doi:10.6041/j.issn.1000-1298.2026.09.008

基于轻量级检测网络与显著性分割网络级联的鲜食甘薯轮廓提取方法

赵　昂[1,2]　张　健[1,2]　潘永菲[2,3]　朱宏飞[2,3]　文辉辉[1,2]　刘　洋[1,2]

（1.海南大学机电工程学院，海口 570228；2.农业农村部热带高效农业智能装备重点实验室，儋州 571737；
3.海南大学信息与通信工程学院，海口 570228）

摘要：鲜食甘薯是海南省重要的经济作物，采用机器人进行捡拾作业是提高其收获效率的有效手段。然而，在复杂的田间环境下，目标轮廓的准确提取仍存在鲁棒性较差和分割精度不高的问题。为此，提出了一种级联改进的检测—分割方法：以轻量化的 YOLO 11n 作为检测骨干，并与 BASNet（Boundary-aware salient Net）级联，旨在提高田间轮廓识别的准确性与稳定性。首先用轻量化的 FasterNet 替换 YOLO 11n 的原有主干，显著减少模型参数并提升推理速度；其次，在主干网络中引入多尺度、可变扩张率的 LSKA（Large separable kernel attention）模块，以灵活扩大感受野，增强对小目标及被植被遮挡目标的响应能力；然后，在中小目标检测头之前插入 SENetV2（Squeeze-and-excitation version 2）通道注意力模块，通过全局特征重标定进一步提升复杂背景下的目标信号提取能力；最后，将 BASNet 显著性分割网络级联至检测模块，对检测框区域进行像素级精细分割，有效剔除背景干扰。对比试验表明，改进后的 FLS－YOLO 11n 在显著降低模型体积与参数量、提高推理帧率的同时，实现了召回率提升 4.1 个百分点、平均精度提升 1.8 个百分点。级联至 BASNet 后，分割的平均绝对误差（MAE）约降低 40%，而最大 F 测度、最大结构相似性测度、最大增强对齐测度分别提升约 3.37%、3.31% 与 4.19%。最后，在机器人上进行试验，抓取成功率达到 90.6%。试验结果表明，该网络能够在复杂收获环境中实现较高的轮廓识别精度，为甘薯捡拾机器人在类似场景的工程化应用提供可行的技术路径。

关键词：鲜食甘薯；轮廓提取；捡拾机器人；YOLO 11n；语义分割

中图分类号：TP391.4　　**文献标识码**：A　　**文章编号**：1000-1298(2026)09-0082-11　　**OSID**：

Cascaded Lightweight Detection and Saliency Segmentation for Field Sweet-potato Contour Extraction

ZHAO Ang[1,2]　ZHANG Jian[1,2]　PAN Yongfei[2,3]　ZHU Hongfei[2,3]　WEN Huihui[1,2]　LIU Yang[1,2]

(1. *School of Mechanical and Electrical Engineering, Hainan University, Haikou* 570228, *China*
2. *Key Laboratory of Tropical High-efficiency Agricultural Intelligent Equipment, Ministry of Agriculture and Rural Affairs, Danzhou* 571737, *China*
3. *School of Information and Communication Engineering, Hainan University, Haikou* 570228, *China*)

Abstract: Fresh sweet potato is an important economic crop in Hainan Province, and the use of robots for picking operations is an effective means to improve harvesting efficiency. However, in complex field environments, accurate extraction of target contours still suffers from poor robustness and low segmentation accuracy. To address this issue, a cascaded improved detection-segmentation method was proposed: lightweight YOLO 11n was used as the detection backbone, cascaded with boundary-aware salient net (BASNet) to enhance the accuracy and stability of contour recognition in the field. Firstly, the original backbone of YOLO 11n was replaced with lightweight FasterNet, significantly reducing model parameters and improving inference speed. Secondly, a large separable kernel attention (LSKA) module with multi-scale and variable dilation rates was introduced into the backbone network to flexibly expand the receptive field, enhancing the response capability to small targets and targets occluded by vegetation. Then, the squeeze-and-excitation version 2 (SENetV2) channel attention module was inserted before the

收稿日期：2025－11－25　修回日期：2025－12－08

基金项目：国家自然科学基金项目(52265029)

作者简介：赵昂(1999—)，男，硕士生，主要从事农业机器人技术与装备研究，E-mail：1370660179@qq.com

通信作者：张健(1982—)，男，副教授，博士生导师，主要从事现代农业智能装备研究，E-mail：zhangjian_qau@163.com

small and medium target detection heads to further improve target signal extraction capability in complex backgrounds through global feature recalibration. Finally, the detector's bounding boxes were passed to BASNet for pixel-level salient segmentation, removing background noise and refining object contours. On benchmark comparisons, the improved FLS－YOLO 11n achieved a 4.1 percentage points increase in recall and a 1.8 percentage points gain in mAP while substantially reducing model volume and parameter count and improving inference FPS. After cascading with BASNet, segmentation MAE was reduced by roughly 40%, and $maxF_\beta$, $maxE_\varphi$, and $maxS_m$ was increase by 3.37%, 3.31% and 4.19%, respectively. Field trials on a harvesting robot produced a 90.6% grasp success rate. Results demonstrated that the proposed pipeline attained high contour recognition accuracy in complex harvesting environments, offering a practical technical path for engineering deployment of sweet potato picking robots.

Key words: fresh sweet potato; contour extraction; picking robotics; YOLO 11n; semantic segmentation

0　引言

甘薯是热带地区的重要粮食作物，尤其在亚洲、非洲及太平洋岛国广泛分布，占全球产量的95%以上[1]。中国是世界上最大的甘薯生产国。海南作为中国唯一的热带省份，历来是中国南方薯区的重要产区，所产甘薯多为鲜食型，对收获效率和品质控制具有较高要求。海南省中部以山地为主，向外逐步过渡至丘陵、台地和平原，呈环状分布[2]。该地区复杂的地形对提升热带地区农业生产效率与可持续发展造成明显制约，尤其在甘薯收获环节，大型机械难以高效作业，因此甘薯收获仍主要依赖人工。研究适用于海南省多样地形环境的甘薯捡拾机器人，对于提升收获效率、减轻劳动强度具有重要意义。

由于非结构化农业环境中的特殊工作条件，目标视觉信息的获取已成为限制采摘机器人在生产中应用的主要瓶颈之一[3-7]。近年来，为了实现采摘的高精度与高效率，众多科研人员针对各类农业机器人不断开发与完善视觉系统[8-10]，研究多在理想或模拟环境中开展，未充分考虑复杂田间现场的光照变化、背景干扰及物体遮挡等问题。

近几年，基于深度相机与二维轮廓识别的抓取点定位技术逐渐兴起[11-12]。然而，在复杂背景下准确提取作物轮廓依然困难。近年来，图像分割算法在作物轮廓提取方面的性能显著提升[13-14]。然而，模型均为针对不同作物及采摘需求进行定制，尚缺乏针对甘薯形态特征与现场环境的专用视觉系统。鉴于此，针对复杂地形与田间环境中的甘薯目标检测与轮廓识别问题，亟需研发高效且鲁棒的视觉方案。

本文旨在为甘薯收获场景构建高效、模块化的轮廓识别系统。首先，以 YOLO 11n[15] 作为级联检测模块的基线，并将其原有主干替换为轻量化 FasterNet[16]，以显著降低模型参数量、内存占用量和计算复杂度；同时，在主干网络中用多尺度可变扩张率的 LSKA 模块[17] 替换 YOLO 11 原有的注意力机制模块，以增强对被植被遮挡甘薯目标的检测能力；然后，在中小目标检测头前插入 SENetV2 注意力模块[18]，通过全局特征重标定进一步提升复杂背景下的信号提取效果，最终构建 FLS－YOLO 11n 检测网络，以在保持检测性能的同时显著降低模型复杂度并增强复杂背景下的特征响应能力。最后，选用 BASNet[19] 作为级联的显著性分割网络以实现像素级轮廓细化。将 FLS－YOLO 11n 与 BASNet 显著性分割网络级联，验证检测与分割联合体在甘薯识别与轮廓提取任务中的可行性，并将该算法作为装备级感知模块与机械系统对接，以期为甘薯捡拾机器人在复杂收获现场的轮廓识别提供技术路径和方法基础。

1　图像获取与预处理

针对热带地区主要的红壤土与砂壤土环境，本文使用 D435i 型深度相机，分别于 2024 年 10 月在海南大学智能农机装备研发基地红壤土试验田采集栽培品种为“广薯 87”的甘薯彩色图像和 2025 年 5 月在海南省儋州市海头镇海头地瓜省级现代农业产业园的砂壤土试验田采集栽培品种为“高系 14 号”的甘薯彩色图像，采集图像分辨率为 1 920 像素×1 080 像素，共采集 1 032 幅图像。采集后，剔除重复性高或拍摄效果不佳的图像。按红壤土与砂壤土各 500 幅的规模汇总整理，并随机抽取其中 800 幅作为训练集、100 幅作为验证集、100 幅作为测试集，确保两类土壤在各子集中的比例均为 1∶1。随后，通过翻转、模糊等数据增强手段随机组合对原始图像进行扩充，使总数据集数量扩充至 5 000 幅；最终扩充后的训练集达到 4 000 幅，验证集 500 幅，测试集 500 幅。针对遮挡

情况,对甘薯目标的遮挡程度进行了分级处理,并对各级别在数据集中的分布进行了统计与记录。遮挡分级覆盖从无遮挡到严重遮挡的多个等级,数据集中同时包含大量无遮挡样本与不同程度遮挡样本,以反映田间实际的复杂情况并提升模型在多种遮挡情形下的鲁棒性。

标注阶段,采用开源工具 LabelMe 对所有图像手工标注 YOLO 训练框与分割真值图,以保证标注的准确性和可重复性。数据集示例如图 1 所示。



图 1 数据集示例

Fig. 1 Dataset example

2 材料与方法

2.1 YOLO 11 网络

YOLO 11 模型主要由主干网络(Backbone)、颈部网络(Neck)和检测头部(Head)3 部分组成。主干网络承担图像特征提取的功能,检测头部延续了 YOLO v8 的基本架构,并引入了一种可配置的模块 C3k2。该模块通过布尔参数 C3k 控制其结构:当 C3k 设置为 False 时,C3k2 等同于标准的 C2F 模块;若设置为 True,则切换为 C3 模块,以更高效地完成特征提取。这一可切换机制增强了模型在不同任务和场景下的适应性与灵活性。在主干网络中,C2PSA 模块是一个关键组成部分。该模块在 C2F 结构基础上融合了 PSA(Pyramid split attention)模块,通过对输入特征施加前馈神经网络与多头自注意力,实现对重要信息的强化与冗余信息的抑制,从而提升模型对目标关键特征的捕捉能力,进一步提高检测精度。

2.2 甘薯检测算法改进

甘薯实际捡拾作业环境复杂,且甘薯与土壤颜色相近,导致目标特征提取难度较高,对检测网络的特征表达能力提出更高要求。同时,完整的级联检测算法需应对实际情况中对检测速度与模型体积的要求。为此,本文主要验证级联的可行性以及级联网络前端的轻量化设计,首先以 YOLO 11 中最轻量化的 n 系列作为该级联网络中前端网络的基线模型,进行级联网络前端轻量化设计。本文主要从以下 3 个方面进行优化改进,以在提升检测性能的同时显著降低计算成本。改进后的模型结构如图 2 所示。

首先,为解决模型参数量及内存占用量过大与

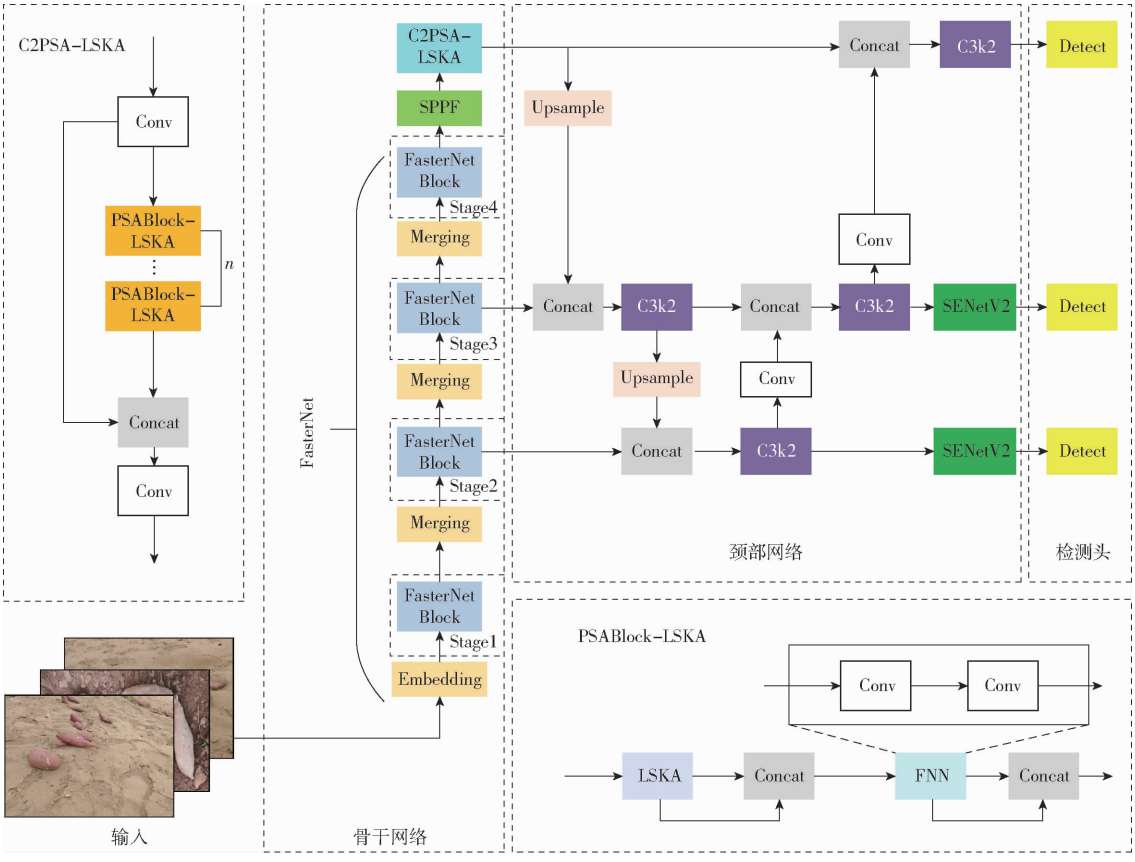


图 2 FLS-YOLO 11n 结构

Fig. 2 Structure of FLS-YOLO 11n

边缘截断或部分遮挡情况下对甘薯识别的鲁棒性不足的问题，将原有主干替换为轻量化的 FasterNet；其次，为应对复杂田间环境中的噪声干扰，将多尺度、可变扩张率的 LSKA 模块插入主干网络的 C2PSA 注意力机制中，以替代原有注意力层，其中 n 为 PSABlock－LSKA 串联个数，本文取 1；最后，针对中小目标识别率较低的问题，在中小目标检测头之前的颈部网络插入 SENetV2 通道注意力模块，从而构建了 FLS－YOLO 11n 检测网络。

2.2.1　FasterNet 主干网络

YOLO 11 原有的主干网络，存在复杂模型对简单任务的计算冗余，并且在边缘截断或部分遮挡情况下对甘薯识别的鲁棒性不足。而 FasterNet 中的部分卷积（Partial convolution，PConv）能够有效解决上述问题。如图 2 所示，FasterNet 包含 4 个 stage，每个 stage 之前都有一个 stride 为 4 的普通卷积（Embedding layer）或一个 stride 为 2 的普通卷积（Merging layer）用于下采样或增加通道数。每个 stage 都包括若干 FasterNet blocks。FasterNet 的核心在于引入 PConv 模块。PConv 仅在输入特征图的一部分通道上执行常规卷积，而将其余通道保留不变，从而同时降低浮点运算量并减少内存访问次数，进而显著提升移动端和嵌入式设备的实际推理吞吐率。如图 3 所示，PConv 的浮点运算量 F_p 计算公式为

$$F_p = hwk^2c_p^2 \tag{1}$$

式中　h——特征图在竖直方向上的尺寸

w——特征图在水平方向上的尺寸

c_p——参与部分卷积操作的通道数

k——卷积核尺寸

为了充分而高效地利用来自所有通道的信息，FasterNet 在 PConv 的后面添加了一个 PWConv。这两个结合起来使在输入上的有效感受野类似于 T 型卷积（TConv），如图 4 所示，与普通卷积相比它更关注于中心区域。TConv 的浮点运算量 F_T 计算公式为

$$F_T = hw[k^2c_p^2c + c(c - c_p)] \tag{2}$$

式中　c——输入特征图的通道数

应用了该卷积的 FasterNet 有助于降低模型复杂度、抑制背景噪声并强化目标边缘特征，特别是在甘薯边缘被截断或部分遮挡时，仍能完成精确的特征提取与定位。因此，本文基于 FasterNet 构建了一种适用于复杂田间环境下甘薯抓取点检测的轻量化网络，期望在保证高精度的同时实现更低的延迟和能耗。

2.2.2　LSKA 模块

由于 C2PSA 模块难以完全应对复杂田间环境

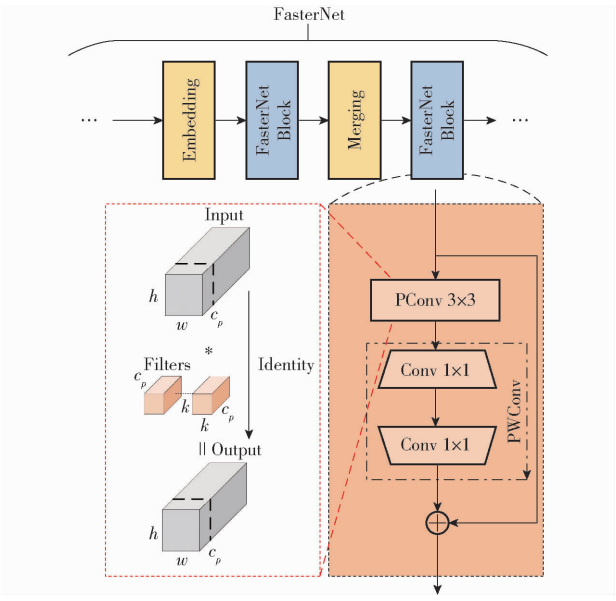


图 3　FasterNet 结构

Fig. 3　Structure of FasterNet

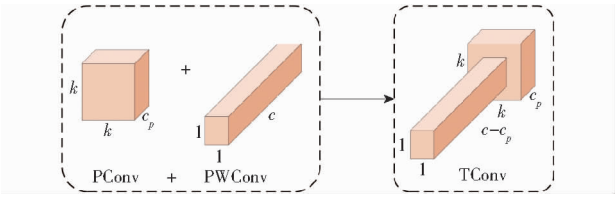


图 4　T 型卷积

Fig. 4　T-shaped Conv

中的噪声干扰，本文将 LSKA 模块替换 C2PSA 模块中 PSA 的注意力层。LSKA 模块是在原始 LKA（Large kernel attention）模块基础上改进而来，其结构如图 5 所示。该模块将 LKA 的两层大核操作分解为 4 层，每层由两个一维卷积组成：先使用水平方向的一维卷积对输入特征图进行卷积，再对其输出使用垂直方向的一维卷积；随后通过 1×1 卷积生成注意力图，最后将注意力图与原始特征图逐元素相乘以得到最终输出。LSKA 模块的浮点运算量 P 和参数量 F 计算方法为

$$P = 2(2d-1)C + 2\left\lfloor \frac{k}{d} \right\rfloor C + C^2 \tag{3}$$

$$F = \left(2(2d-1)C + 2\left\lfloor \frac{k}{d} \right\rfloor C + C^2\right)HW \tag{4}$$

式中　d——扩张率　　C——通道数

H——特征图高度

W——特征图宽度

这种设计在保持相同感受野的前提下，使参数量与浮点运算量明显下降，显著提升了模块的轻量化与实时性。更重要的是，LSKA 中 4 层可变扩张率（dilation rates 为 1、2、4、8）的设置，使网络具备自适应捕捉能力，可提取从细微局部特征到全局上下文的多层次信息，进一步增强对不同尺度目标的注

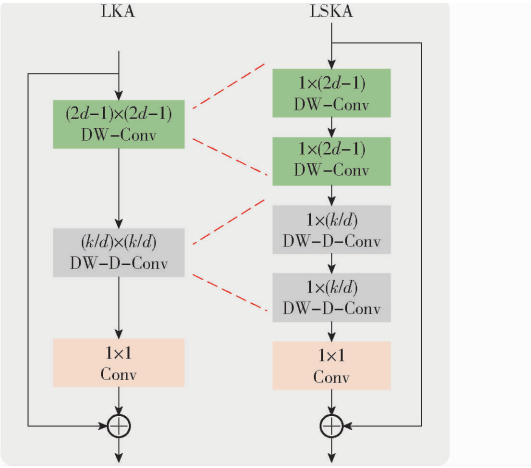


图 5 LSKA 模块

Fig. 5 LSKA module

意力响应。

将主干网络中 C2PSA 的注意力层替换为 LSKA 后,4 层一维卷积分解不仅保持了模型的轻量化与实时性,还能显著扩大感受野并灵活调节扩张率,提高了对复杂田间环境中噪声干扰的应对能力。

2.2.3 SENetV2 注意力模块

原颈部网络对背景干扰的抑制能力较弱,导致传输到检测头前的特征中仍保留较多无效信息,从而影响中小目标的识别效果。因此,本文在中小目标检测头之前的颈部网络中插入 SENetV2 模块以提升特征重标定能力。SENetV2 结构如图 6 所示。在经典通道注意力模块的基础上,使用全局平均池化与轻量化的 1×1 卷积替代全连接层,并对特征图逐通道重加权,以增强有用特征并抑制噪声。鉴于甘薯中小目标尺寸较小且与背景对比弱,将 SENetV2 插入中小目标检测头之前,可动态放大对目标细节响应显著的通道、削弱土壤与植被的干扰,从而显著提升中小尺寸甘薯的检测率与定位精度。

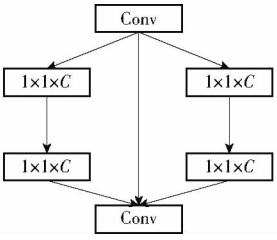


图 6 SENetV2 注意力模块

Fig. 6 Squeeze-and-excitation version 2 attention module

2.3 BASNet 显著性目标检测网络

由于甘薯捡拾存在收获环境复杂、背景杂物多且与土壤颜色接近等问题,因此需要更高的特征提取能力与边缘识别能力。其结构简图如图 7 所示。BASNet 将编码器解码器结构的分割能力与预测模块到精细化模块的分段式检测相结合,并通过融合 BCE、SSIM 和 IoU 3 重混合损失,有效突出显著区域并精准捕捉曲折边界。此外,基于 BASNet 的改进模型如 U^2 − Net[20]、Mobile BASNet[21] 提供了丰富的预训练权重与工程示例,有利于训练与推理的可复现性与工程化部署。因此,本文选用 BASNet 作为级联分割网络的基础模型,以兼顾边界精度与实际工程适用性。

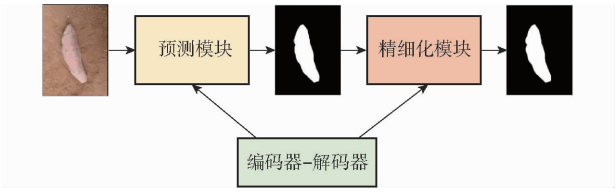


图 7 BASNet 结构简图

Fig. 7 Schematic of BASNet architecture

2.4 级联网络

本文中的级联网络首先在轻量级的 FLS − YOLO 11n 网络上快速定位甘薯目标,然后将检测框映射回原图并裁剪出高分辨率子图,交由 BASNet 进行精细分割。这种两阶段处理既显著降低了对整图高分辨率分割的计算开销,又最大程度保留了细节特征,从而提升轮廓重建的准确性。同时,检测与分割模块相互独立、可灵活替换与升级,能够降低复杂背景下的误检与漏检概率;并依托成熟的开源模型与预训练权重,保证流程的稳定性与可复现性。整体而言,该级联方案在速度、精度、灵活性与工程可用性之间实现了良好平衡,更适合对显著目标轮廓要求较高的场景。

如图 8 所示,原始图像输入网络前通常被统一缩放为 640×640 分辨率。该下采样会丢失部分细节,进而影响后续分割效果。为兼顾网络输入统一性与细节保留,本文采用先缩放检测、再回投裁剪的策略:将原图缩放为 640×640 后输入 FLS − YOLO 11n 进行检测;随后对检测到的框坐标进行归一化并映射回原始图像,以裁剪出对应的高分辨率子图,最后将这些子图传入 BASNet 进行精细分割。该流程既保证了输入尺寸统一,又最大限度保留了原始图像的细节信息,从而提高分割与轮廓恢复的精度。

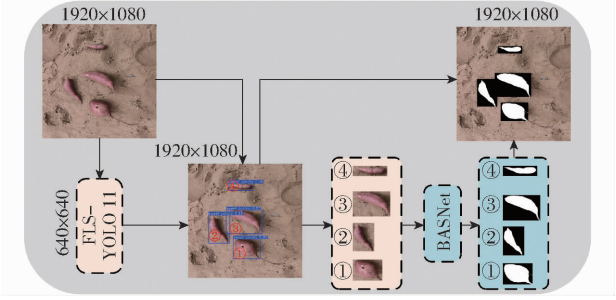


图 8 级联网络原理图

Fig. 8 Schematic of cascaded network

2.5　视觉模块与装备系统集成

如图 9 所示，本文将所提 FLS－YOLO 11n＋BASNet 级联网络作为甘薯捡拾机器人的核心感知单元，结合深度相机完成 3 维抓取点的标定，实现了从感知到执行的工程对接。

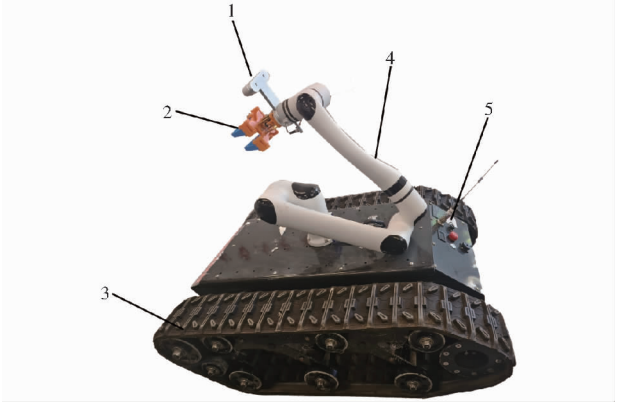


图 9　平台结构图

Fig. 9　Platform structural diagram

1. 深度相机　2. 鱼鳍式机械爪　3. 履带底盘　4. 机械臂　5. 控制与信号接收系统

该系统的采集模块为 Intel RealSense D435i 型深度相机，相机通过刚性安装固定于机械臂末端。采用标准的相机-机械臂外参标定方法建立像素坐标到机器人末端坐标系的映射，保证像素级轮廓可以准确映射到 3 维抓取点。彩色图像经 FLS－YOLO 11n 实时检测，检测框被回投至原图并裁剪后由 BASNet 进行高分辨率分割；分割结果经边缘提取后转化为封闭轮廓。

抓取点基于分割轮廓与深度信息计算得到，系统会估算目标质心、主轴方向以及若干候选抓取位，并通过几何分析确定甘薯轮廓的主轴线以选择最稳健的抓取方案。机械爪选用市面上应用普遍的鱼鳍式机械爪。视觉模块的输出经控制器转换为机械臂轨迹与夹爪开合指令，夹爪实行位置/力闭环控制，进行甘薯的抓取。

2.6　深度相机抓取点三维坐标提取

试验采用型号为 D435i 的深度相机对甘薯进行定位点及抓取方向的确定，如图 10 所示。甘薯的形状可近似为纺锤体，以果实的质心（点 E）为坐标系原点建立甘薯横截面坐标系（X_p，Y_p，Z_p），以相机视觉点 O_c 为原点建立图像像素坐标系（X_c，Y_c，Z_c）。点 A、B 为甘薯长轴端点，点 D 为靠近相机一侧果实外表面的投影中心对应的空间点。根据投影原理，甘薯截面的点 A、B、C、D 在显著图中的投影分别为 a、b、c、d。主轴线 PP'通过主成分分析确定，cd 为通过显著图分析获得的机械手抓取方向。深度相机可提供点 D 的深度，该深度用于精确控制机械手的抓取深度与位姿。

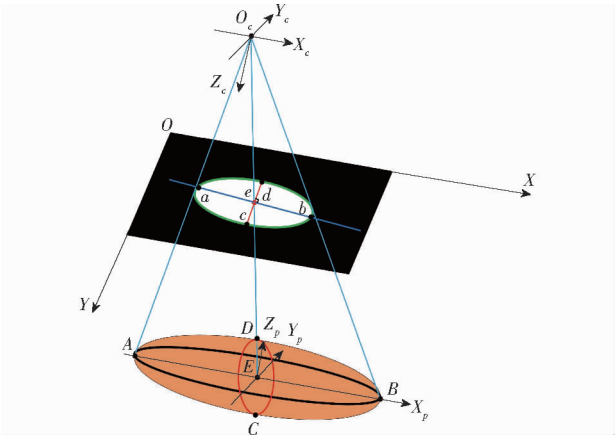


图 10　甘薯抓取点识别示意图

Fig. 10　Identification diagram of sweet potato grasping point

3　模型训练与评价指标

3.1　试验环境和参数配置

试验平台为 Windows 11 系统，采用 PyTorch－1.13.1 学习框架。CPU 型号为 Intel（R）Xeon（R）Silver 4316 CPU @ 2.30GHZ；GPU 型号为 NVIDIA GeForce RTX 4090，24 GB 显存；CUDA 版本为 12.7.29；Python 版本为 3.10。检测网络训练时，兼顾训练效率及硬件显存大小，设置训练批次（Batch size）为 36；本文注意到当训练轮数（Epochs）为 200 轮时，损失值曲线趋于收敛状态，因此将检测网络训练轮数设定为 200。分割网络训练时，首先将每幅图像的大小调整为 512×512，然后随机裁剪为 224×224。部分编码器参数由 ResNet－34[22] 模型初始化。在不使用验证集的情况下训练网络直到损失收敛。

3.2　评价指标

在检测网络方面，本文采用精确率与召回率来衡量模型的检出准确性与完整性，并以平均精度（Mean average precision，mAP）来综合评估定位与分类性能；同时考察模型参数量和浮点运算量以衡量存储及计算开销，并通过帧率评价检测的实时性。

对于分割网络，本文引入 F 测度[23]、增强对齐测度[24]、结构测度[25] 和平均绝对误差（Mean absolute error，MAE）[26] 等多维度指标来量化预测掩码与真实掩码在像素级差异、精确性与召回性、结构一致性及整体形状相似度上的表现，从而全面评估在加入检测模块后对分割精度的影响。

其中，F 测度可以有效反映模型在检测到的轮廓是否完整和准确，适用于对模型综合性能的评估。计算公式为

$$F_\beta = \frac{(1+\beta^2)P_r R_1}{\beta^2 P_r + R_1} \tag{5}$$

式中 P_r——查准率 R_1——查全率

β——常系数,取 0.3

增强对齐测度是一种更全面的评价指标,结合了图像的像素级别和区域级别的相似性,能更好地反映全局感知质量。在甘薯轮廓提取中,它能够反映模型对整体形状和结构的感知能力,适用于评估模型在处理复杂背景和形状变化时的鲁棒性。计算公式为

$$E_\phi = \frac{1}{WH}\sum_{i=1}^{W}\sum_{j=1}^{H}\phi(P(i,j),G(i,j)) \tag{6}$$

式中 $P(i,j)$——显著性检测结果在位置(i,j)的灰度

$G(i,j)$——真值图在位置(i,j)的灰度

$\phi(P,G)$——增强一致性函数

结构测度用于衡量显著性检测结果的结构相似性。它特别适合评估模型在保持整体结构的前提下进行准确提取的能力,对于甘薯这种具有明确边界和形状的目标尤为重要。计算公式为

$$S = \alpha S_o + (1-\alpha)S_r \tag{7}$$

式中 S_o——对象相似性指数

S_r——区域相似性指数

α——权重因子,取 0.5

4 试验

4.1 骨干网络引入不同注意力性能对比

为验证引入轻量级 LSKA 模块的有效性,选取同样轻量级且具代表性的空间注意力模块 CAA[27](Coordinate anchor attention)、SEAM[28](Separated and enhancement attention module)和 EMA[29](Enhanced multi-scale attention)在相同试验设置下与 LSKA 模块进行对比试验。对比结果如表 1 所示。

表 1 骨干网络替换不同注意力的模型测试结果

Tab.1 Test results of model for substitution of different attention by backbone network

模型	精确率/%	召回率/%	mAP/%	浮点运算量	帧率/(f·s^{-1})
YOLO 11n	93.2	90.7	95.9	6.4×10^9	204
YOLO 11n-CAA	92.2	92.2	95.8	6.3×10^9	277
YOLO 11n-EMA	92.2	93.2	95.6	6.3×10^9	254
YOLO 11n-SEAM	92.5	92.0	95.1	6.3×10^9	223
YOLO 11n-LSKA	93.4	93.9	96.3	6.3×10^9	285

试验结果表明,与其他主流注意力机制对比,LSKA 模块在保持计算开销少量提升的前提下,对目标轮廓敏感度和漏检率的综合提升最为显著。在主干网络的 C2PSA 模块的注意力层替换为 LSKA 后,精确率、召回率和 mAP 均有所提高。同时在替换注意力层后在参数量变化不大的情况下,替换后的帧率提升至 285 f/s,鉴于此提升对实时性要求高的机器人检测与抓取场景十分有利,故本文选择以 LSKA 替换原注意力层。

4.2 原模型对比实验

为验证各改进模块在甘薯检测上的有效性,以 YOLO 11n 为基准模型,对各改进模块进行消融试验,主要以精确率、召回率、mAP、参数量、浮点运算量、帧率作为参考指标,结果如表 2 所示。

表 2 消融试验对比结果

Tab.2 Comparative results of ablation experiment

FasterNet	LSKA	SENetV2	精确率/%	召回率/%	mAP/%	参数量	帧率/(f·s^{-1})	浮点运算量
×	×	×	92.2	91.7	94.9	2.58×10^6	204	6.4×10^9
√	×	×	91.8	91.5	94.2	2.37×10^6	249	5.5×10^9
×	√	×	93.4	93.9	96.3	2.55×10^6	285	6.3×10^9
×	×	√	93.7	93.3	95.2	2.29×10^6	193	6.2×10^9
√	×	√	93.1	93.2	95.1	2.31×10^6	225	5.5×10^9
√	√	×	93.2	92.7	95.4	2.34×10^6	267	5.5×10^9
√	√	√	93.9	95.8	96.7	2.27×10^6	251	5.5×10^9

注:√代表使用该模块,×代表不使用该模块。

消融试验结果表明,本文所采用的多种模块改进对模型性能产生了显著影响。具体而言,将轻量化主干网络 FasterNet 引入模型后,参数量减少了 8.1%,浮点运算量降低了 14.1%,但 mAP 也相应下降了约 0.7 个百分点。这一结果说明,尽管轻量化设计有效降低了模型复杂度和计算负担,但也部分削弱了模型对关键特征的提取能力。单独引入 LSKA 模块则能够在计算量增加不显著的情况下,显著提升多项性能指标:精确率提升 1.2 个百分点,召回率提高 2.2 个百分点,mAP 增长 1.4 个百分点,显示出该模块在增强特征表征能力方面的优势。而 SENetV2 的引入使参数量减少 11.2% 的同时,还带来了精确率 1.5 个百分点和 mAP 0.3 个百分点的小幅提升,但其帧率从 204 f/s 略微下降至 193 f/s,

说明其在计算效率上存在一定权衡。而将 FasterNet 与 LSKA 或 SENetV2 组合使用，能够有效弥补单一模块的局限性。例如，FasterNet 与 LSKA 组合在保持浮点运算量为 5.5×10^9 的同时，将帧率提高至 267 f/s，mAP 恢复至 95.4%，显示出良好的速度-精度平衡。同样，FasterNet 与 SENetV2 的组合在相同计算开销下将 mAP 提升至 95.1%。

最终，同时集成 FasterNet、LSKA 和 SENetV2 的改进模型，在多个关键指标上均优于基线模型：精确率提升 1.7 个百分点、召回率提高 4.1 个百分点、mAP 增长 1.8 个百分点，参数量减少 12.0%，浮点运算量降低 14.1%，帧率从 204 f/s 提升至 251 f/s。这一综合结果充分证明了各模块间的协同效应，能够在控制模型复杂度的同时显著提升检测性能。

本文进一步在不同土壤环境下对模型性能进行了可视化对比分析。试验结果如图 11 所示。

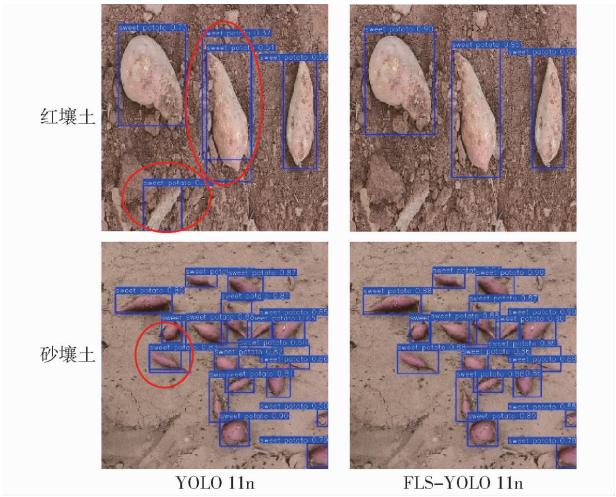


图 11　YOLO 11n 改进后图像检测结果

Fig. 11　Image detection results before and after YOLO 11n improvement

原始 YOLO 11n 模型在不同土壤条件中均出现明显的重复检测与误检现象，尤其是在目标边缘模糊或存在遮挡时更为显著。相比之下，FLS－YOLO 11n 不仅在整个测试场景中表现出更高的检测置信度，误检和漏检数大幅降低。这些可视化对比充分表明，所提出的改进模块能够有效提升模型在复杂农业环境中的泛化能力和可靠性。

4.3　主流模型性能对比

为全面评估本文提出的改进模型 FLS－YOLO 11n 在甘薯检测任务中的性能，将其与当前多种主流目标检测模型进行了详细对比，包括两阶段经典模型 Faster R－CNN[30]、轻量级单阶段检测器 SSD[31]、基于 Transformer 结构的 Rt－DETR[32]，以及 YOLO 系列近期代表作 YOLO v8s[33] 和 YOLO v10n[34]。对比结果（表 3）表明，FLS－YOLO 11n 在轻量化与综合检测性能方面均表现最优：其浮点运算量仅为 5.5×10^9，参数量约 2.2×10^6，帧率达 251 f/s；在检测精度方面，mAP 达到 96.7%，精确率为 93.9%，召回率为 95.8%，在所有对比模型中处于领先水平。相比之下，YOLO v10n 的 mAP 为 92.9%，帧率为 221 f/s；YOLO v8s 的 mAP 略高，为 96.2%，但帧率仅为 201 f/s。Rt－DETR 虽然凭借全局建模能力在部分场景中表现良好，但其结构计算冗余较多，帧率仅为 152 f/s，约为 FLS－YOLO 11n 的 60.5%，这一差距突显了本文轻量化设计的实际价值。与传统重型检测模型相比，FLS－YOLO 11n 的优势更为明显：其计算量约为 Faster R－CNN 的 4.1%、SSD 的 17.1%，推理速度提升 4.6～5.6 倍，同时所有精度指标均显著优于这两类模型。

综上所述，FLS－YOLO 11n 不仅在参数量、计算效率和推理速度方面表现出优秀的轻量化特性，同时还在精确率与召回率方面全面提升，在真实农业环境中具备显著的应用潜力与竞争力。

4.4　轮廓识别系统验证

为评估级联网络对分割精度的影响，本文将其生成的拼接图像进行二值化处理，并与原始分割网络的输出结果、典型端到端模型 Deeplabv3 +[35]、与 YOLO 11 系列端到端模型 YOLO 11m－seg[15] 进行比较。对比指标包括最大 F 测度、最大结构相似性测度、最大增强对齐测度以及平均绝对误差（MAE），结果如表 4 所示。

表 3　主流模型对比

Tab. 3　Comparison of mainstream models

模型	精确率/%	召回率/%	mAP/%	参数量	帧率/(f·s^{-1})	浮点运算量
Faster R－CNN	92.5	92.6	93.1	4.18×10^7	45	1.343×10^{11}
SSD	91.3	91.5	92.2	2.47×10^7	70	3.21×10^{10}
Rt－DETR	92.3	92.8	93.1	2.01×10^7	152	6.12×10^{10}
YOLO v8s	91.6	93.5	96.2	9.8×10^6	201	2.33×10^{10}
YOLO v10n	88.3	88.8	92.9	2.2×10^6	221	6.5×10^9
FLS－YOLO 11n	93.9	95.8	96.7	2.2×10^6	251	5.5×10^9

表 4 原模型对比

Tab. 4 Comparison of original model

模型	MAE	最大 F 测度	最大结构相似性测度	最大增强对齐测度
BASNet	0.170	0.940 4	0.932 3	0.946 6
YOLO 11m - seg	0.147	0.914 4	0.891 5	0.913 6
Deeplabv3 +	0.132	0.921 4	0.934 4	0.925 4
FLS - YOLO 11n + BASNet（低分辨率）	0.143	0.957 2	0.951 8	0.951 1
FLS - YOLO 11n + BASNet（高分辨率）	0.102	0.972 1	0.971 3	0.979 7

由表 4 所示的试验结果可知，本文提出的级联 FLS - YOLO 11n 与 BASNet 的融合模型在轮廓识别任务中实现了多项性能提升：在保持高精度分割能力的前提下，其最大 F 测度达到了 0.972 1，MAE 从 0.170 降低至 0.102，降幅达 40%；同时，最大结构相似性测度和最大增强对齐测度也分别提升了 4.2% 和 3.5%。而进行级联后的网络参数明显高于 Deeplabv3 + 与 YOLO 11m - seg 网络。这些量化指标一致表明，所采用的级联策略有效增强了对目标边界的识别能力。

为进一步验证性能提升，图 12 给出了改进模型与基准模型的可视化对比。将预测结果二值化并列展示后可见：原始 BASNet 存在明显的漏检，尤以目标的遗漏和边界处的模糊最为突出；端到端模型 Deeplabv3 + 、YOLO 11m - seg 在边缘定位上不够完整，且易产生重复识别。相比之下，所提级联模型在视觉上显著增强了目标轮廓的完整性与清晰度，其分割结果与人工标注表现出更高的一致性。

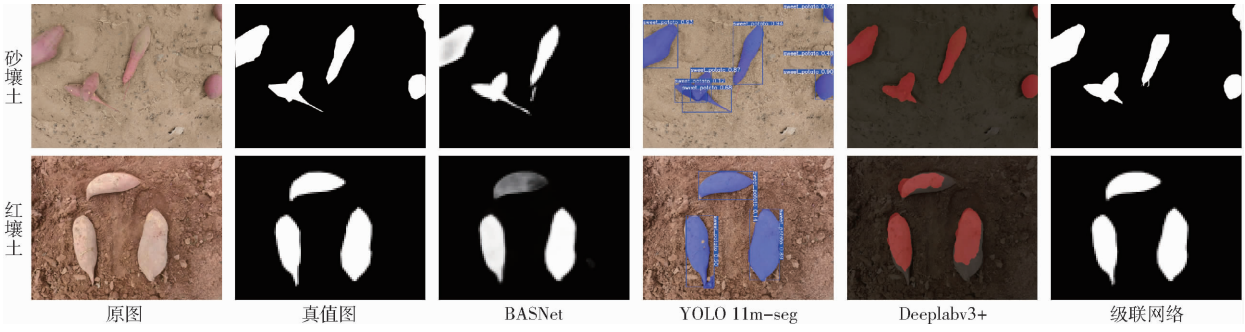


图 12 分割可视化对比图

Fig. 12 Comparison of segmentation visualization

为进一步探究性能提升的来源，设计了一组对照试验，旨在判断在级联 BASNet 后分割性能的提升是由 BASNet 本身较强的边界刻画能力驱动，还是源自“先检测并裁剪高分辨率区域”的处理流程。分别对 YOLO 剪裁后直接分割与 YOLO 剪裁后经高分辨率归一化再分割进行了对比试验。结果（表 4）显示，低分辨率分割在各项指标上均低于高分辨率分割。为进一步分析其原因，对分割流程进行了可视化展示，如图 13 所示。

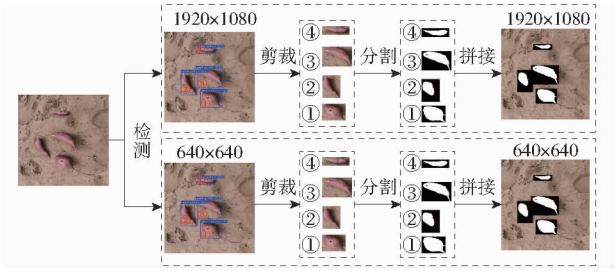


图 13 分割流程可视化展示

Fig. 13 Visualization of segmentation process

由可视化分析可知，YOLO 剪裁后直接分割的处理方式，会因识别阶段的分辨率损失而导致分割物体的边缘模糊等特征遗失现象，从而削弱分割效果。

4.5 平台部署与验证

本文基于 2.5 节所述平台，选取 100 个甘薯样本开展抓取试验。试验分别在海南大学智能农机装备研发基地的红壤土试验田与砂壤土模拟场地进行，现场环境如图 14 所示。



(a) 砂壤土 (b) 红壤土

图 14 甘薯抓取试验

Fig. 14 Grasping experiment of sweet potato

为准确统计抓取结果并系统评估系统性能，将甘薯样本划分为 10 组，每组 10 个，并随机抛于地面，以模拟实际田间条件下甘薯的分布特征。为进一步考察系统在不同环境光照下的稳定性，选择在 07:00（早）、12:00（中）与 18:30（晚）3 个典型光照时段分别对各组样本执行抓取任务。

抓取成功的判定标准为：机械手能够将目标甘薯从初始位置完整拾起，并稳定、准确地转移至收集

箱内。为保障试验结果的统计可靠性及可重复性，每组样本在每个光照时段均重复进行 3 次抓取试验，累计完成 300 次有效抓取。田间试验结果显示，系统整体抓取成功率约为 90.6%。其中不同土壤抓取结果见表 5。

表 5　抓取试验结果

Tab. 5　Results of capture experiment　%

模拟时间	不同土壤成功率	
	砂壤土	红壤土
早	91.3	89.7
中	93.7	92.0
晚	89.3	87.7
平均值	91.4	89.8

由表 5 可知，试验环境与试验时间均对抓取实验结果产生影响。具体分析如下：

在试验环境方面，在砂壤土环境中的抓取成功率高于红壤土。数据分析显示，砂壤土条件下的图像分割效果更优，抓取点识别成功率也更高。红壤土环境中因存在较多杂草、枯叶等地面干扰物，会遮挡甘薯的可见区域，进而影响其质心估计，导致偏抓或漏抓。尽管如此，整体抓取成功率仍达到 90.6%，表明本文所提出的级联抓取模型在面对复杂环境背景时具有良好的稳定性和鲁棒性。

在时间因素方面，不同采集时段的抓取成功率存在差异，其中 12:00 的成功率高于 07:00，而 18:30 的成功率最低。试验中发现，光照条件变化主要影响深度相机的信息提取效果，从而对抓取操作产生影响。尽管如此，改进后的系统在不同自然光照条件下仍能保持一定的抓取性能，说明其对自然光变化具备一定鲁棒性。此外，汇总试验数据时发现，个别甘薯的抓取成功率极低，经分析其原因是目标物体的物理尺寸超出夹爪的有效抓取范围。

综上所述，本文提出的级联抓取模型在面对不同形态的甘薯和复杂的农田环境时，表现出良好的抓取稳定性与鲁棒性，验证了其在实际应用中的可行性与推广价值。

5　结论

（1）对复杂田间环境中甘薯轮廓识别易受背景干扰和遮挡影响的问题，本文提出了一种级联 FLS－YOLO 11n 与 BASNet 的融合架构。研究结果表明，结合检测框引导的 BASNet 分割后，MAE 从 0.170 降至 0.102，最大 F 测度、最大结构相似性测度、最大增强对齐测度分别提升约 3.37%、3.31% 和 4.19%，有效缓解了漏分割与边界模糊问题。

（2）在对级联网络前端进行轻量化改造后，与原始 YOLO 11n 相比，精确率提升约 1.7 个百分点、召回率提升约 4.1 个百分点、mAP 提升约 1.8 个百分点；同时参数量约减少 12.0%，帧率约提升 23.0%，浮点运算量约降低 14.1%。

（3）所提出的 FLS－YOLO 11n＋BASNet 视觉模块已在甘薯捡拾机器人平台上完成端到端工程验证，结果表明该模块在热带复杂地形与光照条件下能够与机械臂-夹爪系统可靠对接。该系统在田间测试中的抓取成功率约为 90.6%，验证了所提视觉-抓取流水线在复杂收获环境中的实用性与工程可行性。

参　考　文　献

[1] NOREEN S, NIAZI M K, SHEHZADI S, et al. Extend the emphasis to encompass sweet potatoes health advantages, industrial applications, and nutritional qualities[J]. CyTA-Journal of Food, 2024, 22(1): 1－19.

[2] 秦建军，戴起伟，汪翔，等. 海南甘薯产业发展特征与前景[J]. 农业展望，2020，16(10)：77－81.
QIN Jianjun, DAI Qiwei, WANG Xiang, et al. Characteristics and prospects of the Hainan sweet potato industry[J]. Agricultural Outlook, 2020, 16(10): 77－81. (in Chinese)

[3] WANG Z H, XUN Y, WANG Y K, et al. Review of smart robots for fruit and vegetable picking in agriculture[J]. International Journal of Agricultural and Biological Engineering, 2022, 15(1): 33－54.

[4] YANG R, ZHAO A, LV D, et al. Sweet potato grasping point recognition in complex backgrounds: a method based on salient object detection[J]. Frontiers of Agricultural Science and Engineering, 2026, 13(2): 25653 (in Chinese)

[5] PAN Y, YANG R, ZHANG H, et al. Experiment and optimization of the potato-soil separation and conveying device for a harvester using RecurDyn－EDEM coupling simulation[J]. International Journal of Agricultural and Biological Engineering, 2025, 18(4): 149－156. (in Chinese)

[6] 安麒麟，汪风珠，刘阳春，等. 基于轻量级 RepVIT 的农机具工况识别方法研究[J]. 农业机械学报，2025，56(2)：187－194，205.
AN Qilin, WANG Fengzhu, LIU Yangchun, et al. Lightweight RepVIT-based working condition recognition method for agricultural implements[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(2): 187－194, 205. (in Chinese)

[7] 孔德航，刘云强，崔巍，等. 基于改进 YOLO v8n 的辣椒穴盘育苗播种质量轻量级检测方法[J]. 农业机械学报，2025，56(2)：381－392.
KONG Dehang, LIU Yunqiang, CUI Wei, et al. Lightweight detection method for seeding quality of chili seedling trays based on

improved YOLO v8n[J]. Transactions of the Chinese Society for Agricultural Machinery,2025,56(2):381 - 392. (in Chinese)
[8] GUO C M, ZHU C H, LIU Y, et al. End-to-end lightweight Transformer-based neural network for grasp detection towards fruit robotic handling[J]. Computers and Electronics in Agriculture, 2024, 221:109014.
[9] WANG F, URQUIZO R C, ROBERTS P, et al. Biologically inspired robotic perception-action for soft fruit harvesting in vertical growing environments[J]. Precision Agriculture, 2023, 24(4):1072 - 1096.
[10] LU A, LV X L, MA L Z, et al. An effective picking point localization method for multi-posture lotus pods based on three-view depth vision observation[J]. Computers and Electronics in Agriculture, 2024, 227(1):109492.
[11] 罗陆锋,邹湘军,叶敏,等 基于双目立体视觉的葡萄采摘防碰空间包围体求解与定位[J]. 农业工程学报, 2016, 32(8):41 - 47.
LUO Lufeng, ZOU Xiangjun, YE Min, et al. Calculation and localization of bounding volume of grape for undamaged fruit picking based on binocular stereo vision[J]. Transactions of the CSAE, 2016, 32(8):41 - 47. (in Chinese)
[12] ZHU Y, ZHANG T, LIU L, et al. Fast location of table grapes picking point based on infrared tube[J]. Inventions, 2022, 7(1):27.
[13] HUANG X, PENG D, QI H, et al. Detection and instance segmentation of grape clusters in orchard environments using an improved Mask R - CNN model[J]. Agriculture, 2024, 14(6):918.
[14] WANG Z F, ZHANG Z H, LU Y Q, et al. SE - COTR: a novel fruit segmentation model for green apples application in complex orchard[J]. Plant Phenomics, 2022, 2022:0005.
[15] JJOCHER G,NISHIMURA K,MINEEVA T,et al. YOLO 11[EB/OL]. 2024. https://github.com/ultralytics/ultralytics.
[16] CHEN J, KAO S H, HE H, et al. Run, Don't walk: chasing higher FLOPS for faster neural networks[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 9622 - 9631.
[17] LAU K W, PO L M, REHMAN Y A U. Large separable kernel attention: rethinking the large kernel attention design in CNN [J]. Expert Systems with Applications, 2023, 236: 121352.
[18] NARAYANAN M. SENetV2: aggregated dense layer for channelwise and global representations[J/OL]. CoRR, 2023, abs/2311.10807.
[19] QIN X, ZHANG Z, HUANG C, et al. BASNet: boundary-aware salient object detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, 2019: 7471 - 7481.
[20] LIU Y, WANG J, ZHANG H, et al. SEU^2 - Net: squeeze-and-excitation enhanced U^2 - Net for liver lesion segmentation in ct images[J]. Medical Image Analysis, 2022, 82: 102583.
[21] NGUYEN T, NGUYEN B, PHAM V, et al. Mobile BASNet: a lightweight boundary-aware salient object detection network [C]// Proceedings of the IEEE International Conference on Image Processing (ICIP). Anchorage, 2021: 1234 - 1238.
[22] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, 2016: 770 - 778.
[23] ACHANTA R, HEMAMI S, ESTRADA F, et al. Frequency-tuned salient region detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2009:1597 - 1604.
[24] FAN D P, GONG C, CAO Y, et al. Enhanced-alignment measure for binary foreground map evaluation[C]// Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI). Stockholm: IJCAI, 2018:698 - 704.
[25] FAN D P, CHENG M M, LIU Y, et al. Structure-measure: a new way to evaluate foreground maps[J/OL]. arXiv preprint, 2017, arXiv:1708.00786.
[26] HE K, CHEN X, XIE S, et al. Masked autoencoders are scalable vision learners[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE,2022:4065 - 4074.
[27] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021:8159 - 8168.
[28] XIE J, XIE X, WU X, et al. An improved YOLOv8-based method for detecting pests and diseases on cucumber leaves in natural backgrounds[J]. Sensors, 2025, 25(5):1551.
[29] OUYANG D, HE S, ZHANG G, et al. Efficient multi-scale attention module with cross-spatial learning[J/OL]. arXiv preprint, 2023, arXiv:2305.13563.
[30] REN S, HE K, GIRSHICK R, et al. Faster R - CNN: towards real-time object detection with region proposal networks[J]. Advances in Neural Information Processing Systems,2015,29: 91 - 99.
[31] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]// European Conference on Computer Vision (ECCV). LNCS 9905, 2016:21 - 37.
[32] ZHAO Y, LV W, XU S, et al. DEtrs beat yolos on real-time object detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024:7263 - 7271.
[33] Ultralytics. YOLOv8: high-performance real-time object detection models[R]. Ultralytics, 2023.
[34] WANG A, CHEN H, LIU L, et al. YOLOv10: real-time end-to-end object detection[J/OL]. arXiv preprint,2024,arXiv: 2405.14458.
[35] 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), 2018: 801 - 818.