

基于 YOLO v8n-seg 与轻量化 CNN 的草莓成熟度级联检测方法

洪旗^{1,2} 孔鸣宇¹ 贺磊磊¹ 孔朔琳¹ 杨蜀秦^{1,2} 傅隆生^{1,2}

(1. 西北农林科技大学机械与电子工程学院, 陕西杨凌 712100; 2. 农业农村部农业物联网重点实验室, 陕西杨凌 712100)

摘要: 针对自然环境下草莓成熟度识别受枝叶遮挡及光照不均影响导致分类精度低的问题,本研究提出一种融合改进的 YOLO v8 实例分割与轻量化 CNN 分类网络的级联检测方法。该方法首先使用 GhostNet 轻量化结构和注意力机制改进的 YOLO v8n-seg 模型实现果实定位与分割,在模型参数量较低(2.6×10^6)的前提下,实现在复杂背景中对目标的准确分割。其次,利用分割掩码从原始图像中裁剪出果实目标区域,并输入至基于 EfficientNet-Lite 构建的 CNN 分类网络,该网络借助 MBConv (Mobile inverted bottleneck convolution) 模块和 Softmax 分类器,专注于捕捉果实表面细微的颜色与纹理特征变化,有效提升了模型对“转熟”等过渡阶段果实的辨识能力。试验结果表明,本研究提出的解耦式空间定位与属性识别协同策略显著增强了系统的鲁棒性,成熟度识别平均准确率达到 88.1%,整体系统参数量为 7.3×10^6 (EfficientNet-Lite 参数量为 4.7×10^6),推理速度为 71.9 f/s,在保证实时处理的同时实现了对未熟、转熟及成熟草莓的准确分级,为草莓智能采摘机器人的视觉系统提供了技术支撑。

关键词: 草莓; 成熟度检测; YOLO v8n-seg; EfficientNet-Lite; CNN; 实例分割

中图分类号: TP391.4; S24

文献标识码: A

文章编号: 1000-1298(2026)11-0354-10

OSID:



Strawberry Ripeness Cascade Detection Method Based on YOLO v8n-seg and Lightweight CNN

HONG Qi^{1,2} KONG Mingyu¹ HE Leilei¹ KONG Shuolin¹ YANG Shuqin^{1,2} FU Longsheng^{1,2}

(1. College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China

2. Key Laboratory of Agricultural Internet of Things, Ministry of Agriculture and Rural Affairs, Yangling, Shaanxi 712100, China)

Abstract: Aiming to address the problem of low classification accuracy in strawberry ripeness recognition caused by branch and leaf occlusion and uneven lighting under natural conditions, a cascade detection method that integrated an improved YOLO v8 instance segmentation with a lightweight CNN classification network was proposed. This method firstly used the GhostNet lightweight structure and an attention mechanism to improve the YOLO v8n-seg model for fruit localization and segmentation, achieving accurate target segmentation in complex backgrounds with a relatively low model parameter count (2.6×10^6). Next, fruit target areas were cropped from the original images by using the segmentation masks and input into a CNN classification network built on EfficientNet-Lite. This network, with mobile inverted bottleneck convolution (MBConv) modules and a Softmax classifier, focusing on capturing subtle color and texture feature changes on the fruit surface, effectively improved the model's ability to recognize fruits in transitional ripening stages. Experimental results showed that the proposed decoupled spatial localization and attribute recognition coordination strategy significantly enhanced system robustness. The average ripeness recognition accuracy reached 88.1%, the overall system parameter count was 7.3×10^6 (EfficientNet-Lite parameters accounted for 4.7×10^6), and the inference speed was 71.9 f/s, achieving accurate grading of unripe, turning, and ripe strawberries while ensuring real-time processing. The research result can provide technical support for the vision system of intelligent strawberry-picking robots.

Key words: strawberry; maturity detection; YOLO v8n-seg; EfficientNet-Lite; CNN; instance segmentation

收稿日期: 2026-01-28 修回日期: 2026-02-13

基金项目: 国家自然科学基金项目(32572201)

作者简介: 洪旗(1989—),男,副教授,博士,主要从事农业智能化技术与装备研究,E-mail: h407460979@163.com

通信作者: 傅隆生(1984—),男,教授,博士生导师,主要从事智慧农业技术与装备研究,E-mail: fulsh@nwafu.edu.cn

0 引言

草莓作为一种深受消费者喜爱且经济效益显著的高附加值作物,在全球农业生产中占据重要地位。在草莓等水果生产管理链条中,果实成熟度的精准判断是确定最佳采摘窗口、保障果实品质与提升商品价值的关键环节^[1]。然而,目前生产实践中主要依赖人工经验进行成熟度判别,这种方式不仅劳动强度大、效率较低,且成熟度分级结果易受主观因素的影响,难以适应现代农业规模化、精细化的需求。实际的田间场景呈现高度非结构化特征(光照条件、枝叶遮挡和果实姿态)导致传统依赖人工经验或简单颜色阈值的判别方法难以满足自动化作业的要求^[2]。针对上述问题,研究一种在复杂非结构化环境下兼具高精度与高鲁棒性的草莓成熟度检测方法,对于推动该产业智能化转型具有重要意义^[3]。

近年来,研究者从算法精度提升、轻量化设计及复杂环境适应性等多个维度对果实检测任务开展了相关研究。在目标检测方面,YOLO 模型因其兼顾检测准确性和实时性成为主流选择^[4-5]。为适应嵌入式设备的低算力问题,轻量化网络研究受到人们的广泛关注^[6-8]。随着 YOLO v8 的发布,其在果实成熟度分级与机器人视觉领域得到了大量的应用与研究^[9-10]。在实例分割领域,LI 等^[11]提出了一种非模态实例分割方法,通过改进 Mask R-CNN 头部结构并结合数据增强策略,实现了对被枝叶严重遮挡果实完整形态的精准预测,有效解决了遮挡导致的果实定位偏差问题。李丹等^[12]综述了多模态融合技术在果蔬识别中的应用,指出融合深度信息与彩色图像特征可显著提升分割精度。JIA 等^[13]针对果园中的青苹果检测难题,通过增强 Mask R-CNN 对颜色相近目标的边缘特征提取能力,显著降低了背景与果实颜色混淆带来的误检率。

尽管上述研究在单项任务上取得了显著进展,但针对草莓采摘机器人的“检测—分割—成熟度判别”一体化流程仍面临 3 大技术瓶颈:①背景噪声干扰问题突出,目前主流的检测方法输出包含枝叶、地膜等冗余信息,严重干扰了对“转熟期”果实微细颜色特征的判断,且传统基于颜色空间的方法在复杂光照下鲁棒性不足。②模型的精度与实时性需要进一步平衡,以 Mask R-CNN 为代表的两阶段分割算法虽然精度较高,但其庞大的参数量难以满足机器人实时作业需求;而过度轻量化的模型对果实边缘细节的提取能力不足,易引发属性识别的误差累积。③细粒度分类能力受限,现有的一阶段检测网络通常缺乏针对性的特征提取方式,难以精准捕捉

“未熟”与“转熟”之间模糊的颜色过渡特征,从而限制了系统的分级性能。

针对上述问题,本文提出一种融合改进 YOLO v8n-seg 实例分割与轻量化 CNN 分类的级联检测方法。该策略采用“定位-属性”解耦的设计思路:首先利用 GhostNet 与注意力机制对 YOLO v8n-seg 模型进行改进,在低算力平台上实现对果实的定位与背景剔除;随后,将感兴趣区域(Region of interest,ROI)输入至轻量级 EfficientNet-Lite 网络,通过深层特征挖掘实现对草莓未熟、转熟及成熟 3 个阶段的精准辨识。以期通过该级联架构,在复杂检测背景下提升系统鲁棒性的同时,有效平衡检测精度与推理效率,为农业机器人的实时采摘作业提供一种技术支撑。

1 材料与方法

1.1 材料

1.1.1 数据集构建

为确保模型可以识别不同的草莓品种以及进一步提高模型的泛化能力,本研究采用双源数据构建策略。鉴于公共数据集在场景多样性方面存在不足,缺乏遮挡、光照变化等复杂样本,除引入 Training Data Pro 公共数据集(<https://www.kaggle.com/datasets/trainingdatapro/ripe-strawberries-detection>)作为数据集 1 外,还在陕西省咸阳市杨陵区草莓种植基地(34°16'N,108°04'E)开展实地图像采集,构建数据集 2。图像采集设备选用搭载于机器人末端的高清摄像头(Raspberry Pi camera module v3),其分辨率为 4 608 像素×2 592 像素,存储格式为 JPG。本研究参考联合国欧洲经济委员会标准 FFV-35^[14]及中国农业行业标准 NY/T 2785—2015 关于草莓外观色泽的等级规定^[15],结合机器人田间作业的实际需求,划分了 3 类草莓成熟度:①成熟:果实表面着红面积在 90% 以上,呈现深红色,符合鲜食上市标准,为机器人采摘的首要目标。②转熟:果实表面着红面积在 20%~90% 之间,呈现粉红或红白相间特征。尽管该阶段果实并非当前采摘对象,但对其精准识别具有重要的农艺价值:一方面,转熟果实是未来 24~48 h 内的潜在收获产能,识别其数量有助于实现短期产量的精准预估;另一方面,转熟期果实组织已趋于柔软,将其识别为“软障碍物”可引导机械臂在规划路径时主动避让,防止因误触造成的机械损伤^[15]。③未熟:果实表面着红面积在 20% 以下,呈现绿色或白色。

基于上述标准,采集过程覆盖不同拍摄视角与光照条件,共获取 600 幅涵盖成熟、转熟与未熟 3 类

的草莓图像(图 1),统计结果如表 1 所示。最终构建的草莓数据集样本示例如图 2 所示。

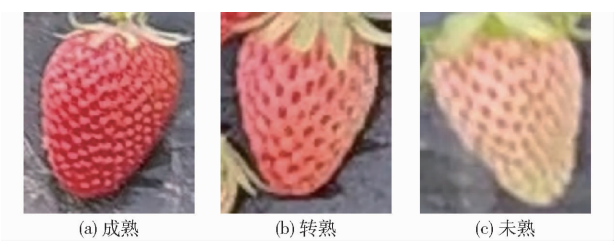


图 1 不同成熟度草莓种类示例

Fig. 1 Diagrams of strawberry varieties with different maturity levels

表 1 草莓数据集样本统计

Tab. 1 Detailed distribution of strawberry instance in dataset					
数据集来源	图像 总数	样本 总数	成熟 数量	转熟 数量	未熟 数量
公共数据集(数据集 1)	500	892	745	85	62
采集数据集(数据集 2)	600	2 002	1 086	495	421
合计	1 100	2 894	1 831	580	483
占比/%			63. 3	20. 0	16. 7

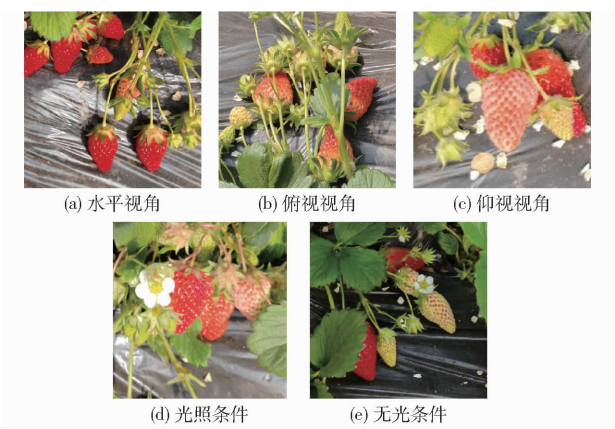


图 2 不同拍摄视角与光照条件数据集样本示例

Fig. 2 Schematics of data under different shooting angles and lighting conditions

数据集中草莓果实均采用 Labelme v5. 6. 0 版本进行标注^[16],按照比例 8: 2 划分为训练集和验证集。

1. 1. 2 试验场景与硬件环境配置

(1) 试验场景构建与数据采集

为验证算法在非结构化农业环境下的鲁棒性,本研究在陕西省咸阳市杨陵区草莓种植基地构建了模拟采摘试验场景。如图 3 所示,试验采用履带式智能巡检机器人作为移动载体,搭载高清摄像模组,安装位置模拟机械臂末端执行器视角,保持主要作业距离在 20 ~ 50 cm。

试验涵盖晨间、正午及傍晚等不同时段,通过多时序采集引入强光直射、阴影变化及枝叶重叠等复

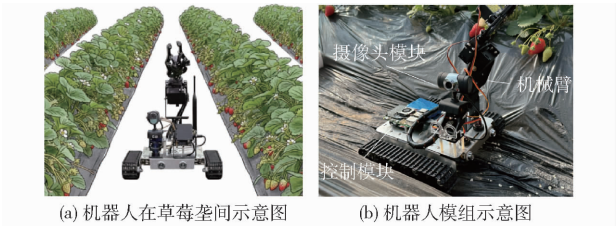


图 3 实地试验场景

Fig. 3 Field experiment scenarios

杂环境干扰,以充分测试改进 YOLO v8n-seg 模型的分割性能。在试验过程中,机器人沿垄间匀速巡航,系统对实时视频流执行“检测—分割—识别”级联处理:首先利用分割掩码精准提取果实 ROI 区域,随后驱动 EfficientNet-Lite 网络完成成熟度判别。该场景设计实现了空间定位与属性辨识的解耦,为模拟真实的机器人采摘作业提供了高保真的数据支撑与试验基础。

(2) 软硬件环境配置

试验计算平台配置如下:处理器为 Intel Core i5-14600KF,显卡为 NVIDIA GeForce RTX 5060 Ti(16 GB 显存),内存 32 GB。软件环境基于 Windows 11 操作系统,深度学习框架采用 PyTorch 2. 0 配合 CUDA 11. 8 加速,目标检测与分割算法基于 Ultralytics YOLO v8 框架部署,Python 版本为 3. 9。

1. 2 草莓果实分割算法选择与改进

为在复杂田间环境下实现草莓果实的精准分割,算法选型需在检测精度与推理效率之间寻求最佳平衡。首先对当前主流实例分割算法进行了对比分析。

以 Mask R-CNN 为代表的两阶段算法虽然边缘分割精度较高,但其庞大的参数量与复杂的计算流程导致推理延迟高^[17],难以满足农业采摘机器人对实时性的需求。基于 Transformer 架构的 RT-DETR-seg 虽然在全局特征捕获方面具有优势^[18],但其计算复杂度高,在算力有限的嵌入式设备上部署较为困难。

相比之下,YOLO v8 系列模型凭借 Anchor-free 检测头与 C2f 模块设计,在保持高精度的同时具备推理速度与轻量化优势。预试验结果表明,原始 YOLO v8n-seg 推理速度虽满足要求,但在枝叶密集遮挡场景下的分割细节仍显不足。鉴于此,本研究选取 YOLO v8n-seg 为基准模型,针对田间作业环境进行结构优化,构建了一种改进型果实分割模型。

具体改进策略包括 3 个方面:①轻量化。引入 GhostNet^[19]重构主干网络,利用 Ghost 模块的轻量级运算特性,在保证模型特征表达能力的同时降低

参数量。② 注意力增强。在 C2f 模块中嵌入 CBAM^[20]与 CoordAtt^[21]混合注意力机制,强化模型对果实空间位置与关键通道信息的感知能力。③ 损失函数优化。引入 Focal Loss 与 Dice Loss 组合损失函数^[22],重点解决难分样本挖掘及边缘模糊问题。

改进后的 YOLO v8n-seg 模型结构如图 4 所示,图中,GhostConv 表示基于 GhostNet 的轻量化卷积模块,用于降低参数量与计算开销;CBAM 和 CoordAtt 表示卷积块注意力模块,用于增强模型对目标区域的特征提取能力;Focal Loss 与 Dice Loss 作为分割分支的组合损失函数,有助于提升小目标检测与边界分割的精度。其余模块含义与原始 YOLO v8 保持一致,包括 Conv(卷积)、C2f(特征融合模块)、SPPF(快速空间金字塔池化)、Concat(特征拼接)、Upsample(上采样)以及 Detect(检测头)。

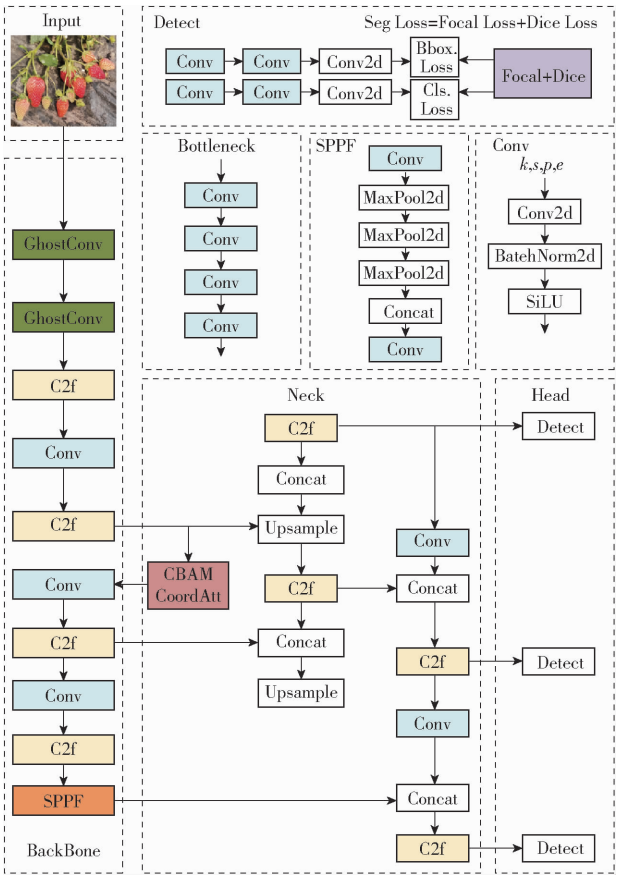


图 4 改进的 YOLO v8n-seg 模型结构示意图
Fig.4 Schematic of improved YOLO v8n-seg model structure

1.3 CNN 成熟度识别模型设计

为了实现草莓果实的自动化成熟度识别,本文在 YOLO v8-seg 分割模型输出的基础上构建了一个轻量化 CNN 分类器,用于判断草莓果实的成熟阶段。在实际应用中,分类器需要同时具备高精度、低计算量与强泛化能力^[23]。为此,本文选用 EfficientNet-Lite 作为主干网络结构,对草莓 ROI 进行特征提取与分类。EfficientNet-Lite 是由 Google

提出的高效轻量化网络,其通过复合缩放策略与高效卷积结构实现了精度与速度的平衡,特别适合部署于移动端与实际农业场景中^[24]。

以下将从 CNN 原理、EfficientNet-Lite 主干结构以及分类机制 3 个方面进行详细阐述。

CNN 是一类具有局部连接、参数共享与分层表达能力的神经网络。其核心思想是通过多层卷积与下采样逐级提取图像的局部纹理、边缘、颜色与高层语义特征。CNN 的基本组成包括卷积层、激活函数、批归一化层、池化层、全连接层与 Softmax 分类器。

(1) 卷积层

卷积层是 CNN 的核心结构,其主要功能是从局部区域中提取边缘、纹理和颜色等低级特征。卷积操作可表示为

$$Y(i,j,k') = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \sum_{c=1}^C X(i+m,j+n,c) \cdot W(m,n,c,k') + b_k \quad (1)$$

式中 X ——输入特征图
 W ——卷积核 b_k ——偏置项
 C ——输入通道数
 k' ——卷积核数量
 c ——输入特征图的通道索引(遍历变量)
 M,N ——卷积核的高度和宽度
 m,n ——卷积核在高度和宽度方向上的偏移量

卷积层通过局部连接与参数共享机制显著降低了网络参数量,使其适用于嵌入式农业场景。

(2) 批归一化

批归一化层用于缓解深层网络中的梯度消失与梯度爆炸问题,同时提升模型的训练稳定性和泛化能力。其标准化公式为

$$y = \gamma \hat{x} + \beta \quad (2)$$

其中
$$\hat{x} = \frac{x - \mu}{\sqrt{\sigma^2 + \varepsilon}} \quad (3)$$

式中 $\hat{x}$ ——归一化后的特征值
 μ, σ ——mini-batch 的均值和标准差
 γ, β ——可学习参数
 ε ——防止分母为零引入的数值稳定常量
 x ——批归一化层的输入特征值
 y ——经过归一化及线性缩放、平移后的输出特征值

BN 在光照变化较大的自然环境图像处理中特别有效。

(3) 激活函数

为增强网络的非线性表达能力,模型采用 ReLU

或 ReLU6 激活函数,其定义为

$$\text{ReLU}(x) = \begin{cases} x & (x > 0) \\ 0 & (x \leq 0) \end{cases} \quad (4)$$

激活函数能够有效提升模型对复杂颜色差异(例如草莓由绿色逐渐转为红色)的表征能力,是成熟度分类任务中不可或缺的组成部分。

(4) 池化层

池化层用于降低特征图分辨率、减少参数量并增强特征鲁棒性,其计算式为

$$y(i,j) = \max_{(m,n) \in \Omega} X(i+m,j+n) \quad (5)$$

式中 $y(i,j)$ ——池化操作后输出特征图在 (i,j) 位置的值

Ω ——池化核的作用区域

池化层有助于过滤噪声与局部光照反射对分类的干扰,使网络更关注于稳定的颜色与纹理特征。

(5) EfficientNet - Lite 主干与 MBConv 模块结构

为在保证模型轻量化的同时提升特征提取能力,本文将传统卷积堆叠结构替换为 EfficientNet - Lite 主干网络。该网络以 MBConv (Mobile inverted bottleneck convolution) 为核心^[25],其通过倒残差结构、深度可分离卷积与通道注意力机制协同作用,实现高效的特征表达,其基本结构如图 5 所示,图中 C_{in} 表示输入通道数, C_{out} 表示输出通道数, H 、 W 分别表示特征图的高度和宽度。

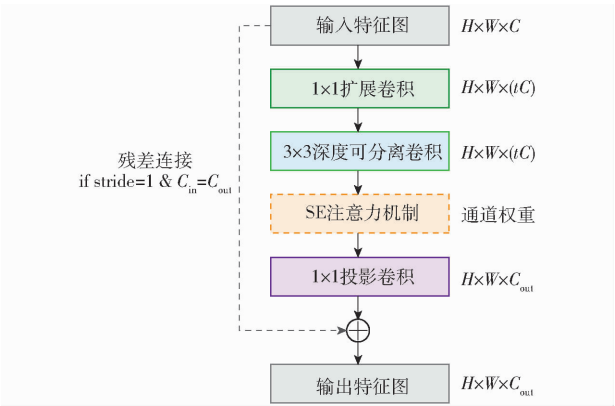


图 5 MBConv 模块结构
Fig. 5 MBConv module structure

EfficientNet - Lite 的核心模块 MBConv 包含：
①扩展卷积,将输入通道升维,增强特征表达。
②深度可分离卷积,逐通道卷积,极大减少运算量。
③逐点卷积,恢复通道数,实现特征融合。
④SE 注意力机制 (Squeeze-and-excitation, SE) 强化颜色与光照敏感区域。

这一结构使得 EfficientNet - Lite 在检测草莓成熟度时具备更高的颜色判别力。

在 MBConv 模块中,首先通过扩展卷积将输入通道数按系数 t 进行升维,从而在轻量计算的前提

下增强特征空间的表达能力。扩展过程可表示为 $C_{\text{expand}} = tC_{\text{in}}$,其中 C_{expand} 表示扩展后的特征通道数。随后,采用深度可分离卷积对各通道分别进行卷积操作,即对第 k 个通道独立执行 $Y_k = X_k W_k$,其中 Y_k 表示第 k 个通道卷积处理后的输出特征图, X_k 表示第 k 个通道的输入特征图, W_k 表示独立作用于第 k 个通道的深度卷积核,极大降低了计算量,使网络在保障实时性的同时保持较高分类精度。接着,引入 SE 注意力机制^[26] 对特征进行自适应重标定,其基本流程如图 6 所示。

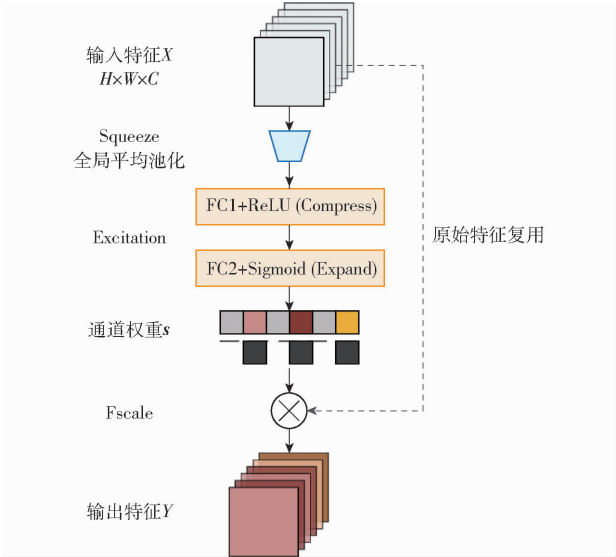


图 6 SE 注意力机制特征标定流程
Fig. 6 SE attention mechanism feature calibration process

该机制首先通过全局平均池化进行通道压缩

$$Z_c = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W X_c(i,j) \quad (6)$$

式中 $X_c(i,j)$ ——输入特征图第 c 个通道在空间位置 (i,j) 上的像素值

随后通过激励操作获得通道权重

$$s = \text{Sigmoid}(W_2 \delta(W_1 z)) \quad (7)$$

式中 s ——最终生成的通道权重向量

δ ——ReLU 激活函数

W_1 、 W_2 ——第 1、2 个全连接层的权重矩阵

z ——上一步得到的全局描述符向量

并将其作用于原始特征图,实现

$$s = s_c X_c \quad (8)$$

式中 s_c ——激励操作得到的第 c 个通道权重标量

从而增强与成熟度变化相关的关键颜色与纹理信息。最后,模块利用 1×1 的逐点卷积实现通道融合与降维,使输出恢复到与模块输入一致的维度。在步长为 1 且输入输出通道一致的情况下,MBConv 引入残差连接,以保持梯度流动并提升稳定性,整体可表达为

$$Y = X + F(X) \quad (9)$$

式中 Y ——MBConv 模块的最终输出特征图
 X ——MBConv 模块的初始输入特征图
 F ——包含扩展卷积、深度可分离卷积、SE 注意力机制和逐点投影卷积在内的一系列主干操作

上述结构的组合使 EfficientNet-Lite 在模型参数量较低条件下依然具备优秀的特征表达能力,特别适用于草莓成熟度识别过程中对细微颜色差异的建模。

(6) 全局平均池化 (Global average pooling, GAP)

GAP 将二维特征图压缩为单一通道特征,计算式为

$$g_c = \frac{1}{HW} \sum_{i,j} X_c(i,j)$$
 (10)

式中 g_c ——第 c 个通道经过全局平均池化后得到的输出标量值

GAP 不仅减少了参数量,还能够有效保留整体颜色趋势,有助于区分草莓不同成熟阶段的色彩分布。

(7) 全连接层与 Softmax 分类器

全连接层执行线性变换

$$z_{fc} = W_i g + b$$
 (11)

式中 W_i ——全连接层的权重矩阵
 g ——经过全局平均池化处理后展平的一维特征向量
 z_{fc} ——全连接层输出
 b ——偏置项向量

Softmax 将输出映射为 3 类成熟度的概率分布为

$$P_k = \frac{e^{z_k}}{\sum_{i=1}^3 e^{z_i}}$$
 (12)

式中 P_k ——模型预测该草莓属于第 k 个成熟度类别的概率
 z_i, z_k ——全连接层输出向量 z 中,对应第 i 和第 k 个类别的得分

最终,模型输出未熟、转熟与成熟 3 个类别的概

率值,用于草莓成熟度判定。

1.4 YOLO 与 CNN 融合检测策略

本文在 YOLO v8n-seg 的基础上,结合轻量级 CNN 实现草莓果实的成熟度检测。整体思路是:首先利用 YOLO v8-seg 模型检测并分割出图像中的草莓区域;然后,将每个分割出的 ROI 提取出来,输入到训练好的 CNN 分类模型中,判断果实的成熟状态,从而实现检测与分类相结合的自动化分析流程。

系统的主要工作流程为:①YOLO 分割检测。通过 YOLO v8n-seg 模型识别图像中的草莓,并输出每个目标的掩码信息。②ROI 提取。利用分割掩码从原图中裁剪出每个草莓的区域图像。③CNN 分类。将提取的草莓区域输入到轻量级 CNN 网络中进行分类,输出成熟度类别。④结果显示与统计。在原图上为每个目标叠加成熟度标签,并统计各类果实数量与比例。

2 结果与分析

2.1 不同分割模型性能对比试验

为验证本文改进算法的先进性,在相同草莓数据集与试验环境下,将本文提出的改进 YOLO v8n-seg 模型与当前主流的实例分割模型进行对比。对比模型包括:经典两阶段算法 Mask R-CNN、Transformer 架构算法 RT-DETR-seg,以及单阶段算法 YOLO v5n-seg 和 YOLO v7-tiny-seg^[3]。

试验结果如表 2 所示。从表 2 可以看出,Mask R-CNN 凭借强大的 Transformer 主干,mAP_{0.5}达到了最高值 87.5%,但其参数量高达 8.6×10^7 ,推理速度仅为 18.0 f/s,无法满足实时检测需求。RT-DETR-seg 虽然精度较高,但计算量依然较大。相比之下,本改进模型在引入注意力机制与轻量化模块后,mAP_{0.5}达到 88.9%,与 Mask R-CNN 差距极小(仅高 1.4 个百分点),但参数量仅为 2.6×10^6 ,不足其 1/30,推理速度高达 113.0 f/s。这表明本文方法在大幅降低计算负载的同时,有效保持了分割精度,实现了速度与精度的最佳平衡,对于草莓自动采摘系统具有较高应用潜力^[27]。

表 2 不同主流分割模型在草莓数据集上的性能对比

Tab.2 Performance comparison of different mainstream segmentation models					
模型方法	主干网络	mAP _{0.5} /%	mAP _{0.5:0.95} /%	参数量	推理速度/(f·s ⁻¹)
Mask R-CNN	Swin-T	87.5	63.2	8.6×10^7	18.0
RT-DETR-seg	HGNetv2-L	87.1	63.0	3.25×10^7	32.0
YOLO v5n-seg	CSPDarknet	83.5	58.2	2.1×10^6	95.0
YOLO v7-tiny-seg	ELAN	85.2	60.5	6.8×10^6	67.0
YOLO v8n-seg (基线)	CSPDarknet	84.5	59.5	3.4×10^6	115.0
YOLO v8n-seg (改进)	GhostNet	88.9	63.9	2.6×10^6	113.0

2.2 CNN 成熟度分类结果分析

基于检测分割得到的 ROI, 本文采用 EfficientNet - Lite0 进行成熟度识别。该模型以 MBConv 为核心构建深层网络, 利用复合缩放策略平衡网络深度、宽度与分辨率。具体流程为: 首先将 ROI 图像调整为 224 像素 × 224 像素, 经 Stem 层与 7 个 MBConv 阶段提取高层语义特征; 随后通过 GAP 将特征压缩为一维向量, 最后经 Dropout 层与全连接层, 由 Softmax 函数输出 3 类成熟度的概率分布^[28-29]。

为验证本研究选用 EfficientNet - Lite0 作为成熟度分类器的合理性, 在相同草莓 ROI 数据集下, 将其与经典的残差网络 ResNet18 以及轻量化网络 MobileNetV3 - Small 进行了基线对比试验。

试验结果表明, ResNet18 虽然特征提取能力强, 分类准确率最高 (88.5%), 但其参数量高达 1.17×10^7 , 计算量较大, 不利于后续的嵌入式部署; MobileNetV3 - Small 虽然参数量仅为 1.5×10^6 且推理速度最快, 但由于网络深度与宽度受限, 对转熟期模糊特征的代表能力不足, 准确率仅为 85.2%。相比之下, 本文选用的 EfficientNet - Lite0 在保持 88.1% 准确率的同时, 参数量控制在 4.7×10^6 , 取得了精度与计算成本的最佳平衡, 证明了其作为级联分类基线模型的优越性。

CNN 模型训练过程如图 7 所示, 训练集的损失值 (Loss) 随迭代轮数逐渐下降, 并在第 25 轮左右趋于稳定, 验证集准确率 (Accuracy) 持续上升, 最终收敛于 88.1%。

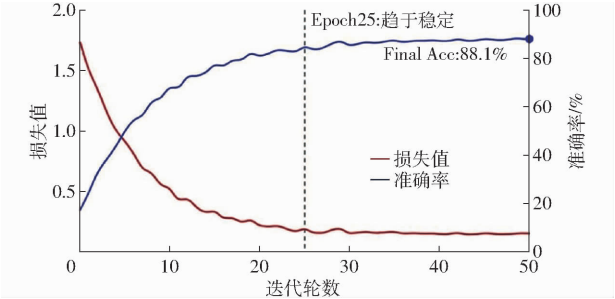


图 7 CNN 模型训练过程曲线
Fig. 7 Training curves of CNN model

为进一步验证模型分类性能, 本文在测试集上绘制了混淆矩阵, 结果如图 8 所示。模型在未熟与成熟类别上表现优异, 准确率分别达到 95.0% 和 91.0%; 转熟类样本由于颜色过渡模糊, 存在少量误判。总体分类准确率达 88.1%, F1 分数为 88.0%。

同时, 模型的详细性能指标如表 3 所示, 模型在未熟、转熟、成熟 3 类中的精确率分别为 85.7%、81.8%、100%, 召回率分别为 75.0%、100%、90.9%, 表明模型在成熟与未熟样本的识别上表现

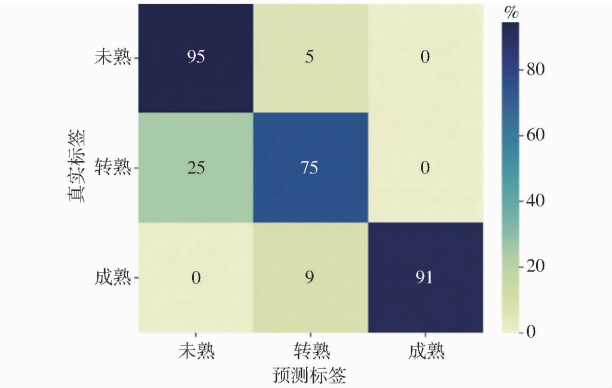


图 8 混淆矩阵结果图

Fig. 8 Confusion matrix result

表 3 CNN 模型各类别性能指标

Tab. 3 Performance metrics of CNN model by category

类别	精确率	召回率	F1 分数	mAP _{0.5:0.95}
未熟	85.7	75.0	80.0	60.1
转熟	83.3	100	90.9	62.0
成熟	100	90.9	95.2	63.4
平均值	89.7	88.6	88.7	63.9

最为稳定。模型对中间阶段的识别略有偏差, 主要原因在于样本间过渡模糊及光照差异较大。

2.3 模型可视化与特征解释

为进一步探究轻量化 CNN 模型在草莓成熟度判定中的内部决策机制, 验证模型是否准确学习到了果实表面的关键特征而非背景噪声, 本研究采用 Grad - CAM 技术对模型进行可视化分析。Grad - CAM 通过计算最后一层卷积层特征图对分类得分的梯度, 生成类激活热力图, 从而直观展示网络在分类时关注的图像区域。

可视化结果如图 9 所示, 图中选取了不同成熟度阶段的典型样本, 分别展示了原始图像、特征掩码图及叠加热力图。图中颜色越趋向深红色的区域, 代表模型在该位置的特征响应越强, 即该区域对模型做出分类决策的贡献越大。

模型在不同成熟度阶段表现出显著的响应差异, 体现了其对细粒度特征的敏锐捕捉能力。在成熟阶段, 模型的热力值高度集中于果实表面呈现深红色的区域, 表明“全红”的颜色语义是判定成熟样本的关键特征; 在转熟阶段, 针对颜色过渡最为复杂的特征, 热力图的高亮区域精准分布在果实由白转红的渐变地带, 这证明模型能够有效识别“红白相间”的混合颜色特征, 并将其作为区分转熟果实的重要标志; 在未熟阶段, 模型的关注点均匀分布在泛白或淡绿色的果面上。值得注意的是, 模型准确避开了果萼的干扰, 证明其具备区分果实组织与营养器官的能力。

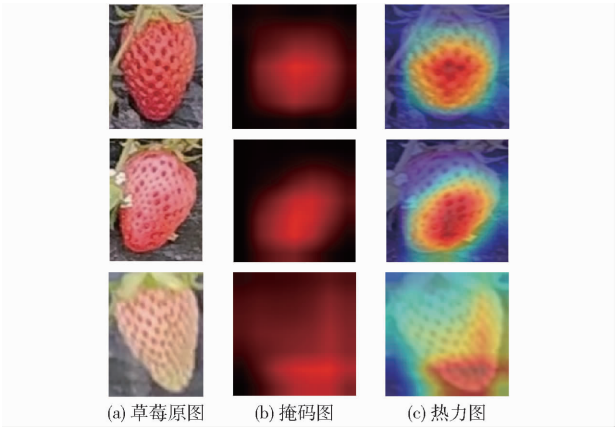


图 9 基于 Grad-CAM 的草莓成熟度特征可视化

Fig.9 Feature visualization of strawberry maturity based on Grad-CAM

可视化分析进一步验证了“分割引导—分类”级联策略的有效性。对比原图与热力图可以发现,由上一级 YOLO v8n-seg 模块裁剪产生的黑色背景区域在热力图中始终表现为冷色调,意味着背景信息对最终分类结果几乎无贡献。这种现象表明,通过前期分割剔除枝叶、地膜等复杂背景干扰,迫使分类网络专注于纯净的果实 ROI 区域,从根源上消除了背景噪声对细粒度分类的影响。这种解耦设计不仅提升了模型的分类精度,更显著增强了系统的可解释性与可靠性。

2.4 级联系统综合性能评价与对比

为全面评估“改进 YOLO v8n-seg + EfficientNet-Lite”级联系统的实际应用价值,本研究从系统实时性与综合识别准确率 2 个维度进行了量化评价,并与当前主流的单阶段检测 (YOLO v8n-cls)、两阶段分割 (Mask R-CNN) 及传统图像处理方法 (HSV 颜色阈值) 进行了对比^[8]。

级联模型的端到端推理耗时主要由实例分割推理、ROI 裁剪与预处理,以及成熟度分类推理 3 个阶段组成。为了验证系统的实时性,本研究在相同硬件平台 (RTX 5060 Ti) 上对单帧图像的处理流程进行了耗时分解测试。

测试结果表明,系统处理单帧图像的总耗时约为 13.9 ms。其中,改进 YOLO v8n-seg 模型完成果实定位与分割 (含 NMS 后处理) 的平均耗时为 8.9 ms,占总耗时的 64.0%;CPU 执行 ROI 坐标映射与图像裁剪耗时约 1.5 ms;EfficientNet-Lite 分类网络对裁剪后的果实区域进行推理仅需 3.5 ms。虽然引入分类网络增加了少量计算,但得益于其轻量化设计,系统整体运行速度仍满足农业机器人实时作业要求,能够流畅应对动态采摘场景。

为了验证级联架构的优势,将本文方法与“YOLO v8n 直接分类”及“Mask R-CNN + ResNet”

等方法进行对比。定义级联系统准确率为:仅当目标被正确分割 (IoU 大于 0.5) 且成熟度分类正确时,判定为识别成功。

由表 4 试验结果表明,相比于直接使用 YOLO v8n 进行端到端检测,本文的级联策略虽然速度略有下降,但综合准确率达 88.1%,提升了 8.1%。这是因为“先分割、后分类”的解耦策略通过掩码有效剔除枝叶、地膜等复杂的田间背景噪声,使分类网络能专注于果实本身的细微纹理特征。与高精度的 Mask R-CNN 相比,本文方法在准确率更高的情况下,推理速度提升了 4 倍,达到 71.9 f/s;参数量为 7.3×10^6 ,减少了 91%,实现了精度与速度的最佳平衡。

表 4 不同草莓成熟度检测策略的综合性能对比

Tab.4 Comprehensive performance comparison of different strawberry ripeness detection strategies

方法	策略	综合准确率/%	参数量	推理速度/ (f·s ⁻¹)	抗干扰能力
HSV 阈值法	颜色分割	81.2		120.0	差
Mask R-CNN	分割 + 分类	87.5	8.6×10^7	18.0	中
	一体化				
YOLO v8n	检测 + 分类	81.5	3.2×10^6	110.0	中
本文方法	改进分割 + 轻量分类	88.1	7.3×10^6	71.9	优

此外,针对自然环境下常见的“枝叶遮挡”与“光照不均”难点问题,本研究进行了专项抗干扰性能验证。从测试集中筛选出具有典型强遮挡 (遮挡面积大于 30%) 与复杂光照 (强逆光或阴影) 特征的样本各 100 幅进行测试。结果显示:

(1) 在强遮挡场景下:得益于引入的 CBAM 注意力机制对可见区域关键特征的聚焦能力,模型仍保持了 85.3% 的识别准确率,相比未改进的基准模型提升了 6.8%。

(2) 在复杂光照条件下:级联架构中的 EfficientNet-Lite 通过色彩增强与归一化处理,有效缓解了阴影对颜色判断的干扰,准确率保持在 86.5%。

上述数据表明,本文提出的改进策略不仅在常规场景下保持一定的识别精度,在非结构化复杂环境下同样具备较强的鲁棒性,有助于解决传统方法因环境干扰导致分类精度下降的问题。

2.5 嵌入式部署与实时性验证

为了验证本文提出的级联检测算法在低成本农业设备上的可行性,本研究构建了基于树莓派 5 的嵌入式视觉检测平台。

平台搭建与实测效果如图 10 所示。其中,图 10a 为硬件试验环境,系统主要由树莓派 5 计算核心、散热模块、便携式显示屏组成,配置 Python 3.11 运行环境与 OpenCV 库,利用 ONNX Runtime 引擎对导出的模型进行加速推理;图 10b 为系统实际运行界面与检测结果。

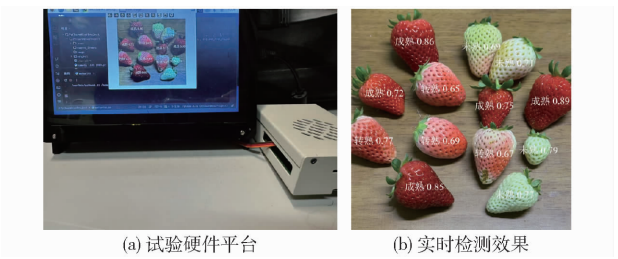


图 10 树莓派 5 嵌入式部署环境与检测结果

Fig. 10 Embedded deployment environment and detection results on Raspberry Pi 5

测试结果表明,得益于改进算法的轻量化设计与树莓派 5 的算力支持,系统运行流畅,能够实时输出检测结果。从图 10b 可见,系统将所有草莓成熟度全部准确检测出来,检测置信度在 0.75 左右,最高置信度达 0.89,模型内存占用量 760 MB 左右。算法具有极高的鲁棒性,能够精准识别视野内不同成熟度的草莓。

该试验证明了本文提出的改进 YOLO v8n-seg + EfficientNet-Lite 级联架构在低功耗嵌入式设备上具备良好的实时性与泛化能力,能够满足智能农业装备的实际应用需求。

3 结论

(1) 针对自然环境下枝叶遮挡及光照不均导致草莓成熟度识别精度低、鲁棒性不足的问题,本研究

提出一种融合改进的 YOLO v8n-seg 实例分割与 EfficientNet-Lite 分类网络的级联检测方法。通过解耦空间定位与属性辨识 2 个任务,实现了复杂非结构化场景下的高精度实时检测。

(2) 构建了轻量化且高精度的草莓实例分割模型。针对 YOLO v8n-seg 进行改进,通过引入 GhostNet 轻量化主干网络显著降低了计算冗余;嵌入 CBAM 与 CoordAtt 混合注意力机制,增强了模型在复杂背景下对果实关键特征的提取能力;采用 Focal 和 Dice 联合损失函数优化了难分样本与掩码边缘的分割质量。试验表明,改进模型参数量为 2.6×10^6 ,推理速度达 113.0 f/s, mAP_{0.5} 提升至 88.9% (较基准模型提高 4.4 个百分点),实现了检测精度与推理速度的最佳平衡。

(3) 实现了基于 ROI 的成熟度精细化识别策略。采用“检测引导—ROI 提取—精细分类”的级联架构,利用分割掩码精准裁剪果实 ROI 并输入 EfficientNet-Lite0 网络。借助 MBConv 模块的高效特征提取能力与迁移学习策略,有效克服了传统单阶段检测器在转熟等颜色过渡阶段识别能力受限的问题。测试结果显示,成熟度识别综合准确率达到 88.1%,验证了级联架构在细粒度分类任务中的有效性。系统整体参数量仅约为 7.3×10^6 ,推理速度为 71.9 f/s,实现了算法精度与计算负载的有效平衡。

(4) 开发了可视化的智能检测系统。基于 PyQt5 和树莓派搭建了草莓成熟度检测软件平台,集成果实定位、掩码分割及成熟度判定功能,实现了检测结果的实时可视化输出。系统运行稳健,交互性强,可为农业采摘机器人的视觉伺服与果园产量智能预估提供可靠的技术支撑。

参 考 文 献

[1] 张强,孙涛,刘志刚. 基于深度学习的草莓分级与质量评价方法研究[J]. 农业机械学报, 2023, 54(2): 137-145.
ZHANG Qiang, SUN Tao, LIU Zhigang. Research on strawberry grading and quality evaluation method based on deep learning [J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(2): 137-145. (in Chinese)

[2] 张莉,李伟,黄涛. 基于深度学习的果蔬目标检测研究进展[J]. 农业机械学报, 2022, 53(6): 1-12.
ZHANG Li, LI Wei, HUANG Tao. Research progress on fruit and vegetable object detection based on deep learning [J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(6): 1-12. (in Chinese)

[3] GONG L, YU M, JIANG S, et al. A review on image-based automatic detection of fruit diseases and pests[J]. Computers and Electronics in Agriculture, 2023, 209: 107833.

[4] 傅隆生,家博,刘小娟,等. 基于文献计量学的农业工程技术发展动态分析[J]. 农业机械学报, 2025, 56(10): 249-276.
FU Longsheng, JIA Bo, LIU Xiaojuan, et al. Development trends of agricultural engineering technology based on bibliometrics [J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(10): 249-276. (in Chinese)

[5] SU Y, XIE Z, WANG Y, et al. Real-time detection of tomato at different stages of maturity in complex greenhouse environments using improved YOLOv8[J]. Computers and Electronics in Agriculture, 2024, 222: 109012.

[6] LI H, LI J, WEI H, et al. A lightweight improved YOLOv5s model and its deployment for detecting pitaya fruits in daytime and nighttime light-supplement environments[J]. Computers and Electronics in Agriculture, 2024, 220: 108914.

[7] LIU Y, ZHANG H, CHEN Z, et al. Real-time apple detection in orchards using improved YOLOv7 with lightweight backbone [J]. Computers and Electronics in Agriculture, 2023, 210: 107997.

[8] REN Z, WANG K, ZHOU J. Lightweight YOLOv8 with transformer attention for cucumber detection under greenhouse environment[J]. Frontiers in Plant Science, 2024, 15: 1423567.

[9] 田有文, 秦尚升, 闫玉波, 等. 基于改进 YOLOv8 的田间复杂环境下蓝莓成熟度检测[J]. 农业机械学报, 2024, 55(12): 128 – 137.

TIAN Youwen, QIN Shangsheng, YAN Yubo, et al. Blueberry maturity detection under complex field environment based on improved YOLOv8 [J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55 (12): 128 – 137. (in Chinese)

[10] AGUIAR A S, MAGALHÃES S A, DOS SANTOS F N, et al. Grape bunch detection at different growth stages using YOLOv8 for harvesting robots[J]. Computers and Electronics in Agriculture, 2024, 219: 108780.

[11] LI G, VOUGIOUKAS S G. Improved amodal fruit segmentation via data augmentation and Mask R – CNN head modification [J]. Computers and Electronics in Agriculture, 2022, 197: 106951.

[12] 李丹, 郑伟, 韩雪. 基于多模态融合的果蔬智能识别与分级技术研究进展[J]. 农业工程学报, 2024, 40(7): 1 – 13.

LI Dan, ZHENG Wei, HAN Xue. Research progress on multimodal fusion for intelligent fruit and vegetable recognition and grading[J]. Transactions of the CSAE, 2024, 40(7): 1 – 13. (in Chinese)

[13] JIA W, ZHANG Y, ZHENG Y, et al. Real-time instance segmentation of green apples in complex orchards based on the improved Mask R – CNN[J]. Computers and Electronics in Agriculture, 2024, 221: 108960.

[14] UNECE. Standard FFV – 35 concerning the marketing and commercial quality control of strawberries[S]. New York and Geneva: United Nations, 2023.

[15] 中华人民共和国农业部. NY/T 2785—2015 草莓等级规格[S]. 北京: 中国农业出版社, 2015.

[16] 孙磊, 高鹏, 王志伟. 基于轻量化 YOLO 模型的温室黄瓜检测研究[J]. 农业装备与机械化, 2024, 45(3): 66 – 74.

SUN Lei, GAO Peng, WANG Zhiwei. Greenhouse cucumber detection based on lightweight YOLO model[J]. Agricultural Equipment and Mechanization, 2024, 45(3): 66 – 74. (in Chinese)

[17] MILLETARI F, NAVAB N, AHMADI S A. V – Net: fully convolutional neural networks for volumetric medical image segmentation[C]//3D Vision (3DV), 2016: 565 – 571.

[18] 陈亮, 何鹏, 张磊. 基于深度学习的草莓病害识别与检测[J]. 农业科学学报, 2023, 42(9): 130 – 138.

CHEN Liang, HE Peng, ZHANG Lei. Strawberry disease identification and detection based on deep learning[J]. Journal of Agricultural Science, 2023, 42(9): 130 – 138. (in Chinese)

[19] HAN K, WANG Y, TIAN Q, et al. GhostNet: more features from cheap operations[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020: 1580 – 1589.

[20] WOO S, PARK J, LEE J Y, et al. CBAM:convolutional block attention module[C]//Computer Vision-ECCV, 2018: 3 – 19.

[21] WANG L, XIAO J, PENG X, et al. Mango inflorescence detection based on improved YOLOv8 and UAV – RGB images[J]. Forests, 2025, 16(6): 896.

[22] WANG C Y, YEH I H, LIAO H Y M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023: 7464 – 7475.

[23] FU L, GAO F, WU J, et al. Application of consumer RGB – D cameras for fruit detection and localization in field: a critical review[J]. Computers and Electronics in Agriculture, 2020, 177, 105687.

[24] VERMA P R, PANTOLA D, SINGH N P. Plant disease detection with transfer learning: evaluating the impact of various loss functions and explainable AI[J]. Journal of Agricultural, Biological and Environmental Statistics, 2025,30: 1 – 23.

[25] 梁熙峰, 魏志伟. 基于改进 CycleGAN 与 YOLOv8 的夜间番茄茎、枝分割方法[J]. 农业机械学报, 2025, 56(3): 145 – 154.

LIANG Xifeng, WEI Zhiwei. Nighttime tomato stem and branch segmentation method based on improved CycleGAN and YOLOv8[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(3): 145 – 154. (in Chinese)

[26] HUANG Y, MA X, LIU R, et al. Multi-task learning with YOLOv8 for fruit detection and size estimation[J]. Sensors, 2025, 25(4): 3562.

[27] JIANG H, ZHAO J, MA F, et al. Mobile – YOLO: a lightweight object detection algorithm for aquatic organisms[J]. Fishes, 2025, 10(7): 348.

[28] GENÉ-MOLA J, SANZ-CORTÉS M, MARTIN-GORRIZ B, et al. Fruit detection, segmentation and 3D localisation using the MobileNetv3 – YOLOv4 architecture[J]. Precision Agriculture, 2023, 24(1): 278 – 299.

[29] LU Y, YOUNG S, HU X, et al. Practical assessment of deep learning-based detection for apple harvesting robots[J]. Computers and Electronics in Agriculture, 2023, 211: 108037.