doi:10.6041/j.issn.1000-1298.2025.09.052

基于改进 YOLO v11n 的轻量化葡萄穗梗采摘点定位方法与试验

庄　煜　徐琨林　李美琪　戴济泽　王金峰
（东北农业大学工程学院，哈尔滨 150030）

摘要：针对葡萄穗梗形态纤细且与环境色差微弱导致的采摘点定位问题，本文提出了一种基于改进 YOLO v11n 的轻量化分割模型 YOLO v11n－MBCa。构建了复杂环境下绿色葡萄高质量数据集，并引入轻量化 MobileNetV4 网络替换 YOLO v11n 的骨干网络（Backbone），减少模型参数量以及在特征融合层（Neck）引入加权双向特征金字塔网络（BiFPN），增强多尺度特征融合能力。增加坐标注意力机制（CA），使模型能更准确地定位和聚焦感兴趣区域。分割葡萄穗梗，通过优化质心算法计算出最佳采摘点。试验结果表明，改进后模型在葡萄数据集上的精确率、召回率和平均精度均值（mAP）达到 90.1%、93.2% 和 93.6%，模型参数量与模型内存占用量仅为 2.2×10^6 和 4.9 MB。研究结果可为采摘机器人智能化采摘提供理论支持。

关键词：葡萄穗梗；图像分割；YOLO v11n－MBCa 模型；采摘点；轻量化
中图分类号：TP391　　**文献标识码**：A　　**文章编号**：1000-1298(2025)09-0618-10　　**OSID**：

Lightweight Grape Cluster Stalk Picking Point Localization Method and Test Based on Improved YOLO v11n

ZHUANG Yu　XU Kunlin　LI Meiqi　DAI Jize　WANG Jinfeng
(*College of Engineering*, *Northeast Agricultural University*, *Harbin* 150030, *China*)

Abstract: Aiming to address the problem of picking point localization caused by slender stem morphology and weak chromatic contrast with the environment, a lightweight segmentation model was proposed based on the improved YOLO v11n architecture, designated as YOLO v11n－MBCa. Firstly, a high-quality dataset was constructed and a lightweight MobileNetV4 network was introduced to replace the original backbone network of YOLO v11n. Simultaneously, a weighted bidirectional feature pyramid network (BiFPN) incorporating a coordinated attention (CA) mechanism was implemented in the feature fusion layer (Neck). Finally, image segmentation was performed and the optimal picking points for grape cluster stems were calculated by using an optimized center of mass algorithm. Experimental results demonstrated that the YOLO v11n－MBCa model achieved significant improvements across multiple performance metrics. Compared with the baseline YOLO v11n model, as well as series-equivalent models, including YOLO v5n, YOLO v8, and Small－YOLO, the proposed model exhibited average accuracy increases of 6.2, 34.8, 9.5, and 17.4 percentage points, respectively. Precision (P) metrics showed improvements of 2.7, 21.0, 3.6, and 12.1 percentage points, respectively. With only 2.22×10^6, the model demonstrated balanced optimization between high accuracy and low missed detection rates. The detection speed reached 59.59 f/s, meeting real-time requirements for automated picking equipment. Additionally, four picking scenarios (non-obstructed, obstructed, multi-target, and long-distance) were evaluated through 127 experimental trials, yielding a 91.3% success rate in picking point calculation. These findings validated the effectiveness of the lightweight architecture design, multi-scale feature fusion strategy, and attention mechanism implementation in complex agricultural environments, particularly for detecting elongated grape cluster stems. A technical foundation for future intelligent harvesting systems was established.

Key words: grape cluster stalk; image segmentation; YOLO v11n－MBCa model; picking point; lightweighting

收稿日期：2025－04－23　修回日期：2025－05－10
基金项目：黑龙江省自然科学基金项目（ZL2024E001）和黑龙江省博士后科学基金项目（LBH－Z22078）
作者简介：庄煜（1993—），男，讲师，博士，主要从事智能农业装备研究，E-mail：15776628158@163.com
通信作者：王金峰（1981—），男，教授，博士生导师，主要从事田间机械研究，E-mail：jinfeng_w@126.com

0　引言

葡萄作为典型的浆果类高附加值经济作物，其果实具有果皮薄脆、穗梗纤细等生物学特性[1]，传统机械化采收过程中果实损伤率高，严重制约产业效益提升[2-3]。近年来，采摘机器人视觉系统在果实检测领域取得进展[4]，但自然环境下，穗梗与枝条色差较小、果实遮挡形成的多尺度干扰等问题[5-6]，仍使现有模型定位精度与鲁棒性难以满足实际需求，细小目标识别难度大[7-8]。

国内外学者围绕采摘点视觉定位展开了一系列研究。陈英等[9]提出一种堆叠葡萄果粒尺寸检测算法，基于 8-邻域轮廓跟踪提取果穗轮廓曲线，实现果粒分割和识别，准确率达 72.7% 。LIU 等[10]提出了一种葡萄检测模型 YOLOX-RA，能够快速准确地识别密集生长和遮挡的葡萄串。刘平等[11]提出基于轮廓特征分析的解决方案，通过提取葡萄串轮廓拐点与类圆心特征，结合斜率计算实现目标定位，解决了复杂自然环境下葡萄串检测难题；宁政通等[12]采用深度学习方法实现了葡萄穗梗分割与采摘点定位，结果表明该方法在不同光照条件下穗梗检测平均精确率为 88%，采摘点定位准确率高达 99.43%，单帧图像处理耗时约 4.9 s。罗陆锋等[13-14]提出了一种端到端的主动视觉感知框架，利用深度强化学习（DRL）来增强复杂遮挡情况下葡萄穗梗检测，并提出一种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方法，提升遮挡场景下检测性能。孙俊等[15]提出一种改进的葡萄检测模型，其参数量仅为 7.56×10^6，在测试集上平均精度均值达 97.74%；熊俊涛等[16]提出了“挠性杆-铰链-刚性杆-质量球”模型分析方法，通过对葡萄图像进行阈值分割得到果实和果梗，并计算出各帧图像中葡萄串质心，定位准确率达 80%。WANG 等[17]设计了一个轻量级语义分割模型，将不规则穗梗分割为多个标准化穗梗，在标准化果柄中选择最优果柄精确定位采摘点；然而，现有方法为了提取葡萄穗梗特征往往会加入大量卷积层，使得模型参数量较大，难以满足移动端实时性需求[18-20]；而参数量较少的模型会导致多尺度遮挡场景下准确率显著下降。为提高葡萄采摘机器人在作业过程中的分割精度和鲁棒性，降低因葡萄穗梗分割精度低导致的采摘损伤。同时为满足实际采摘作业实时性需求，亟需开发适用于边缘计算设备轻量化模型算法[21-23]。

通过融合深度学习与计算机视觉技术，针对细小目标检测实时性与精准性需求，本文提出基于一种融合加权双向特征金字塔网络与注意力机制的轻量化分割模型 YOLO v11n-MBCa。使用轻量化网络 MobileNetV4 替换 YOLO v11n 中的特征提取层。并在特征融合层引入注意力机制和加权特征金字塔协同优化，增强对穗梗纹理和颜色特征提取，聚焦穗梗位置，以提高分割准确度。

1　材料与方法

1.1　数据采集

使用的图像采集自甘肃省天水市鲜食葡萄种植基地，采集对象为 Y 形架型培育园阳光玫瑰葡萄，图像采集时段 2024 年 6 月末至 9 月初 10:00—12:00 和14:00—17:00，所有样本均于标准化疏果作业完成后进行采集。图像采用手机拍摄，图像分辨率为 3 000 像素 ×4 000 像素。拍摄过程中，通过模拟田间作业视角，在距果穗 0.3 ~2.5 m 内，以俯视、侧视、仰视等多角度采集图像，涵盖晴天正午强光、多云漫射光及阴影遮挡等典型光照场景，经过筛选，最终保留有效样本图像 800 幅，数据信息如图 1 所示。

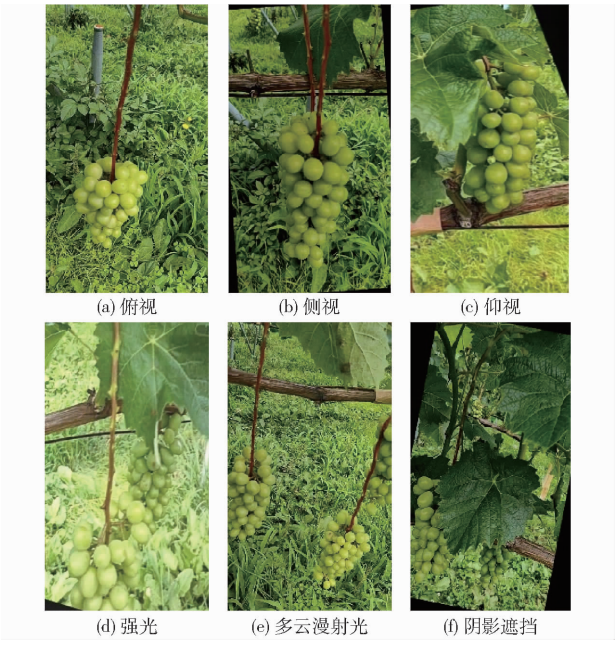


图 1　自制数据集

Fig. 1　Homemade datasets

1.2　数据预处理及数据集制作

从原始数据中随机筛选 640 幅图像作为训练集，80 幅为验证集，80 幅为测试集。将所有图像分辨率统一缩放至 640 像素 ×640 像素。随后采用 Labeling 对图像进行标注，区分果实所属，标注出一整簇葡萄果穗。针对葡萄穗梗标注，只标出主穗梗，有分叉的果实不标注其副穗梗，获得标签矩阵。标注完成后，通过 RoboFlow 平台实施多模态策略增强

数据集包括:随机旋转、透视变换、亮度与对比度调整、噪声注入和运动模糊,提升模型对光照波动与运动模糊鲁棒性。增强后数据集如图 2 所示,经数据增强后,数据集总量扩展至 4 500 幅(训练集 3 600 幅,验证集 450 幅,测试集 450 幅)。

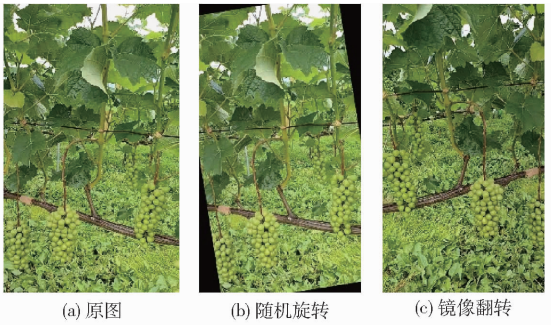


图 2 数据增强

Fig. 2 Datasets augment

2 基于改进 YOLO v11n 的穗梗检测模型

2.1 YOLO v11n 网络结构

YOLO v11 作为第 3 代检测算法,采用动态缩放策略构建 n、s、m、l、x 5 级模型。其网络结构由骨干网络、动态特征融合颈部与检测头 3 部分构成。其骨干网络将上代检测模型 YOLO v8 中的 C2f 模块替换为 C3k2 模块优化特征提取效率,结合 SPPF 模块的多尺度池化与 C2PSA 注意力机制增强全局上下文感知能力;颈部采用改进特征金字塔网络(Feature pyramid network, FPN) + 路径聚合网络(Path aggregation network, PAN)框架,引入动态特征交互策略以自适应融合多尺度特征;检测头基于深度可分离卷积改进。该模型算法在小目标分割与模型泛化性方面性能较好。

2.2 YOLO v11 模型改进

在葡萄自动化采摘过程中,葡萄穗梗展现出显著的细长形态特征,且其特征与周围环境中的枝干特征相似、与周围背景之间的色度差异较小。另外,葡萄叶片具有一定面积,会使穗梗被部分遮挡,增加了分割难度。上述葡萄穗梗几何形态、与环境相似的特征、有限的色彩对比度及枝叶遮挡等不良因素,成为了穗梗分割中的主要难点。另一方面,葡萄采摘工作需要在野外实地开展,要求检测模型通常需部署于移动设备上,例如边缘计算设备,这类设备计算资源有限,对模型尺寸和计算规模提出了更为严格的要求。

将 YOLO v11n 骨干特征提取网络(Backbone)更换为轻量级神经网络 MobileNetV4,大幅降低模型参数量。在 YOLO v11n 特征融合网络(Neck)部分加入 BiFPN(Bidirectional feature pyramid network)模块,该模块能将高层次特征和低层次特征进行双向交互和融合,增强对小目标和被遮挡目标检测精度,使模型实现更准确的分割。在特征融合网络(Neck)部分还引入注意力机制(Coordinate attention, CA),实现了对细小穗梗的多维度特征交互优化,提升定位精度与细粒度特征表达能力。相比其他注意力机制(如 SE、CBAM),CA 模块参数量和计算量更低,不增加推理延迟,适合实时性要求高的采摘场景。在此基础上,采用 YOLOX 中最后 10 轮关闭 Mosiac 数据增强的训练策略,提高检测模型准确性和鲁棒性。改进后轻量化 YOLO v11n 模型结构如图 3 所示。图中 ConvBN 表示卷积、批量归一化;ConvCA 模块通过卷积调整通道数后引入注意力机制;BiFPN 有 2 个输入量,1 个来自低层次特征,另 1 个来自高层次层特征。

2.2.1 MobileNetV4 轻量化网络

MobileNetV4(MNV4)[24] 是最新一代轻量化卷积神经网络,融合通用倒置瓶颈(Universal inverted bottleneck, UIB)和一种额外深度卷积(ExtraDW)及 Mobile MQA 注意力模块,提升了复杂农业场景下特征提取效率与定位精度。MobileNetV4 网络结构如图 4 所示。

UIB 模块[25] 通过动态可配置的深度卷积与感受野扩展策略,实现了多尺度特征高效融合,提高了计算利用率,其动态配置方法为

$$\mathrm{UIB}(\boldsymbol{x}) = \begin{cases} \mathrm{ExtraDW}(\boldsymbol{x}) & (s \leqslant 3) \\ \mathrm{SCMA}(\mathrm{ConvNeXt}(\boldsymbol{x})) & (s > 3) \end{cases} \tag{1}$$

式中 $\boldsymbol{x}$——输入的特征张量

s——网络阶段索引序号

SCMA——空间信息混合操作

该模块中引入 2 个可选的深度卷积模块,根据深度卷积模块的选择差异会产生 4 种 UIB 实例化,分别为倒置瓶颈(IB)、额外深度卷积模块(ExtraDW)、ConvNeXt 模块和前馈网络(FFN)。在 MobileNetV4 中主要使用 ExtraDW 和 ConvNext 模块进行特征提取。在浅层采用 ExtraDW 模块,通过增大感受野强化葡萄串轮廓和穗梗背景色度差异的特征提取能力;在深层采用 ConvNeXt 模块,用大核深

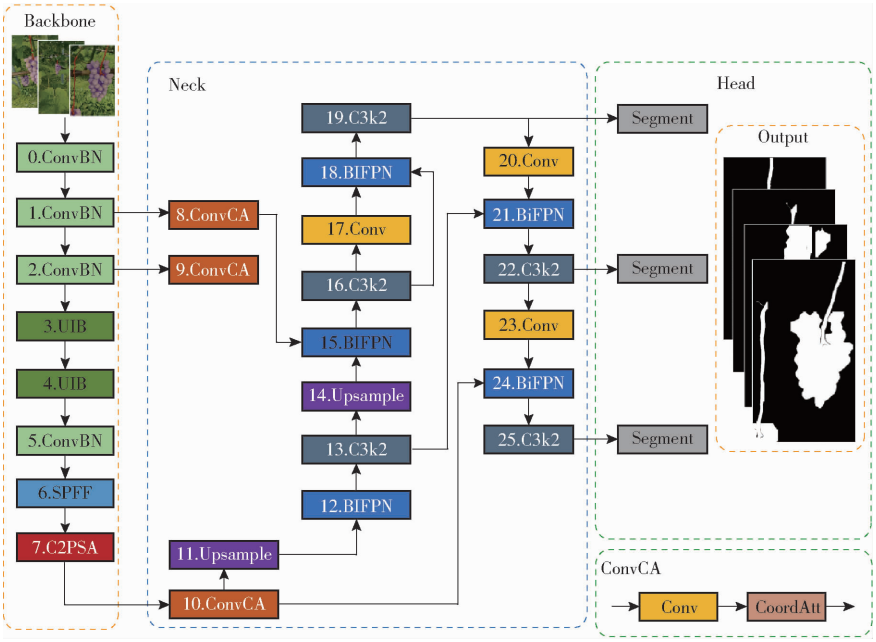


图 3　YOLO v11n－MBCa 改进网络结构图

Fig. 3　Structure diagram of YOLO v11n－MBCa improved network

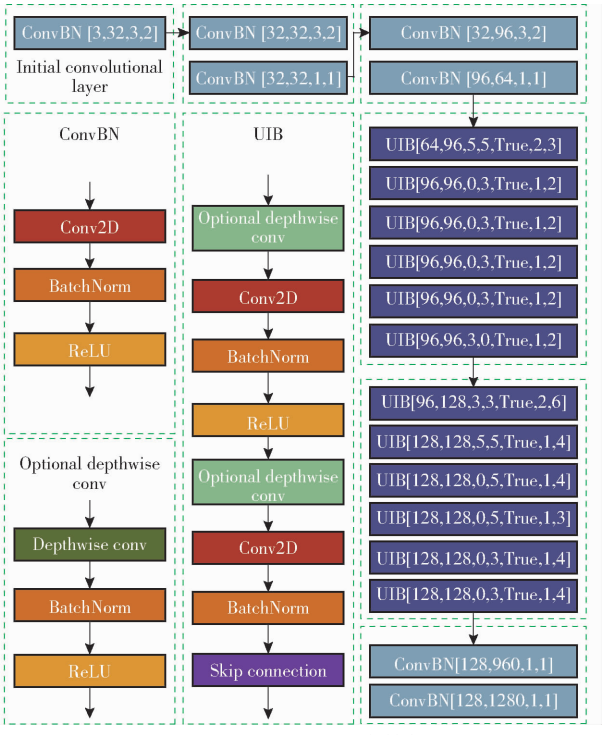


图 4　MobileNetV4 网络结构图

Fig. 4　MobileNetV4 network structure diagram

度卷积进行空间信息混合，降低计算成本，提高特征提取计算利用率[26]。

Mobile Multi－Query Attention（Mobile MQA）模块针对提高葡萄穗梗检测速度，通过空间缩减注意力（Spatial reduction attention，SRA）与键值共享机制减少计算量与内存访问，同时解耦查询（Query）与键值（Key－Value）计算，通过复用键值矩阵避免多头注意力中的冗余参数计算，其乘法累加操作数（multiply accumulate operations，MACs）较基线模型降低 28.6%[24]，显著提高了模型运算强度。Mobile MQA 表达式为

$$\text{Mobile}_{\text{MQA}}(\boldsymbol{X}) = \text{Concat}(\text{Attention}_1, \text{Attention}_2, \cdots, \text{Attention}_n)\boldsymbol{W}^O \tag{2}$$

其中

$$\text{Attention}_j = \text{Softmax}\left(\frac{(\boldsymbol{X}\boldsymbol{W}^{Q_j})(SR(\boldsymbol{X})^{\boldsymbol{W}^K})^{\mathrm{T}}}{\sqrt{d_k}}\right)(SR(\boldsymbol{X})\boldsymbol{W}^V) \tag{3}$$

式中　$\boldsymbol{X}$——输入特征矩阵

$\boldsymbol{W}^{Q_j}$——查询权重矩阵

$\boldsymbol{W}^K$——共享键权重矩阵

$\boldsymbol{W}^V$——共享值权重矩阵

SR——空间缩减（Spatial reduction）操作

d_k——键的维度

Softmax——激活函数

$\boldsymbol{W}^O$——输出线性变换矩阵

在 MobileNetV4 模型架构中，保留 SPPF 模块，在特征图上进行多尺度池化操作，将池化结果拼接起来形成一个融合多尺度信息的特征图，从而增强模型对多尺度目标分割能力。

2.2.2　加权双向特征金字塔网络

针对葡萄穗梗检测中枝叶遮挡及复杂背景干扰导致的多尺度特征退化，采用 BiFPN 模块替代原始 YOLO v11n 的特征金字塔结构（FPN/PANet），网络结构如图 5 所示。FPN 通过上采样将高层语义信息传递到低层，与横向连接的低层特征逐元素相加，形成融合后的特征金字塔，其语义信息仅能自顶向下单向传递，低层定位细节没有被反向传递，因此小目

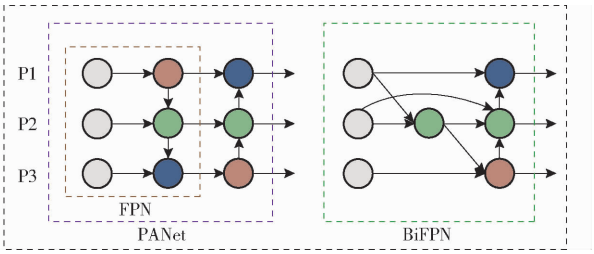


图 5 FPN、PANet、BiFPN 网络结构差异

Fig. 5 Differences in FPN, PANet, BiFPN network structure

标定位不准确。同时,不同层特征在融合时权重固定,无法动态调整不同尺度特征的重要性,可能抑制关键信息。PANet 在 FPN 基础上增加自底向上的路径,通过下采样将低层定位信息传递到高层,与高层特征级联,形成更全面的多尺度特征金字塔,但双向路径和级联操作显著增加参数量和计算成本,尤其在高分辨率特征层,且浅层与深层特征差异较大时,融合效果有限。

在 BiFPN 网络架构中,引入可学习的通道权重系数,其计算方法为

$$w_k = \frac{\mathrm{ReLU}(w_k)}{\sum_i \mathrm{ReLU}(w_i) + \varepsilon} \tag{4}$$

式中 w_k——正在关注的具体通道权重

ReLU——激活函数

w_i——任意一个通道权重

ε——非常小的正数

动态调整不同层级特征图贡献度,以区分不同输入特征重要性,高效地实施加权特征融合处理。同时反复运用自顶向下和自底向上的多尺度特征融合策略,在自上而下与自下而上的双向传播中,通过跳跃连接整合低层高分辨率特征与高层语义特征,通过加权机制避免信息冲突。相较于 PANet,BiFPN 删除只有一个输入节点,简化了网络结构。在原始输入节点和输出节点之间增加 1 条边,在不增加过多计算成本情况下丰富了特征融合。这些特点确保了特征信息能够在不同尺度间更有效地交流,进而增强了模型对于各种尺度目标的检测能力。

2.2.3 坐标注意力机制

为增强模型在复杂环境下对特定目标关键特征的提取能力,并减少背景因素干扰,引入注意力机制捕获位置信息和跨通道信息,使模型更加关注葡萄穗梗的位置。同时因为该机制轻量化能够有效避免显著增加计算成本,其结构见图 6。

注意力机制 CA 首先沿 X 轴和 Y 轴两个方向对输入特征图进行全局平均池化操作,对于第 c 个通道,在第 h 行处与第 w 列处的池化结果分别为

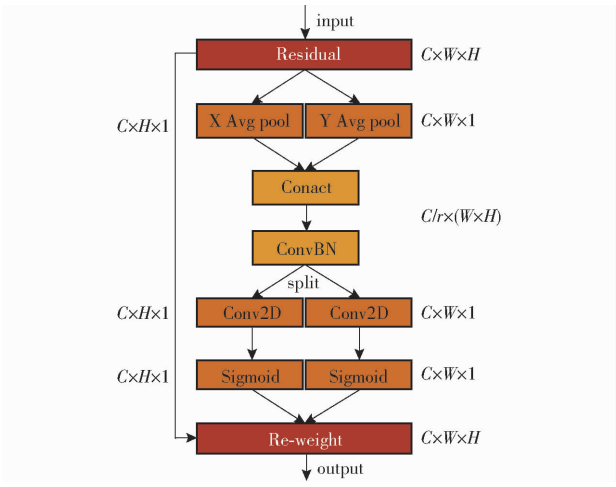


图 6 坐标注意力机制原理图

Fig. 6 Schematic of coordinate attention mechanism

$$\begin{cases} z_c^h(h) = \dfrac{1}{W}\displaystyle\sum_{0\leqslant i\leqslant W} x_c(h,i) \\ z_c^w(w) = \dfrac{1}{H}\displaystyle\sum_{0\leqslant j\leqslant H} x_c(j,w) \end{cases} \tag{5}$$

式中 H——特征图高度

W——特征图宽度

h——特征图上像素侧在行的位置

w——特征图上像素侧在列的位置

生成尺寸为 $C\times H\times 1$ 和 $C\times 1\times W$ 特征图。

将这 2 个方向上特征表示融合(Conact),并通过一个 1×1 卷积层进行降维处理,得到通道数为 C/r 的压缩特征图。特征图经过卷积、批量归一化和非线性激活函数(ConvBN)处理后,得到新的特征表示。将特征表示进行 split 操作,使特征表示分别映射到 X 轴和 Y 轴方向上,利用 1×1 卷积进行升维操作,恢复至原始输入特征图通道数 C,再结合 Sigmoid 激活函数分别计算 X 轴和 Y 轴方向上注意力权重 g^w 和 g^h。最终,通过乘法加权操作将这些注意力权重应用到对应的 X 轴和 Y 轴方向的特征图上,从而生成具有注意力增强效果的输出特征图。

此过程不仅能够提高模型对于特定尺度下特征学习能力,同时也可增强模型对不同空间位置信息的关注度,有助于提升整体准确率。在葡萄与穗梗分割中,CA 模块通过捕捉穗梗在空间坐标上的信息,使模型聚焦于穗梗,提高了复杂环境中的细小目标定位能力,提高检测精度。

2.3 采摘点确定方法

采摘点确定方法基于图像处理与形态学分析实现空间定位。首先,对分割获得的穗梗掩膜进行形态学处理,利用三核开运算消除小噪声,利用无核闭运算填充孔洞,再用高斯模糊与中值滤波使边缘平滑。剔除面积小于设定阈值的区域,排除反光、噪声

区域，防止采摘点误检。处理过程表达式为

$$M'_{stem} = \mathrm{Close}(\mathrm{Open}(M_{stem}, K_3), K_5) \tag{6}$$

$$M''_{stem} = \mathrm{MedBlur}_5(\mathrm{Thresh}_{127}(\mathrm{GaussBlur}_1(M'_{stem}))) \tag{7}$$

式中　K_n——$n \times n$ 结构核

Open——开运算

Close——闭运算

$\mathrm{GaussBlur}_1$——高斯模糊处理

Thresh_{127}——固定阈值为 127 的二值化处理

$\mathrm{MedBlur}_5$——核尺寸为 5×5 中值滤波，消除噪声并平滑边界

图像处理完成后，针对图像中含有多个穗梗与果穗情况，通过主成分分析提取穗梗掩膜的二维主方向向量，确保穗梗与果穗的所属关系，根据掩膜计算出穗梗的质心，由于计算的质心位置可能过于贴近葡萄果实，导致在摘取过程中，机械采摘作业过程中可能会对果实造成损伤。为防止上述问题，对质心进行动态优化，若质心到葡萄轮廓欧氏距离小于安全距离（设定为穗梗长度的 10%，参考《阳光玫瑰葡萄两收果贮藏保鲜技术规范》设定），利用垂直迭代搜索算法，延图像坐标轴 Y 轴负方向进行遍历，选取位于穗梗掩膜内，满足安全距离且最邻近果实轮廓的点作为采摘点。

2.4　模型训练

2.4.1　训练参数设置

计算机资源为 NVIDIA RTX3060 显卡（12 GB），Intel（R）Core（TM）i7CPU 处理器。深度学习框架为 Pytorch 1.10.0，编程语言 Python 3.11.11，CUDA 版本为 12.6，cuDNN 版本为 8.9.7。具体训练参数设置为：图像输入尺寸为 640 像素×640 像素，batch size 为 16，多线程设置为 8，初始学习率为 0.01，最小学习率为 0.000 1，优化器选择 ADAM，训练轮数设置为 300，在最后 10 轮关闭马赛克数据增强。

2.4.2　模型评价指标

为准确评估模型性能，采用准确率（Precision，P）、召回率（Recall，R）及平均精度均值（Mean average precision，mAP）作为评估模型检测精度的指标。此外，模型轻量化效果依据帧率、推理时间和模型参数量进行评估。

3　结果与分析

3.1　检测模型对比试验

将改进模型与检测模型 YOLO v5n、YOLO v8、YOLO v11n 以及其他穗梗轻量化穗梗检测模型进行评估，采用相同的自建数据集进行试验，不同模型精度和性能如表 1、2 所示。

表 1　不同模型精度对比

Tab.1　Comparison of accuracy of different models

%

检测模型	P	R	mAP
YOLO v5n	69.1	46.9	58.8
YOLO v8	86.5	88.9	84.1
YOLO v11n	87.4	86.1	87.4
Small－YOLO	78.0	70.4	76.2
YOLO v11n－MBCa	90.1	93.2	93.6

表 2　不同模型性能对比

Tab.2　Comparison of performance of different models

检测模型	检测速度/（f·s^{-1}）	推理时间/ms	参数量
YOLO v5n	46.56	15.1	2.51×10^6
YOLO v8	48.33	14.7	3.01×10^6
YOLO v11n	50.63	14.5	2.59×10^6
Small－YOLO	25.0	－	－
YOLO v11n－MBCa	59.59	12.0	2.22×10^6

由表 1、2 可知，本文提出的 YOLO v11n－MBCa 模型展现出较好的综合精度，其 mAP 可达 93.6%，较基准模型 YOLO v11n 提升 6.2 个百分点，较同系列模型 YOLO v5n、YOLO v8 以及模型 Small－YOLO 分别提升 34.8、9.5、17.4 个百分点；准确率达 90.1%，相较于基准模型 YOLO v11n、YOLO v5n、YOLO v8 以及模型 Small－YOLO 分别提升 2.7、21.0、3.6、12.1 个百分点。召回率达到 93.2%，且模型参数量仅有 2.22×10^6，验证了高精确度与低漏检率之间的平衡优化效果。提出的模型检测速度达到 59.59 f/s，满足采摘设备实时检测的要求。该结果表明，YOLO 11n－MBCa 的网络架构在特征提取效率、计算资源分配方面和检测的实时性具有显著优势。

3.2　不同注意力机制对比试验

为了分析注意力机制 CA 在葡萄穗梗定位中的性能表现，在模型特征融合层（Neck）加入 CA、CBMA、GAM 和 Focalmodulation 注意力机制，其精度和试验结果如表 3、4 所示。

表 3　各注意力机制精度对比

Tab.3　Comparison of accuracy of each attention mechanism

%

注意力机制	P	R	mAP
YOLO v11n	87.4	86.1	87.4
YOLO v11n＋CA	88.9	92.4	90.0
YOLO v11n＋CBMA	67.9	47.8	58.4
YOLO v11n＋GAM	87.1	92.4	93.1
YOLO v11n＋Focalmodulation	95.2	47.7	72.5

表 4 各注意力机制性能对比

Tab. 4 Comparison of ablation test performance

注意力机制	检测速度/(f·s^{-1})	推理时间/ms	参数量
YOLO v11n	50.63	14.5	2.59×10^6
YOLO v11n + CA	57.81	12.7	2.60×10^6
YOLO v11n + CBMA	47.30	15.1	2.66×10^6
YOLO v11n + GAM	54.90	14.1	4.23×10^6
YOLO v11n + Focalmodulation	37.50	14.2	2.70×10^6

由表 3 可知,坐标注意力机制在农业细粒度图像分割任务中展现出显著优势。相较于基线模型 YOLO v11n,引入 CA 后模型参数量仅微增 0.26%,召回率提升 6.3 个百分点,平均精度均值提升 2.6 个百分点,其通过注意力权重分配机制,在几乎不增加计算复杂度的前提下,有效增强了对细长穗梗目标的定位敏感性。对比其他注意力机制,GAM 虽实现更高 mAP,但参数量激增 63.2% 且精确率下降,凸显其高计算冗余缺陷;而 Focalmodulation 因特征平衡失准导致召回率骤降至 47.7%,验证其在复杂背景干扰下的局限性。

由表 4 可知,在 YOLO v11n 基准模型上引入 CA 后,检测速度提升至 57.81 f/s,推理时间降至 12.7 ms,而参数量仅增加 1×10^4,实现了效率与精度平衡。相较于其他注意力机制,CA 在保持轻量化特性的同时,较 GAM 机制参数量减少 38.5%,相较 CBMA、Focalmodulation 检测速度提升,推理时间减少。CA 机制理论上可聚焦于目标关键区域的特征响应,CA 模块召回率为 92.4%、mAP 为 90.0%,检测速度均衡提升,与复杂背景下对关键区域的稳定感知需求一致,表明其注意力权重合理性。同时,在参数量、检测速度与场景泛化性之间也实现权衡,为轻量化农业检测模型设计提供了参考方案。

3.3 消融试验

由于穗梗形态纤细、易与环境背景混淆,且农业场景对模型的轻量化部署和实时性有严格要求,需通过消融试验验证 MobileNetV4 主干网络是否在降低参数量的同时保留细长目标的纹理特征,BiFPN 结构能否有效融合多尺度特征以提高分割精度,注意力机制 CA 是否准确聚焦关键区域以及模型是否能满足实时性分割的需求。为验证改进模型各模块有效性,对 YOLO v11n - MBCa 进行消融试验,其精度和性能对比结果如表 5、6 所示。

由表 5、6 可知,基于 MobileNetV4 轻量化主干网络、双向加权特征金字塔(BiFPN)与坐标注意力(CA)机制协同优化的 YOLOv11n - MBCa 模型,在复杂农业场景细小图像分割任务中实现了性能与效率的平衡。用 MobileNetV4 更换原始 YOLO v11n 模型的特征提取层(Backbone)后,相较于基线模型 YOLO v11n、YOLO v11n - M 的精确率、召回率和平均精度均值均有所提高,分别提高 0.4、5.8、3.0 个百分点,但精确率提升较少,检测速度有所下降。BiFPN 模块则均衡提升召回率与平均精度值,较基线模型分别提高 4.8、6.2 个百分点,但精确率与检测速度提升不明显。CA 注意力模块则聚焦关键区域,精确率提高 1.5 个百分点,召回率提升 6.3 个百分点,平均精度值提高 2.6 个百分点,同时检测速度大幅提升。在 YOLO v11n - M 基础上加入 BiFPN 模块后,模型召回率和检测速度略有下降,精确率和平均精度均值均有所增加,表明 BiFPN 的引入使得模型预测框和真实框重合效果更好,从而有效提升模型的精确率,且保证了分割精度与检测速度的平衡。而 CA 的嵌入使得该模型在参数量与检测精度得到协同优化,相较于基线模型与模型 2、3、4、5 均有所提升,精确率、召回率、平均精度均值和检测速度均为最高,且参数量仅有 2.22×10^6,模型内存占用量仅为 4.9 MB,验证了轻量化设计、多尺度特征融合和注意力引导机制针对细长葡萄穗梗分割的有效性,为田间智能装备实时部署提供了理论依据。

表 5 消融试验精度对比

Tab. 5 Comparison of ablation test accuracy %

检测模型	*P*	*R*	mAP
YOLO v11n	87.4	86.1	87.4
YOLO v11n - M	87.8	91.9	90.4
YOLO v11n - B	88.2	90.9	93.6
YOLO v11n - CA	88.9	92.4	90.0
YOLO v11n - MB	88.4	91.8	92.1
YOLO v11n - MBCa	90.1	93.2	93.6

表 6 消融试验性能对比

Tab. 6 Comparison of ablation test performance

检测模型	检测速度/(f·s^{-1})	推理时间/ms	参数量
YOLO v11n	50.63	14.5	2.59×10^6
YOLO v11n - M	49.19	15.2	1.80×10^6
YOLO v11n - B	51.54	14.3	2.61×10^6
YOLO v11n - CA	57.81	12.7	2.60×10^6
YOLO v11n - MB	57.06	13.2	2.56×10^6
YOLO v11n - MBCa	59.59	12.0	2.22×10^6

3.4 采摘点计算试验与结果

3.4.1 试验

在自建葡萄数据集上验证了改进 YOLO v11n 模型与优化质心算法性能,涵盖多种复杂场景,如在葡萄穗梗无遮挡条件、远距离小目标条件以及多个葡萄穗梗条件下采摘点计算。规定每个穗梗只检测

出 1 个采摘点且采摘点到葡萄轮廓的欧氏距离大于等于穗梗长度的 10%（安全区域）即为计算成功，计算结果如图 7 所示，分割结果中果实部分为紫色掩膜，穗梗部分为红色掩膜。

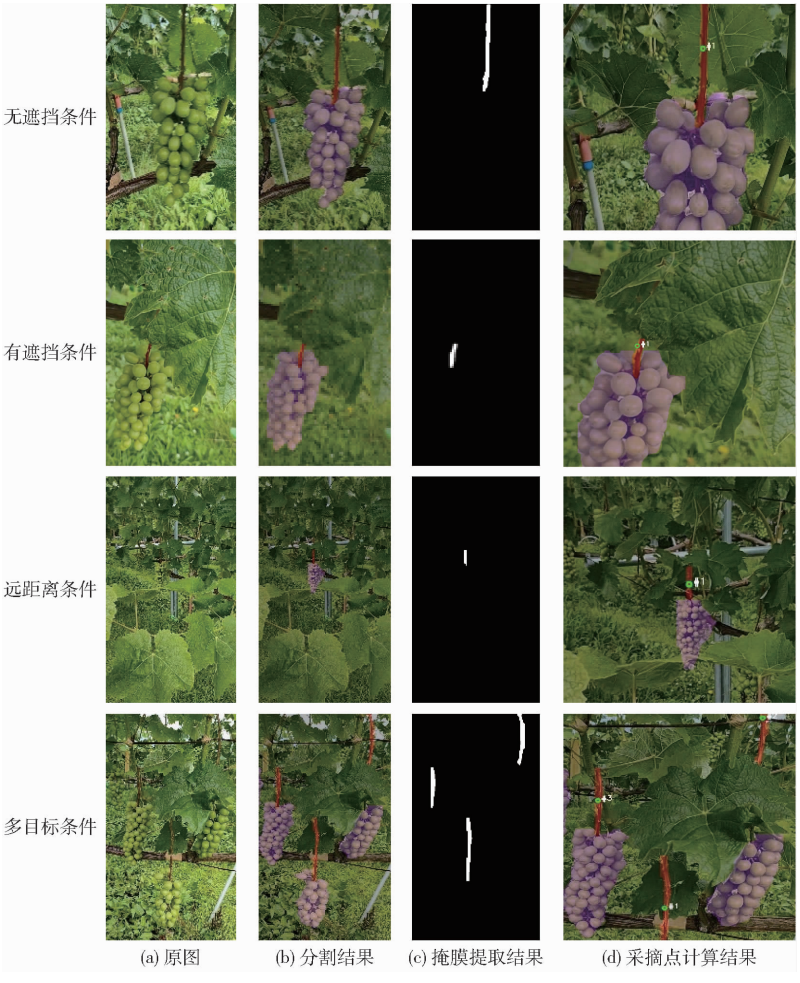


图 7　采摘点计算试验结果

Fig. 7　Picking point calculation test results

3.4.2　试验结果

选取无遮挡、有遮挡、多目标和远距离 4 种采摘条件，共计 127 次试验，计算成功率达 91.3%，试验结果如表 7 所示，表中无遮挡与有遮挡均为单目标条件。

表 7　采摘点计算结果

Tab. 7　Picking point calculation results

采摘条件	计算成功次数	计算失次数	成功率/%
无遮挡	40	2	95.2
有遮挡	35	5	87.5
多目标	25	3	89.3
远距离	16	1	94.1
总计	116	11	91.3

计算失败主要出现在有遮挡和多目标采摘条件下，结合试验结果图像采摘失败原因为：①由于枝干或树叶的遮挡，导致同一根穗梗会被分割出 2 个掩膜，则会计算出 2 个采摘点，如图 8a 所示。②由于穗梗与枝干颜色与特征纹理过于相似，导致枝干被误检为穗梗，导致采摘点会产生在枝干上，如图 8b 所示。

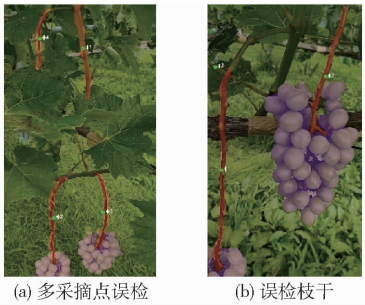


图 8　采摘点计算失败示例

Fig. 8　Failure example of picking point calculation

由于葡萄主要生长方向是纵向生长，穗梗呈长条状，果实串呈下垂姿态，其质心在水平方向上变化较小，纵向方向上差异较大。因此针对该相对稳定的生长规律，结合试验结果，通过优化质心算法计算采摘点的方法可行。

4　结论

（1）提出的 YOLO v11n－MBCa 在自建葡萄数

据集的测试集上精确率达到 90.1%,召回率达到 93.2%,平均精度均值达到 93.6%,检测速度达到 59.59 f/s,模型参数量仅为 2.22×10^6,内存占用量仅为 4.9 MB,相比于基线模型 YOLO v11n,改进模型的检测精度更高,模型更加轻量化。

(2)通过 YOLO v11n - MBCa 分割出的穗梗掩膜,结合形态学开闭运算、阈值法、图像降噪和中值滤波法对掩膜进行优化,得到边缘平滑的完整穗梗掩膜图像。通过提出的优化质心算法在自建葡萄数据集试验成功率达 91.3%。

参 考 文 献

[1] YIN W, WEN H, NING Z, et al. Fruit detection and pose estimation for grape cluster-harvesting robot using binocular imagery based on deep neural networks[J]. Frontiers in Robotics and AI, 2021, 8:626989.

[2] FAHEEM M, LIU Jizhan, CHANG Guozheng, et al. Hanging force analysis for realizing low vibration of grape clusters during speedy robotic post-harvest handling[J]. International Journal of Agricultural and Biological Engineering, 2021, 14(4): 62 - 71.

[3] 刘云玲, 张天雨, 姜明, 等. 基于机器视觉的葡萄品质无损检测方法研究进展[J]. 农业机械学报, 2022, 53(增刊 1): 299 - 308.
LIU Yunling, ZHANG Tianyu, JIANG Ming, et al. Review on non-destructive detection methods of grape quality based on machine vision[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(Supp. 1): 299 - 308. (in Chinese)

[4] TANG Y, CHEN M, WANG C, et al. Recognition and localization methods for vision-based fruit picking robots: a review[J]. Front Plant Sci., 2020, 11:510.

[5] 肖张娜, 罗陆锋, 陈明猷, 等. 基于改进 YOLO - v4 的果园环境下葡萄检测[J]. 智能化农业装备学报, 2023, 4(2): 35 - 43.
XIAO Zhangna, LUO Lufeng, CHEN Mingyou, et al. Detection of grapes in orchard environment based on improved YOLO - v4 [J]. Journal of Intelligent Agricultural Mechanization, 2023, 4(2): 35 - 43. (in Chinese)

[6] 程佳兵, 邹湘军, 陈明猷, 等. 多类复杂水果目标的通用三维感知框架[J]. 自动化与信息工程, 2021, 42(3):15 - 20.
CHENG Jiabing, ZOU Xiangjun, CHEN Mingyou, et al. General 3D perception framework for multiple types complex fruit objects[J]. Automation & Information Engineering, 2021, 42(3): 15 - 20. (in Chinese)

[7] LUO L, TANG Y, LU Q, et al. A vision methodology for harvesting robot to detect cutting points[J]. Computers in Industry, 2018, 99:130 - 139.

[8] LUO L, TANG Y, ZOU X, et al. Vision-based extraction of spatial information in grape clusters for harvesting robots[J]. Biosystems Engineering, 2016, 151: 90 - 104.

[9] 陈英, 李伟, 张俊雄. 基于图像轮廓分析的堆叠葡萄果粒尺寸检测[J]. 农业机械学报,2011,42(8):168 - 172,121.
CHEN Ying, LI Wei, ZHANG Junxiong. Overlapped grapes berry size inspection based on image contour analysis[J]. Transactions of the Chinese Society for Agricultural Machinery,2011,42(8):168 - 172,121. (in Chinese)

[10] LIU B, ZHANG Y, WANG J, et al. An improved lightweight network based on deep learning for grape recognition in unstructured environments[J]. Information Processing in Agriculture, 2024, 11(2):202 - 216.

[11] 刘平, 朱衍俊, 张同勋, 等. 自然环境下贴叠葡萄串的识别与图像分割算法[J]. 农业工程学报, 2020, 36(6): 161 - 169.
LIU Ping, ZHU Yanjun, ZHANG Tongxun, et al. Algorithm for recognition and image segmentation of overlapping grape cluster in natural environment[J]. Transactions of the CSAE, 2020, 36(6): 161 - 169. (in Chinese)

[12] 宁政通, 罗陆锋, 廖嘉欣, 等. 基于深度学习的葡萄穗梗识别与最优采摘定位[J]. 农业工程学报, 2021, 37(9): 222 - 229.
NING Zhengtong, LUO Lufeng, LIAO Jiaxin, et al. Recognition and the optimal picking point location of grape stems based on deep learning[J]. Transactions of the CSAE, 2021, 37(9): 222 - 229. (in Chinese)

[13] LUO Lufeng, LIU Bingpiao, CHEN Mingyou, et al. DRL-enhanced 3D detection of occluded stems for robotic grape harvesting [J]. Computers and Electronics in Agriculture, 2025, 229: 109736.

[14] 罗陆锋, 邹湘军, 王成琳, 等. 基于轮廓分析的双串叠贴葡萄目标识别方法[J]. 农业机械学报, 2017, 48(6): 15 - 22.
LUO Lufeng, ZOU Xiangjun, WANG Chenglin, et al. Recognition method for two overlaping and adjacent grape clusters based on image contour analysis[J]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(6):15 - 22. (in Chinese)

[15] 孙俊, 吴兆祺, 贾忆琳, 等. 基于改进 YOLO v5s 的果园环境葡萄检测[J]. 农业工程学报, 2023, 39(18): 192 - 200.
SUN Jun, WU Zhaoqi, JIA Yilin, et al. Grapevine detection in orchard environment based on improved YOLO v5s[J]. Transactions of the CSAE, 2023, 39(18):192 - 200. (in Chinese)

[16] 熊俊涛,何志良,汤林越, 等.非结构环境中扰动葡萄采摘点的视觉定位技术[J].农业机械学报, 2017,48(4):29 - 33, 81.
XIONG Juntao, HE Zhiliang, TANG Linyue, et al. Visual localization of disturbed grape picking point in non-structural environment[J]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(4):29 - 33,81. (in Chinese)

[17] WANG Jinhai, LIN Xuemin, LUO Lufeng, et al. Cognition of grape cluster picking point based on visual knowledge distillation in complex vineyard environment[J]. Computers and Electronics in Agriculture, 2024, 225:109216.

[18] 罗志聪, 何陈涛,陈登捷, 等. 基于轻量化 YOLO v8s - GD 的自然环境下百香果快速检测模型[J]. 农业机械学报,

2024, 55(8):291 – 300.
LUO Zhicong, HE Chentao, CHEN Dengjie, et al. Passion fruit rapid detection model based on lightweight YOLO v8s – GD [J]. Transactions of the Chinese Society for Agricultural Machinery, 2024,55(8):291 – 300. (in Chinese)

[19] 奚小波, 丁杰源, 翁小祥, 等. 基于轻量化 YOLO v5s – MCA 的番茄成熟度检测方法[J]. 农业机械学报, 2025, 56(3):383 – 391,436.
XI Xiaobo, DING Jieyuan, WENG Xiaoxiang, et al. Tomato maturity detection method based on lightweight YOLO v5s – MCA [J]. Transactions of the Chinese Society for Agricultural Machinery,2025,56(3):383 – 391,436. (in Chinese)

[20] 刘忠, 卢安舸, 崔浩, 等.基于改进 YOLOv8 的轻量化荷叶病虫害检测模型[J].农业工程学报, 2024, 40(19):168 – 176.
LIU Zhong, LU Ange, CUI Hao, et al. A lightweight lotus leaf pest and disease detection model based on improved YOLOv8 [J]. Transactions of the CSAE, 2024, 40(19):168 – 176. (in Chinese)

[21] 武星, 齐泽宇, 王龙军, 等. 基于轻量化 YOLO v3 卷积神经网络的苹果检测方法[J]. 农业机械学报, 2020, 51(8):17 – 25.
WU Xing, QI Zeyu, WANG Longjun, et al. Apple detection method based on Light – YOLO v3 convolutional neural network [J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(8):17 – 25. (in Chinese)

[22] 王巧华, 顾伟, 蔡沛忠, 等.基于改进 YOLO v4 的群体棉种双面破损检测方法[J]. 农业机械学报, 2022, 53(1): 389 – 397.
WANG Qiaohua, GU Wei, CAI Peizhong, et al. Detection method of double side breakage of population cotton seed based on improved YOLO v4 [J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(1):389 – 397. (in Chinese)

[23] 奚小波, 丁杰源, 翁小祥, 等. 基于轻量化 YOLO v5s – MCA 的番茄成熟度检测方法[J]. 农业机械学报, 2025, 56(3): 383 – 391, 436.
XI Xiaobo, DING Jieyuan, WENG Xiaoxiang, et al. Tomato maturity detection method based on lightweight YOLO v5s – MCA [J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(3): 383 – 391, 436. (in Chinese)

[24] QIN D, LEICHNER C, DELAKIS M, et al. MobileNetV4: universal models for the mobile ecosystem [C] // European Conference on Computer Vision. Springer, Cham, 2025.

[25] 袁培森, 欧阳柳江, 翟肇裕, 等. 基于 MobileNetV3Small – ECA 的水稻病害轻量级识别研究[J]. 农业机械学报,2024, 55(1):253 – 262.
YUAN Peisen, OUYANG Liujiang, ZHAI Zhaoyu, et al. Lightweight identification of rice diseases based on improved ECA and MobileNetV3Small[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(1):253 – 262. (in Chinese)

[26] 苗荣慧, 李志伟, 武锦龙. 基于改进 YOLO v7 的轻量化樱桃番茄成熟度检测方法[J].农业机械学报, 2023, 54(10): 225 – 233.
MIAO Ronghui, LI Zhiwei, WU Jinlong. Lightweight maturity detection of cherry tomato based on improved YOLO v7[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(10):225 – 233. (in Chinese)

[27] 李涛, 买买提明·艾尼, 古丽巴哈尔·托乎提, 等. 基于改进 YOLO v8 的小棚架下无核白葡萄穗梗识别[J]. 中国农机化学报, 2025, 46(2): 259 – 263.
LI Tao, Mamtimin·Geni, Gulbahar·Tohti, et al. Seedless white grape stem recognition under small trellises based on improved YOLO v8[J]. Journal of Chinese Agricultural Mechanization, 2025, 46(2): 259 – 263. (in Chinese)