平均精度均值通过在不同交并比阈值下计算各类别平均精度并取均值，综合评估模型在定位和分类上的整体能力。掩码精度和掩码召回率针对像素级分割结果，衡量预测的分割掩码在像素覆盖上的准确程度和对真实实例的捕捉能力。掩码平均精度均值作为实例分割任务的核心指标，进一步在多个掩码交并比阈值下对各类别的分割性能进行综合评价，全面反映模型在精细轮廓分割上的表现。

3.3 试验方法

为全面评估所提出的 YOLO11n-SAL 模型的性能，选取当前 YOLO 系列中具有代表性的轻量化模型作为对比，包括 YOLOv8n 和 YOLO11n。这些模型均针对低资源环境设计，在计算效率、参数量控制和部署便利性之间取得了较好平衡，同时支持实例分割任务，兼具较高的检测精度与实用价值。

此外，本试验采用迁移学习的方案，为 YOLOv8n、YOLOv11n 加载其官方发布的预训练权重，同时，YOLO11n-SAL 模型也采用 YOLO11n 的预训练权重与 YOLOv8n、YOLO11n 进行训练及验证指标的对比。同时，为探究各模块的作用及其实用性，进行了消融试验，最后，为了验证模型的注意力感知及检测分割效果，进行了相关热力图试验和检测分割测试。

3.3.1 不同分割模型训练与验证结果对比

在相同试验环境下使用自建芦笋数据集进行训练。训练过程中，batch 设为 32，epoch 设为 300，并设置了相关数据增强参数及训练配置，如表 2 所示。

各模型在验证集数据的训练对比结果见图 6，验证对比结果见表 3。

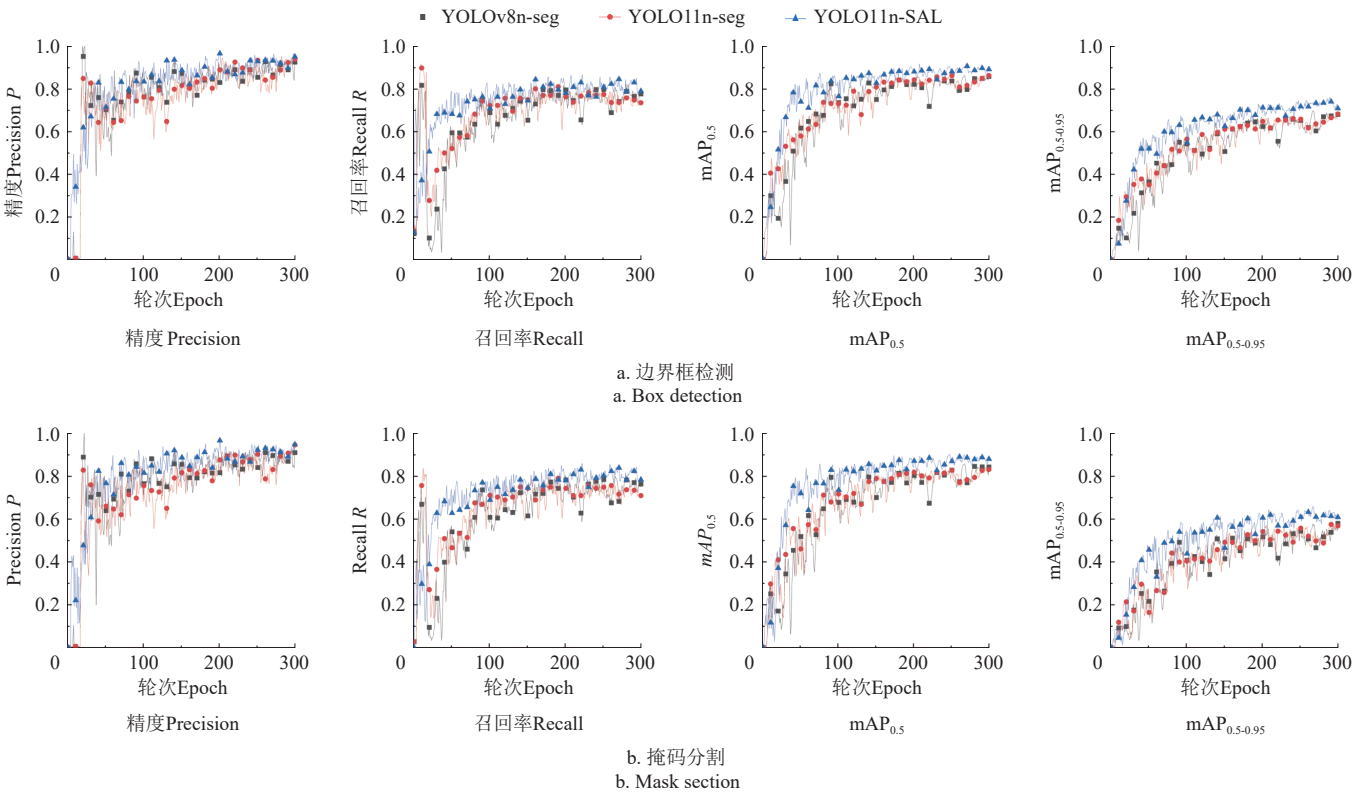
表 2 训练配置

Table 2 Training configuration

参数类型 Hyperparameters types	具体设置 Configuration details
输入图像尺寸 Input image size	640×640
优化器 Optimizer	AdamW
色调调整 Hsv_h	0.5
饱和度调整 Hsv_s	0.5
平移 Translate	0.1
旋转 Degrees	0.1
缩放 Scale	0.1

从模型的各项性能指标对比可以看出，YOLO11n-

SAL 的各项性能都明显优于 YOLOv8n-seg 和 YOLO11n-seg。在相同的试验条件下，YOLO11n-SAL 在边框指标和掩码指标上均有所提升，从边界框检测的各项指标来看，YOLO11n-SAL 的表现最为优异，其边框精度（94.2%）和边框召回率（83.1%）均高于另外两个对比模型，这表明该模型在判断一个区域是否为芦笋（精度）以及找出所有芦笋（召回）的能力上取得了最佳平衡，有效缓解了极端宽高比目标易漏检的问题。更重要的是，其 $mAP_{0.5}$ 达到了 91.2%，远超 YOLOv8n-seg 的 86.5% 和 YOLO11n-seg 的 87.7%；在更严格的 $mAP_{0.5-0.95}$ 指标上也达到了 76.2%。



注： $mAP_{0.5}$ 表示为 IoU 阈值为 0.5 时的平均精度均值， $mAP_{0.5-0.95}$ 表示为 IoU 阈值为 0.5-0.95 时的平均精度均值。下同。
Note: $mAP_{0.5}$ is the mean average precision at IoU threshold 0.5, $mAP_{0.5-0.95}$ is the mean average precision over IoU thresholds from 0.5 to 0.95 (with a step size of 0.05). The same below.

图 6 YOLOv8n-seg、YOLO11n-seg 与 YOLO11n-SAL 的训练结果对比
Fig.6 Comparison of YOLOv8n-seg, YOLO11n-seg and YOLO11n-SAL training result

表 3 模型验证结果对比

Table 3 Comparison of model validation result

模型 Models	边界框检测 Box detection				掩码分割 Mask section				参数量 Parameters/M
	P/%	R/%	$mAP_{0.5}/\%$	$mAP_{0.5-0.95}/\%$	P/%	R/%	$mAP_{0.5}/\%$	$mAP_{0.5-0.95}/\%$	
YOLOv8n-seg	90.8	80.4	86.5	70.7	89.8	80.4	85.9	60.1	3.26
YOLO11n-seg	92.5	79.7	87.7	71.2	89.4	77.0	85.7	58.1	2.80
YOLO11n-SAL	94.2	83.1	91.2	76.2	93.4	77.9	90.7	62.7	3.02

在分割任务的相关指标上，YOLO11n-SAL 的掩码精度高达 93.4%，领先于另外两个模型，说明其生成的像素级分割掩码最为准确。尽管其掩码召回率（77.9%）与 YOLO11n-seg（77.0%）相当，但其分割任务的 $mAP_{0.5}$ 和 $mAP_{0.5-0.95}$ 分别达到了 90.7% 和 62.7%。

3.3.2 消融试验

为了验证各改进模块对性能影响及贡献，设计了 4 组消融试验，结果如表 4 所示。

以 YOLO11n 基线模型为参照，在引入边缘增强模块（MEEM）后，模型边界框召回率从 79.7% 提升至 82.4%，表明该模块有助于增强模型对芦笋目标的感知与召回能力，有效降低了漏检率。边界框精度从 92.5% 下降至 89.0%，同时掩码精度也出现一定下降。值得注意的是，边界框与掩码的 $mAP_{0.5}$ 分别提升至 89.6% 与 89.1%。上述结果表明，边缘增强模块的引入增强了对多尺度目标的感知能力，但同时增加了更多误检目标，导

致边界框精度和掩码精度有所降低。

单独引入分离与增强注意力机制（SEAM）模块后，模型的边界框检测精度和掩码分割精度分别提升至 93.5% 和 90.3%，这表明 SEAM 模块有效增强了模型对芦笋多尺度特征的感知能力。

最终，当共同引入 MEEM 与 SEAM 模块时，模型性能达到了最优。MEEM 模块对芦笋的边缘细节的感知增强，能够检测到更多目标，而 SEAM 模块则感知并细分更多被遮挡的多尺度特征，确保这些特征能被精确检测并分割出来。

表 4 消融试验
Table 4 Ablation test

YOLO11n	MEEM	SEAM	边框精度 $Precision_{Box}/\%$	边框召回率 $Recall_{Box}/\%$	$mAP_{0.5}(Box)/\%$	$mAP_{0.5-0.95}(Box)/\%$	掩码精度 $Precision_{Mask}/\%$	掩码召回率 $Recall_{Mask}/\%$	$mAP_{0.5}(Mask)/\%$	$mAP_{0.5-0.95}(Mask)/\%$
√	×	×	92.5	79.7	87.7	71.2	89.4	77.0	85.7	58.1
√	√	×	89.0	82.4	89.6	74.9	88.3	77.5	89.1	60.1
√	×	√	93.5	82.5	88.7	75.3	90.3	77.1	86.1	60.4
√	√	√	94.2	83.1	91.2	76.2	93.4	77.9	90.7	62.7

注：√ 表示采用该改进方法，× 表示不采用该改进方法。
Note: √ indicates adoption of this improvement method, × indicates non-adoption of this improvement method.

3.3.3 模型特征可视化

热力图通过颜色的变化来直观表示特征分布、及模型对特征的关注意度，其中颜色越明亮，表明模型对该特征的注意力越高。

图 7 为不同视角下的多目标感知结果热力图，由近距离视角下，芦笋重叠程度高，而远距离下芦笋分散，且背景更为复杂。

型因引入了注意力机制及边缘增强，在近景中呈现出更优的特征感知度与细节保留度，生成的分割掩膜更贴合芦笋的实际轮廓，相互紧靠的芦笋也能被清晰地分离和标识出来，显示出其在复杂场景下优异的特征提取与判别能力。而从图 7b 可以看出，在远景多目标场景中，由于目标尺度变小、细节信息减少以及背景环境（如土壤、灌溉管线等）干扰增强，各模型面临的挑战更为严峻。YOLOv8n-seg 模型在此条件下出现部分小目标漏检或轮廓定位模糊的情况，但其热力图响应强度相对较好。YOLO11n-seg 模型通过改进的网络设计，在远景小目标的召回率上有所提升，能够检测到更多远处目标，但其分割边界的精细度因特征抽象而有所损失，热力图响应弱。YOLO11n-SAL 模型在此场景下依旧有较明显的优势，在远景目标上仍能保持较高的激活响应，表明该模型对小尺度芦笋依然具备强大的识别能力，分割轮廓的完整性保持得最好，证明了其在多尺度目标检测方面优秀的鲁棒性。

3.3.4 分割试验对比

试验使用训练后的 YOLOv8n、YOLO11n 的分割模型与 YOLO11n-SAL 进行对比，分别选取远景多目标、母茎茂密、近景多目标、“光头采”多目标及母茎背景较深 5 个场景的图像进行测试，测试效果如图 8 所示。

如图 8 所示，在“光头采”的场景下，背景复杂度低，无明显遮挡时，YOLOv8n-seg、YOLO11n-seg 和 YOLO11n-SAL 三个模型的检测与分割能力均较为优秀，YOLO11n-SAL 检测分割表现最佳，而在更具挑战性的“留母茎采”场景中，各模型性能出现分化：如图 8a 与图 8e 所示，当环境中存在多目标、母茎高大茂密且背景深暗时，YOLOv8n-seg 出现明显漏检，对形态细长的芦笋难以实现有效分割；在母茎茂密、背景干扰强的复杂条件下，YOLO11n-seg 的表现优于 YOLOv8n-seg，未出现明显漏检，但其整体检测置信度与分割质量相较于简单场景仍有明显下降，置信度普遍低于 70%。

相比之下，YOLO11n-SAL 在全部测试场景中均保持较优的综合性能。即便在母茎茂密、遮挡严重的极端条件下，其性能虽有轻微减弱，但仍能维持 70% 以上的高置信度，远优于基线模型。说明 YOLO11n-SAL 所引入的改进机制显著增强了对遮挡、复杂背景及多目标干扰的鲁棒性，在芦笋采摘这一细分任务中展现出更强的实用性与稳定性。

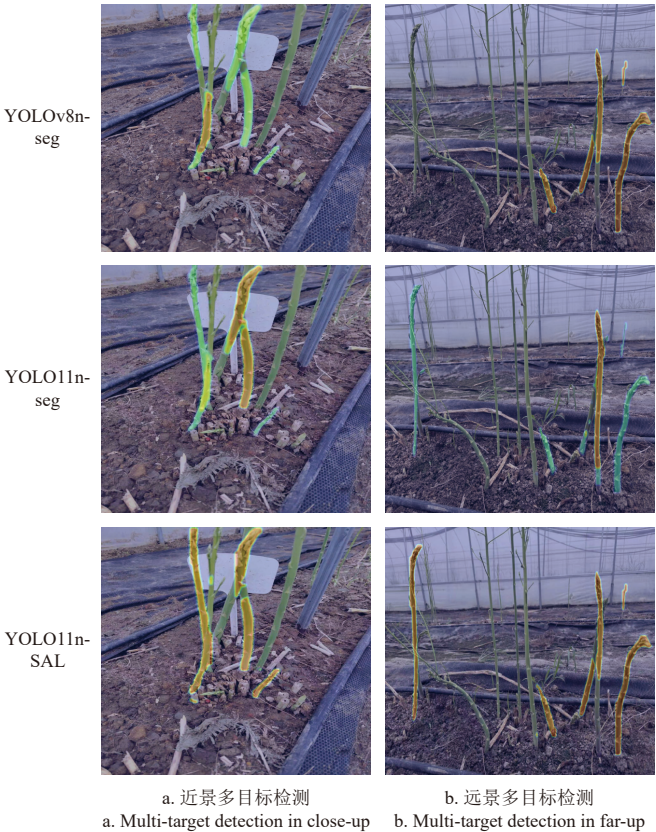


图 7 各模型热力图对比
Fig.7 Comparison of heatmaps for different models

从图 7a 可以看出，在近景多目标场景中，芦笋植株间距相对较小，形态特征较为清晰。YOLOv8n-seg 能够有效识别出多数目标，但其分割边界存在轻微偏差，特别是在相邻植株轻微重叠或遮挡的区域。YOLO11n-seg 在此场景下表现出更优的边界框置信度与轮廓清晰度，对密集目标的区分能力有所增强。而 YOLO11n-SAL 模

3.3.5 芦笋定位及采收抓取试验

为验证模型实际部署的识别、掩码分割及实际定位效果,提出了一种芦笋定位方法,并基于实验室现有的舵机机械臂,设计了一套基于 ROS 集成的眼在手外的机械臂多芦笋定位及抓取系统,如图 9 所示。

首先,基于 YOLO11n-SAL 模型,对芦笋进行实时识别并生成对应的掩码图像,并将芦笋掩码基部的最低

点设定为采收点。通过 Intel RealSense D415 深度相机获取场景的深度信息,并将其转换为三维点云数据。随后,以设定的采收点为中心创建一个圆形分析区域,在该区域内采用 DBSCAN 聚类算法剔除噪声点,提升位置数据的稳定性。进一步利用主成分分析 (principal component analysis, PCA) 方法估计该局部点云所构成平面的法向量,从而实时解算采收点的三维空间坐标。

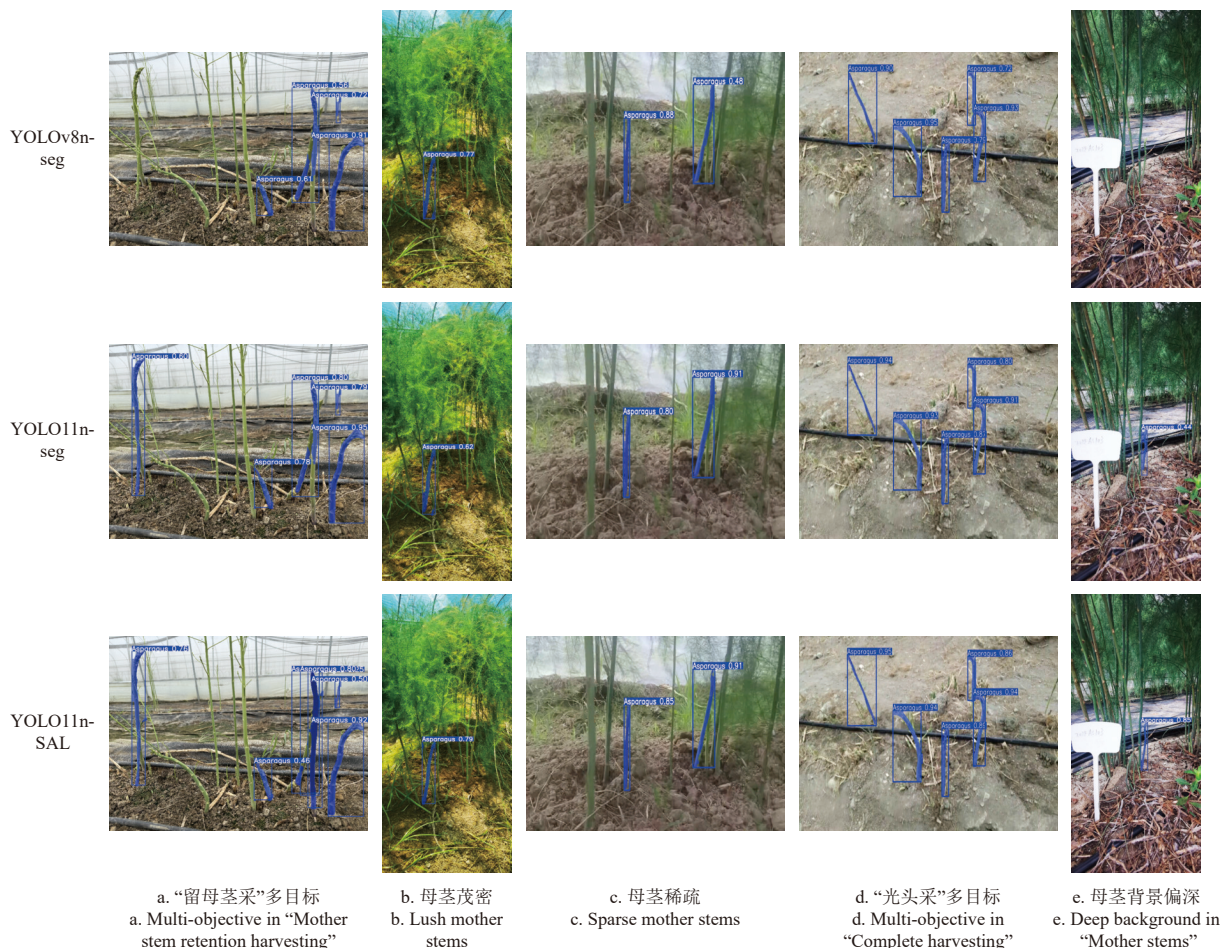


图 8 各模型测试效果对比

Fig.8 Test comparison of different models



图 9 六轴机械臂

Fig.9 Six-axis robotic arm

开发基于 YOLO11n-SAL 的识别与定位节点。该节点实时发布采收点的三维坐标,并通过 RVIZ 可视化工具对识别与定位结果进行直观展示。图 10 所示即为芦笋模型的采收点识别效果图。

其次,采用“眼在手外”(Eye-to-hand)的系统配置方案搭建视觉引导系统。在该配置中,深度相机固定安装于机械臂工作区域外的独立位置,其相机光学坐标系与机械臂基坐标系的相对空间关系保持静止。系统利

用开源工具 EasyHandeye 进行手眼标定:通过在不同位姿下获取 ArUco 码标定板在相机坐标系中的位姿以及通过 MoveIt!读取的机械臂末端执行器在基坐标系下的位姿,采集多组对应数据。随后,基于 $AX=XB$ 的手眼标定模型进行解算,其中 A 和 B 是分别表示相机和机械臂基坐标的齐次变换矩阵。 X 是所求解的手眼矩阵,表示相机与机械臂之间的空间关系。最终得到深度相机光学坐标系到机械臂基坐标系的精确手眼变换矩阵 X 。标定流程如图 11 所示。

最后,利用 MoveIt!框架实现机械臂的实时运动规划与控制。通过订阅上游视觉识别节点发布的采收点三维坐标,并通过所求解的手眼矩阵 X 进行坐标转换,结合预设的 45 度抓取倾斜角,生成包含三维位置与四元数取向的抓取位姿。为提高操作安全性,设置了一个接近位姿,形成完整的采收抓取序列:机械臂首先运动至接近位姿,随后移动至精确抓取位姿并闭合夹爪,接着返回接近位姿,再回到初始位姿,之后运动至放置位姿并释放

目标对象，最后再次返回初始位姿，准备执行下一次任务。

在运动规划层面，系统调用 OMPL (the open motion planning library) 运动规划库，并选用其内置的 RRTConnect 算法进行路径规划。该算法基于随机采样的树扩展策略，适用于复杂空间中的运动规划问题。为提升对多个芦笋目标的采收效率，系统在识别到多个采收点后，会依据视觉模块输出的识别置信度以及机械臂当前位姿到各目标点的运动距离进行综合排序，优先采置信度高、运动路径短的目标，从而优化整体采收序列，减少空程移动和时间消耗。抓取流程如图 12 所示，其中，芦笋的识别与定位工作完全由眼在手外 (eye-to-hand) 的深度相机完成，而眼在手上 (eye-in-hand) 的深度相机并未参与该过程。



图 10 芦笋采收点识别与定位效果

Fig.10 Effects of asparagus harvesting point identification and localization



图 11 手眼标定流程

Fig.11 Hand-Eye calibration process

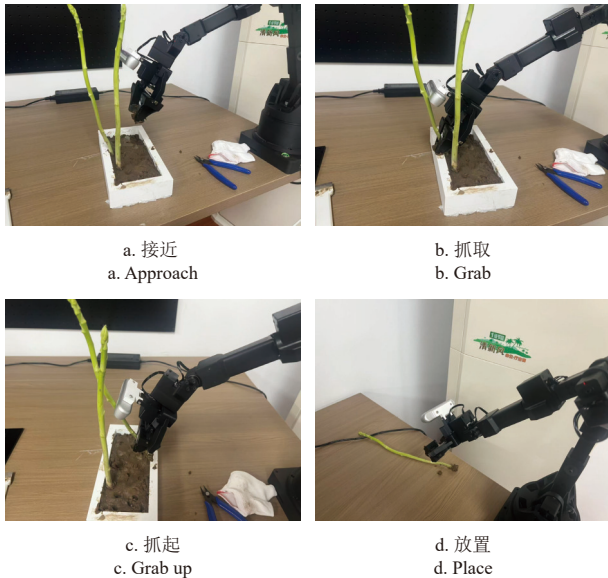


图 12 芦笋抓取流程

Fig.12 Asparagus grasping process

在试验设计上，为模拟真实环境下的芦笋，将芦笋插至 20 cm×9 cm 的土壤中，为测试芦笋定位及采收抓取序列效果，如图 13 所示，分别设置了 20 组不同位姿的单棵芦笋识别采收抓取、20 组 2 棵芦笋识别采收抓取以及 20 组 3 棵芦笋识别采收抓取，试验布置如图 13 所示。

最后，进行采收抓取试验，记录并统计定位及抓取效果，如表 5 所示，其中，定位准确次数是指成功获取所有芦笋采收点的试验次数，抓取成功次数是指成功抓取所有芦笋的试验次数。

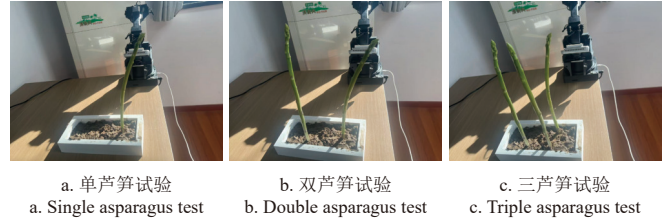


图 13 芦笋试验布置

Fig.13 Asparagus grasping experiment arrangement

表 5 抓取试验
Table 5 Grasping test

试验组 Test group	试验次数 Number of tests	定位准确 次数 Successful localizations	抓取成功 次数 Successful grasps	定位准确率 Localization success rate/%	抓取成 功率 Grasping success rate/%
单芦笋采收抓取	20	19	18	95	90
双芦笋采收抓取	20	18	18	90	90
三芦笋采收抓取	20	18	17	90	85

根据表 5 可知，所部署模型在机械臂上的整体定位效果良好，但随着目标数量的增加略有下降。在单根芦笋抓取任务中，系统表现稳定，定位成功率达到 95%，抓取成功率为 90%，其中出现一次因机械臂奇异点导致的抓取规划失败。当目标增加至 2 棵时，定位成功率降至 90%。在 3 棵芦笋采收抓取任务中，抓取成功率进一步下降至 85%。除奇异点问题外，还存在因关节运动误差与定位累积误差导致机械臂未能准确到达预定位置，从而造成抓取失败的情况。

试验结果表明，随着目标数量增多与环境复杂度上升，机械臂抓取操作受奇异点、关节误差及定位累积误差等因素的影响逐渐显现，系统整体成功率出现一定程度下降。尽管如此，各场景下抓取成功率均保持在 85% 以上，说明该系统已具备基本的实用性与鲁棒性。然而，针对多目标同步定位与抓取任务的精度与可靠性，仍有进一步优化和提升的空间。

4 结 论

本研究针对机器人在芦笋采收过程中存在芦笋细长、特征集中在头部、目标小且易被母茎遮挡导致特征提取、检测分割难度大及复杂环境适应力不足的问题，基于 YOLO11n-seg 改进出 YOLO11n-SAL 模型，通过在模型结构中引入 MEEM、SEAM 模块，研究成果如下：

1) 建立了“光头采”和“留母茎采”任务下多种场景的综合数据集，并通过数据增强技术，提升了模型的泛化能力，为采收机器人对芦笋识别检测及分割提供了良好的数据基础。

2) 引入多尺度边缘增强模块和分离与增强注意力机制，提升了模型对多场景芦笋的检测与分割能力，试验表明，YOLO11n-SAL 在各项检测分割训练、验证及多项测试指标上相较于 YOLOv8n-seg 和 YOLO11n-seg 都

有明显提升, 其平均精度均值 $mAP_{0.5}$ 分别提升至 91.2% 和 90.7%, 有效解决了多复杂场景下, 芦笋检测分割难、识别精度差的问题。

3) 通过机械臂采收抓取芦笋试验, 验证了模型的检测及分割定位性能, 部署到机械臂视觉伺服系统上, 有着较好的定位性能, 定位成功率达 90% 以上, 抓取成功率达 85% 以上, 提供了一种有效的芦笋采收抓取方案。

本研究提出了一种有效的绿芦笋检测与分割模型, 并将其部署到机械臂抓取系统上测试了定位效果, 但仍需进一步提升其适用性。未来工作应采集不同时间、不同光照条件下多种类型芦笋的图片及视频数据来扩充数据集, 从而增强模型的泛化能力与鲁棒性。同时, 将通过简化颈部网络结构、模型剪枝等轻量化策略, 进一步提升模型的可部署性, 并在此基础上, 基于扩展后的芦笋数据集, 设计并优化定位算法。

【参 考 文 献】

- [1] 汪小岳, 李为民, 王琳, 等. 基于改进 YOLACT++ 的成熟芦笋检测-判别-定位方法[J]. *农业机械学报*, 2023, 54(7): 259-271.
WANG Xiaochan, LI Weimin, WANG Lin, et al. Method of detection-discrimination-localization for mature asparagus based on improved YOLACT++ algorithm[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2023, 54(7): 259-271. (in Chinese with English abstract)
- [2] 杨光. 面向选择性采收的绿芦笋智能采摘机械臂设计与试验[D]. 泰安: 山东农业大学, 2025.
YANG Guang. Design and Experimentation of an Intelligent Robotic Arm for Selective Harvesting of Green Asparagus[D]. Tai'an: Shandong Agricultural University, 2025. (in Chinese with English abstract)
- [3] 张萍. 白芦笋选择性采收机器人关键技术研究[D]. 泰安: 山东农业大学, 2024.
ZHANG Ping. Research on Key Technology of Selective Harvesting Robot for White Asparagus[D]. Tai'an: Shandong Agricultural University, 2024. (in Chinese with English abstract)
- [4] 吴世琛, 毛玉明, 胡慧中, 等. 基于改进 YOLO11n 的葡萄果叶病害检测方法[J]. *农业工程学报*, 2025, 41(14): 140-147.
WU Shichen, MAO Yuming, HU Huizhong, et al. Detecting grape leaf diseases using improved YOLO11n[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2025, 41(14): 140-147. (in Chinese with English abstract)
- [5] 刘芹, 欧阳婧玟, 周智彬, 等. 基于 YOLO-CRC 的茄子及茄子梗分割与定位方法[J]. *农业工程学报*, 2025, 41(19): 196-205.
LIU Qin, OUYANG Jingwen, ZHOU Zhibin, et al. Segmentation and localization method for eggplants and stems based on YOLO-CRC[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2025, 41(19): 196-205. (in Chinese with English abstract)
- [6] 吴擎, 韦润轩, 周乐, 等. 基于改进 YOLOv5s 的轻量化鲜茶叶识别方法[J]. *智能化农业装备学报 (中英文)*, 2025, 6(1): 1-14.
WU Qing, WEI Runxuan, ZHOU Le, et al. Lightweight fresh tea leaf recognition method based on improved YOLOv5s[J]. *Journal of Intelligent Agricultural Mechanization*, 2025, 6(1): 1-14. (in Chinese with English abstract)
- [7] 熊干, 陈慈发, 张上. QMDF-YOLO11: 复杂场景下水稻害虫检测算法[J]. *计算机工程与应用*, 2025, 61(13): 113-123.
XIONG Gan, CHEN Cifa, ZHANG Shang. QMDF-YOLO11: rice pests detection algorithm in complex scenarios[J]. *Computer Engineering and Applications*, 2025, 61(13): 113-123. (in Chinese with English abstract)
- [8] 陈金荣, 许燕, 周建平, 等. 基于 YOLO-SSAR 的自然环境下红花检测算法[J]. *农业工程学报*, 2025, 41(2): 215-223.
CHEN Jinrong, XU Yan, ZHOU Jianping, et al. Detecting safflower in the natural environment using YOLO-SSAR[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2025, 41(2): 215-223. (in Chinese with English abstract)
- [9] 王少聪, 王会强, 丁小明, 等. 基于改进 YOLOv8n 的草莓温室飞虫虫情监测方法[J]. *农业工程学报*, 2025, 41(17): 184-193.
WANG Shaocong, WANG Huiqiang, DING Xiaoming, et al. Pest monitoring in strawberry greenhouses using improved YOLOv8n[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2025, 41(17): 184-193. (in Chinese with English abstract)
- [10] 吕强, 林刚, 蒋杰, 等. 基于改进 YOLOv5s 模型的自然场景中绿色柑橘果实检测[J]. *农业工程学报*, 2024, 40(18): 147-154.
LYU Qiang, LIN Gang, JIANG Jie, et al. Detecting green citrus fruit in natural scenes using improved YOLOv5s model[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2024, 40(18): 147-154. (in Chinese with English abstract)
- [11] 李天华, 孙萌, 丁小明, 等. 基于 YOLO v4+HSV 的成熟期番茄识别方法[J]. *农业工程学报*, 2021, 37(21): 183-190.
LI Tianhua, SUN Meng, DING Xiaoming, et al. Tomato recognition method at the ripening stage based on YOLO v4 and HSV[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2021, 37(21): 183-190. (in Chinese with English abstract)
- [12] 任锐, 孙海霞, 张淑娟, 等. 基于改进 YOLOv8n 的不同栽培模式下玉露香梨轻量化检测[J]. *农业工程学报*, 2025, 41(5): 145-155.
REN Rui, SUN Haixia, ZHANG Shujuan, et al. Lightweight detection method for 'Yuluxiang' pear under different cultivation modes based on improved YOLOv8n[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2025, 41(5): 145-155. (in Chinese with English abstract)
- [13] 黄铭杰, 蔡伟强, 张子健, 等. 基于改进 YOLO11 的荔枝果实品种实时精准识别算法[J]. *农业工程学报*, 2025, 41(11): 156-164.
HUANG Mingjie, CAI Weiqiang, ZHANG Zijian, et al. Realtime accurate recognition algorithm for litchi fruit varieties based on improved YOLO11[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2025, 41(11): 156-164. (in Chinese with English abstract)
- [14] 戈刚, 杨珺, 刘毅, 等. 改进 YOLOv8n 模型对复杂农业场景下西红柿目标的检测[J]. *农业工程学报*, 2025, 41(9): 143-153.
GE Gang, YANG Jun, LIU Yi, et al. Detecting tomatoes in complex agricultural environments using improved YOLOv8n[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2025, 41(9): 143-153. (in Chinese with English abstract)
- [15] GAI R, CHEN N, YUAN H. A detection algorithm for cherry fruits based on the improved YOLO-v4 model[J]. *Neural Computing and Applications*, 2021, 35(19): 1-12.
- [16] HU H M, KAIZU Y, ZHANG H D, Xu Y W, et al. Recognition and localization of strawberries from 3D binocular cameras for a strawberry picking robot using coupled YOLO/Mask R-CNN[J]. *International Journal of Agricultural and Biological Engineering*, 2022, 15(6): 175-179.
- [17] CHEN C, LI J, LIU B, et al. A robust vision system for measuring and positioning green asparagus based on YOLO-seg and 3D point cloud data[J]. *Computers and Electronics in Agriculture*, 2025, 230: 109937.
- [18] SHEN Q, ZHANG X, SHEN M, et al. Multi-scale adaptive YOLO for instance segmentation of grape pedicels[J].

- Computers and Electronics in Agriculture*, 2025, 229:109712.
- [19] 张勤, 庞月生, 李彬. 基于实例分割的番茄串视觉定位与采摘姿态估算方法[J]. *农业机械学报*, 2023, 54(10): 205-215. ZHANG Qin, PANG Yuesheng, LI Bin. Visual positioning and picking pose estimation of tomato clusters based on instance segmentation[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2023, 54(10): 205-215. (in Chinese with English abstract)
- [20] 王文波, 单韵竹, 胡甜甜, 等. 基于目标区域语义分割的苹果采摘点定位方法[J]. *农业工程学报*, 2024, 40(24): 172-178. WANG Wenbo, SHAN Yunzhu, HU Tiantian, et al. Locating apple picking points using semantic segmentation of target region[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2024, 40(24): 172-178. (in Chinese with English abstract)
- [21] CUBUK E D, ZOPH B, SHLENS J, et al. Randaugment: practical automated data augmentation with a reduced search space[C] //2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle, WA, USA: IEEE, 2020: 3008-3017.
- [22] GAO S X, ZHANG P P, YAN T Y, et al. Multi-scale and detail-enhanced segment anything model for salient object detection[C]//Proceedings of the 32nd ACM International Conference on Multimedia. Melbourne, VIC, Australia: ACM, 2024.
- [23] YU Z, HUANG H, CHEN W, et al. YOLO-FaceV2: A scale and occlusion aware face detector[J]. *Pattern Recognition*, 2024, 155: 110714.

Multi-target segmentation, recognition, and localization for an asparagus-harvesting robot based on YOLO11n

LIANG Mingxu^{1,2,4}, XIA Xianfei^{1,4*}, YUAN Changrong², LYU Juntong^{1,4}, GONG Qingshuo^{1,4}, JIA Lei^{1,3,4}

(1. *Nanjing Institute of Agricultural Mechanization, Ministry of Agriculture and Rural Affairs, Nanjing 210014, China*; 2. *School of Mechanical Engineering, Nanjing Institute of Technology, Nanjing 211167, China*; 3. *College of Mechanical and Electronic Engineering, Northwest A & F University, Yangling 712100, China*; 4. *East China Agricultural Science and Technology Center, Suzhou 215300, China*)

Abstract: Asparagus harvesting can be confined to the efficacy of robotic vision in recent years. Asparagus spears are characterized by a slender morphology in their natural growth state. These tender stems are highly prone to mutual occlusion and overlapping when growing densely in field conditions. Furthermore, the stout mother stems can simultaneously present as the complex background interference. Collectively, it is often required for the high accuracy of the multi-target segmentation and recognition using machine vision. In this study, the lightweight instance segmentation model (YOLO11n-seg) was adopted as a baseline, in order to improve the precise positioning and harvesting performance of the robotic end-effector. Consequently, an optimized model named YOLO11n-SAL was also proposed to specifically tailor the slender, occluded targets with high fidelity. Two modules were introduced to enhance the feature extraction and attention mechanisms in the architectural framework. Firstly, the multi-scale edge enhancement Module (MEEM) was conceptually designed and integrated in order to mitigate the challenge wherein the edge features of the slender asparagus targets were inherently weak and easily lost during convolutional operations. Multi-scale decomposition was performed on the convolutional feature maps. The MEEM effectively extracted and intensified the edge and contour information before feature fusion. The sensitivity to the target boundaries was significantly elevated for the high segmentation precision, thereby enhancing the perceptual capability of the targets with the slender morphological structures. Secondly, the separated and enhancement attention module (SEAM) was introduced to rectify the feature confusion and data incompleteness caused by inter-target occlusion. Attention separation over both channel and spatial dimensions was also utilized to adaptively perceive the local and global features of the occluded asparagus at the varying scales. These features were selectively enhanced and effectively fused to better position the visible subjects of the partially masked targets, while suppressing the background noise and distractor information. The robust performance of the detection and recognition was maintained even within the complex and cluttered environments. A series of experiments was conducted to verify the effectiveness of the improved model. Quantitative evaluation results indicate that the improved YOLO11n-SAL model achieved significant gains over all key performance indicators, compared with the baseline model. In the detection task of the target bounding box, the superior performance was achieved with a detection precision of 94.2%, a recall rate of 83.1%, a mean average precision at IoU threshold 0.5 ($mAP_{0.5}$) of 91.2%, and a mean average precision at IoU threshold 0.5-0.95 ($mAP_{0.5-0.95}$) of 76.2%. In the more granular instance mask segmentation, the model also performed impressively. The segmentation precision, recall, $mAP_{0.5}$ and $mAP_{0.5-0.95}$ reached 93.4%, 77.9%, 90.7%, and 62.7%, respectively. Furthermore, the heatmap analysis demonstrated that the YOLO11n-SAL model was markedly improved to perceive the asparagus edge features over different scenarios, with the superior multi-target segmentation and recognition under occluded conditions. The high accuracy of the segmentation and recognition was achieved to reduce the interference in the complex multi-scenario environments, compared with the baseline. Finally, a series of asparagus recognition, positioning, harvesting, and grasping trials were carried out using depth cameras and mechanical arms, in order to validate the cognition and position performance in the actual deployment scenarios. The empirical results showed that a positioning success rate of not less than 90% was accompanied by effective harvesting and grasping performance. These findings can provide reliable technical support for the advancement of robotic harvesting in precision agriculture.

Keywords: multi-object detection; segmentation; green asparagus; YOLO11n improved model; precision harvesting; visual positioning