

基于 YOLO-LTD 的轻量化温室番茄成熟度检测

李全武¹, 杨贝贝¹, 梅俸铜², 唐源³

(1. 郑州信息科技职业学院 郑州 450046; 2. 西安电子科技大学杭州研究院 杭州 310000;

3. 成都理工大学计算机与网络安全学院 成都 610059)

摘要:针对番茄果实成熟度检测中复杂背景和尺度变化引发的漏检与误检问题,以及现有方法在效率和部署上的局限性,提出一种基于 YOLO-LTD 的轻量化温室番茄成熟度检测算法。该模型以 YOLOv11-n 为基线,通过在主干网络中引入交叉注意力模块,减弱叶片、枝干和果实间的遮挡对检测精度的干扰,增强对关键区域的特征提取能力;在颈部网络中采用轻量化的 GSConv 模块替代标准卷积,在保持网络特征表达能力的同时,优化网络结构的计算效率,减少模型的参数数量和计算复杂度;在头部网络中嵌入自适应空间特征融合模块,以缓解不同尺度特征间的不一致性问题,进一步提升模型的鲁棒性和泛化能力。试验结果表明,YOLO-LTD 在平均精度均值(mAP)、召回率 R 和精度 P 上分别达到 94.23%、95.44% 和 92.07%,推理时间为 7.21 ms,模型大小仅为 5.18 Mb。与 YOLOv11-n 相比,mAP、召回率和准确率分别提升 2.50、2.80 和 1.60 个百分点,同时具备更高的效率和更小的模型尺寸。与 Mask R-CNN、Faster R-CNN 及 YOLO 系列模型相比,YOLO-LTD 在精度与效率上均具有优势,展现出在温室环境中广泛应用的潜力,为果园产量估算、作物生长监测、种植管理优化及番茄采摘机器人开发提供了理论依据和技术支持。

关键词:番茄;成熟度检测;YOLO-LTD;YOLOv11;GSConv;注意力机制;自适应空间特征融合;轻量化

中图分类号:S641.2

文献标志码:A

文章编号:1673-2871(2025)11-052-13

Lightweight maturity detection of greenhouse tomato based on YOLO-LTD

LI Quanwu¹, YANG Beibei¹, MEI Fengtong², TANG Yuan³

(1. Zhengzhou Vocational College of Information Technology, Zhengzhou 450046, Henan, China; 2. Hangzhou Institute of Technology, Xidian University, Hangzhou 310000, Zhejiang, China; 3. College of Computer and Network Security, Chengdu University of Technology, Chengdu 610059, Sichuan, China)

Abstract: To address the issues of missed and false detections caused by complex backgrounds and scale variations in tomato fruit maturity detection, as well as the limitations of existing methods in terms of efficiency and deployment, this study proposes a lightweight greenhouse tomato maturity detection algorithm based on YOLO-LTD. Building upon YOLOv11-n as the baseline, the model introduces the following innovations: (1) A cross-attention module is incorporated into the backbone network to mitigate the interference of occlusions between leaves, stems, and fruits on detection accuracy, thereby enhancing feature extraction capabilities for key regions. (2) The lightweight GSConv module replaces standard convolutions in the neck network, optimizing computational efficiency while preserving feature representation, and reducing both model parameter count and computational complexity. (3) An adaptive spatial feature fusion module is embedded in the head network to alleviate inconsistencies between multi-scale features, further improving robustness and generalization. Experimental results demonstrate that YOLO-LTD achieves a mean average precision (mAP), recall, and accuracy of 94.23%, 95.44%, and 92.07%, respectively, with an inference time of 7.21 ms and a compact model size of 5.18 Mb. Compared to YOLOv11-n, YOLO-LTD improves mAP, recall, and accuracy by 2.50 percentage points, 2.80 percentage points, and 1.60 percentage points, respectively, while exhibiting higher efficiency and smaller model size. When evaluated against Mask R-CNN, Faster R-CNN, and other YOLO variants, YOLO-LTD demonstrates superior performance in both accuracy and efficiency, highlighting its potential for widespread application in greenhouse environments. This research provides a theoretical foundation and technical support for orchard yield estimation, crop growth monitoring, cultivation optimization, and the development of tomato-picking robots.

Key words: Tomato; Maturity detection; YOLO-LTD; YOLOv11; GSConv; Attention mechanism; Adaptive spatial feature fusion; Lightweight model

收稿日期:2025-01-09;修回日期:2025-04-01

基金项目:四川省科技计划项目(2021YFN0117)

作者简介:李全武,男,讲师,主要研究方向为图像处理和农业信息工程。E-mail:liqunawu1989@163.com

通信作者:梅俸铜,男,研究员,主要研究方向为深度学习。E-mail:meifengtong1985@163.com

番茄以丰富的营养价值和较高的产量优势成为我国种植面积最大的果蔬之一^[1]。据统计,2022年我国番茄种植面积达111.3万hm²,年产量达6609万t,居全球首位^[2]。然而,尽管产业规模持续扩大,番茄生产的机械化水平仍相对滞后,主要依赖人工采摘,导致效率低下且成本高昂^[3]。近年来,随着农业机械化和自动化技术的快速发展,智能采摘机器人成为解决番茄采摘问题的重要手段,其核心在于果实的精准定位与成熟度识别。因此,开发能够准确区分不同成熟阶段的检测技术,对提升采摘效率、推动农业智能化发展具有重要意义。

传统的果实成熟度检测方法主要依赖于低维视觉特征的提取,通常通过果实的颜色、纹理和形状特征进行识别。例如,孙建桐等^[4]提出了一种基于几何形态学和迭代随机圆的番茄识别方法,利用Canny边缘检测提取果实轮廓,并结合几何形态学处理与迭代随机圆算法,实现了85.1%的识别准确率。冯青春等^[5]基于2R-G-B色差模型提取红色成熟番茄的颜色特征,并采用动态阈值分割法进行检测。Ayuningtyas等^[6]则结合色调、强度和饱和度变换,使用KNN和SVM对番茄成熟度进行分类。然而,这些方法高度依赖手工设计的特征与规则,易受光照变化、遮挡等外部环境因素的影响,难以平衡精度与实时性。相比之下,深度学习模型,特别是卷积神经网络(convolutional neural networks, CNN),能够自动提取和学习图像中的多层次特征,展现出更高的鲁棒性和精度。目前,基于深度学习的果实检测方法主要分为二阶段方法和一阶段方法。二阶段方法如Faster R-CNN^[7]和Mask R-CNN^[8],通过先生成候选区域,再对其进行精细处理,实现高精度检测。然而,这类方法计算复杂,参数量大,运行时间长,难以满足实时检测需求,尤其在农业机器人等实际应用场景中存在局限性。一阶段方法如RetinaNet^[9]和YOLO系列^[10-12],通过单次前向推理同时完成目标定位和分类,大幅缩短了检测时间,因此更适合实时检测需求,逐渐成为果实成熟度检测的研究热点。

在番茄成熟度检测领域,多项研究基于一阶段目标检测方法进行了优化。Chen等^[12]提出了一种基于MTD-YOLOv7的番茄果实及果串成熟度检测方法,通过引入尺度敏感的SIoU损失函数替代CIoU,有效提高了检测精度。常文龙等^[13]结合MobileNetv3与ECANet注意力机制,对YOLOv5s模型进行优化,增强了模型的检测性能和鲁棒性。

Yang等^[14]设计了一种基于改进的YOLOv8s模型的番茄检测方法,引入了深度可分离卷积和双路径注意力门控模块,检测精度提高了1.5%,模型减小了27%。然而,该方法在多尺度目标的检测方面仍存在一定局限性。为解决多尺度目标检测的难题,Gao等^[15]提出了自适应特征提取网络和跨层特征融合网络,有效提升了对多尺度番茄目标的检测精度。此外,杨国亮等^[16]通过设计超分辨率自适应注意力模块,强化了多尺度目标的特征提取能力,使检测精度提升至93.9%。然而,该方法未能充分考虑背景噪声对检测结果的影响。为应对此问题,张俊宁等^[17]在YOLOv5s模型的骨干网络中引入卷积注意力模块,有效提升了绿色番茄在相似背景下的检测精度,展现出更强的抗干扰能力。

尽管现有研究在特征提取、多尺度目标检测和抗噪性等方面取得了重要进展,但番茄成熟度检测仍面临以下挑战:1)叶片和枝条的遮挡以及不同光照条件的变化导致果实定位困难,严重限制了检测精度;2)果实在不同成熟度阶段的尺度特征差异,以及拍摄距离变化引起的目标尺度不一致,使得成熟度检测任务更加复杂;3)考虑到番茄成熟度检测模型需在资源受限的边缘设备上部署,设计高效且轻量级的模型成为关键。为应对这些挑战,本研究提出了一种轻量化温室番茄成熟度检测模型(YOLO-LTD)。该模型在优化检测精度的同时,注重模型轻量化与实时性,通过与Mask R-CNN、YOLOX-s、YOLOv10-n及YOLOv11-n等主流方法的对比,全面验证YOLO-LTD在检测精度、模型复杂度、推理速度以及多场景适应性等方面的综合优势。

番茄果实在生长过程中具有密集分布、异步成熟等特征,目前尚缺乏针对番茄果实成熟度检测的高质量公开数据集。为此,笔者构建了一个温室场景下番茄果实不同成熟度检测数据集,以期为后续研究提供数据支持。

1 材料和方法

1.1 温室番茄数据集

2024年7月数据采集自黑龙江省大庆市林甸县的温室番茄种植基地,涵盖了自然光、人工光、微弱光和钠黄光四种不同光照条件下的图像,分辨率为1280×720像素,共4000张图像,涵盖未成熟、半成熟和成熟3个成熟度等级。该数据集有效反映了番茄果实在空间分布、光照变化及成熟度差异等方面的复杂性。

所有图像均由经验丰富的技术人员使用 LabelImg 软件手工标注,标注内容包括番茄果实的类别和边界框坐标,且严格按照成熟度等级分类,以保证数据质量。数据集按 7:2:1 的比例划分为训练集 D_{train} 、验证集 D_{val} 和测试集 D_{test} ,分别用于模型训练、调参和性能评估。数据集的详细组成和统计信息见表 1。

1.2 检测方法 YOLO-LTD

以 YOLOv11-n 为基线,提出一种改进的轻量化番茄成熟度检测模型,名为 YOLO-LTD,其结构如图 1 所示。

YOLOv11^[18]是当前目标检测领域的先进算法,相较于前几代版本,提供了更快的检测速度、更高

表 1 番茄成熟度检测数据集					
Table 1 Dataset of tomato maturity detection					
数据集 Dataset	图像数量 Images	目标框数量 Target boxes			
		未成熟 Immature	半成熟 Half mature	成熟 Mature	全部 Total
D_{train}	2800	15 843	12 964	11 342	40 149
D_{val}	800	4 807	3 698	3 513	12 018
D_{test}	400	2 353	1 863	2 073	6 289
合计 Total	4000	23 003	18 525	16 928	58 456

的精度及效率。YOLOv11 系列包含 5 种网络模型:YOLOv11-n、YOLOv11-s、YOLOv11-m、YOLOv11-l 和 YOLOv11-x。其中,YOLOv11-n 为轻量级版本,专为资源受限的嵌入式平台上的实时检测任务设计。YOLOv11-n 包括骨干网络(backbone

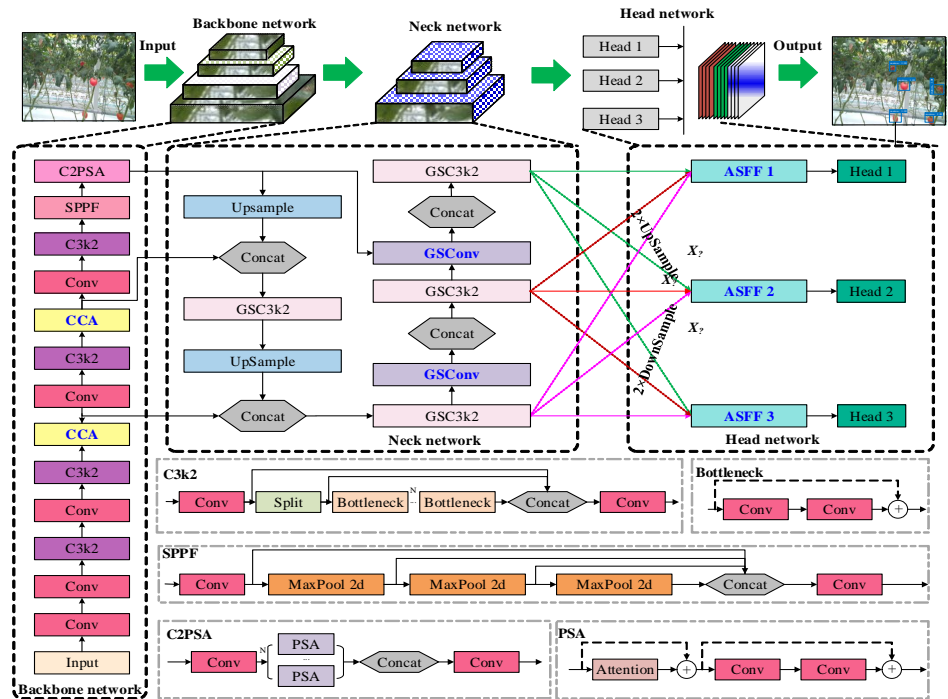


图 1 YOLO-LTD 网络结构
Fig. 1 YOLO-LTD network structure

network)、颈部网络(neck network)和头部网络(head network)。骨干网络通过卷积层和改进的C3k2 模块提取多尺度特征,随后,SPPF(spatial pyramid pooling fast)模块和 C2PSA(cross-layer spatial attention)层进一步加强多尺度特征的传递,并优化高层语义特征与低层细节的融合。颈部网络采用 FPN+PAN 结构,通过结合上采样和下采样操作,进一步提升了对多尺度目标的检测能力。头部网络由 3 个解耦的检测头组成,分别负责分类分数和回归坐标的预测。通过任务对齐机制,头部网络在联合优化分类分数和交并比(intersection of union,

IoU)的同时,有效提升了分类和定位的精度,并抑制了低质量预测框的生成。

尽管 YOLOv11-n 在多种场景下展现了高效性和轻量化优势,但在复杂温室环境中的应用仍存在性能瓶颈。具体表现为在光照变化剧烈和枝叶频繁遮挡的情况下,YOLOv11-n 易出现漏检和误检现象,尤其是在番茄成熟度分类方面的准确性较低。为此,以 YOLOv11-n 为基线,提出一种改进的轻量化番茄成熟度检测模型,名为 YOLO-LTD。主要的改进策略有 3 点:(1)在骨干网络中引入交叉注意力(criss-cross attention,CCA)模块^[19]; (2)在颈部网

络中采用 GSConv 模块^[20]替换标准卷积;(3)在头部网络中嵌入自适应空间特征融合(adaptive spatial feature fusion, ASFF)模块^[21]。

1.2.1 CCA 注意力模块 在温室环境中,番茄果实常常受到叶片、枝条以及果实间的相互遮挡,这明显增加了番茄果实特征的不确定性,进而增加了成熟度检测的难度。引入全局上下文信息有助于更好地描述番茄果实与背景之间的内在关系,从而减轻这些遮挡带来的不确定性对检测性能的影响。为此,本文在骨干网络的第2和第3个 C3k2 模块之后引入具有更强上下文捕捉能力的交叉注意力

(criss-cross attention, CCA)模块,以增强模型对番茄果实图像全局上下文信息的感知能力。CCA 模块的具体结构如图2所示。

该模块实现对全局上下文信息的有效捕捉过程如下:

首先,对于骨干网络获取的输入特征 $F \in \mathbb{R}^{C \times H \times W}$,通过2个 1×1 的卷积层(Conv1d_1 和 Conv1d_2)进行处理,减少特征通道数至 C' ,从而生成查询特征图 $Q \in \mathbb{R}^{C' \times H \times W}$ 和键特征图 $K \in \mathbb{R}^{C' \times H \times W}$ 。随后,利用仿射变换计算查询特征图 Q 与键特征图 K 中所有像素之间的相关性矩阵 $D \in \mathbb{R}^{(H+W-1) \times W \times H}$,如公式(1)所示:

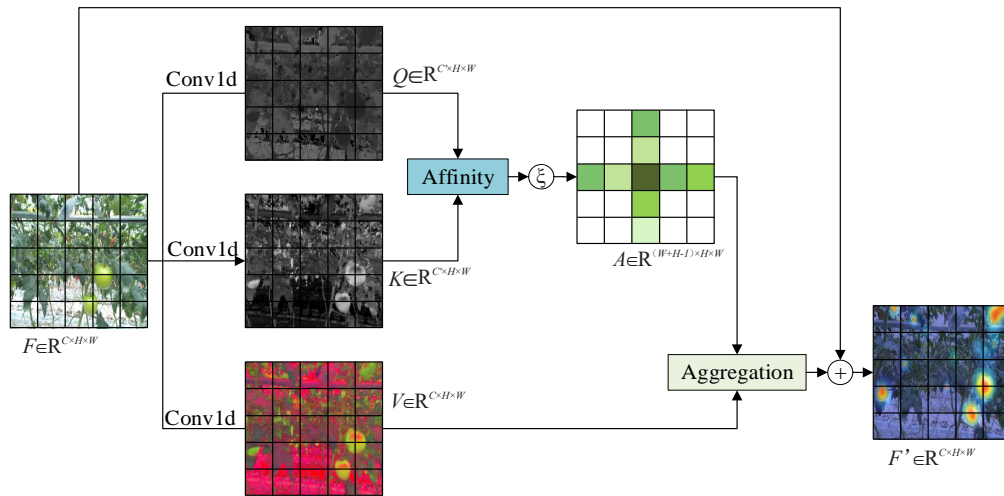


图2 CCA 模块的网络结构

Fig. 2 Network structure of CCA module

$$d_{p,q} = Q_p \times K_q^T \quad (1)$$

式中: Q_p 表示查询特征图 Q 中位置 p 处的特征向量, K_q 表示键特征图 K 中与位置 q 同行同列的特征向量, $d_{p,q} \in D$ 表示 Q_p 和 K_q 之间的相似性。

接着,对相关性矩阵 D 进行 Softmax 归一化处理,生成注意力特征图 $A \in \mathbb{R}^{(H \times W) \times (H+W-1)}$,如公式(2)所示:

$$A = \text{softmax}(D) \quad (2)$$

进一步对输入特征 F 通过一个尺寸为 1×1 的卷积层(Conv1d_3)处理,生成用于特征适配的值特征图 $V \in \mathbb{R}^{H \times W \times C'}$ 。最后,将注意力特征图 A 与值特征图 V 加权求和,与输入特征 F 相加得到融合了上下文信息的细化特征 F' ,如公式(3)所示:

$$F' = \sum_{q=1}^{H+W-1} A_{p,q} \times V_p + F \quad (3)$$

式中: $A_{p,q}$ 表示注意特征图中的位置 p 和 q 处的标量权重。 $V_p \in \mathbb{R}^C$ 表示值特征图 V 中像素 p 对应的

特征向量。

通过引入 CCA 模块,模型能够更有效地捕捉番茄果实与背景之间的全局上下文关系,提升在复杂温室环境下的检测精度。

1.2.2 GSConv 模块 在番茄果实成熟度检测任务中,模型的高效性对满足实时检测需求以及在资源受限的嵌入式平台上部署至关重要。近期研究表明,采用 GSConv 替代传统的标准卷积(standard conv, SConv),可以明显降低模型复杂度,同时保持甚至提升模型的检测精度,因而成为优化轻量级模型的一种有效策略^[20]。图3展示了 GSConv 的网络结构。

如图3所示,GSConv 模块主要由三部分组成:标准卷积(convolution, Conv)、深度可分离卷积(depthwise separable convolutions, DWConv)和通道洗牌(shuffle)。在该模块中,输入特征同时通过标准卷积和深度可分离卷积,以提取不同层次的特征信息。随后,利用通道洗牌策略,将2种卷积产生

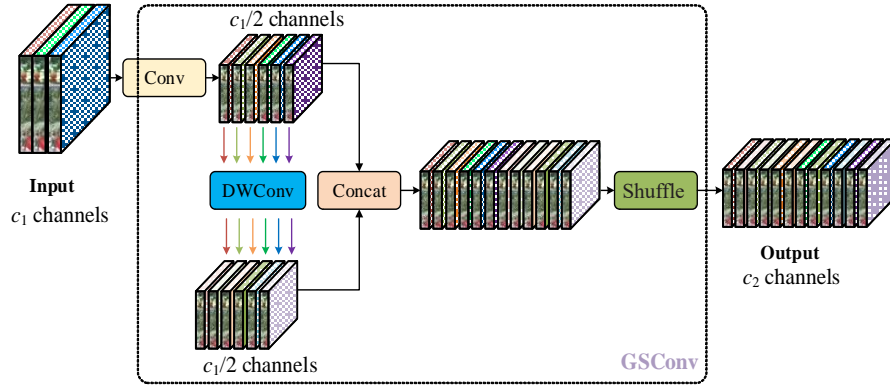


图3 GSConv 的网络结构

Fig. 3 Network structure of GSConv

的特征进行融合,增强特征表达的多样性和丰富性。这种融合方法不仅降低了计算成本,还在保持与标准卷积相近的特征提取能力的同时,提升了模型的非线性表达能力,使其更适合应用于轻量化目标检测器。标准卷积和 GSConv 的计算代价分别如公式(4)和(5)所示:

$$GFLOPs_{Conv} = W'H'K_1'K_2'C_1'C_2; \quad (4)$$

$$GFLOPs_{GSConv} = W'H'K_1'K_2' \frac{C_2}{2} (C_1+1)。 \quad (5)$$

式中: W 和 H 表示特征映射的宽度和高度的维数; K_1 和 K_2 表示卷积核大小; C_1 和 C_2 分别表示输入和输出的通道数。由上述公式可以看出,GSConv 的计算量较标准卷积减少了约 50%,从而有效实现模型的轻量化。

尽管 GSConv 在降低计算复杂度和保持精度方面表现出色,但在模型中并未完全替代所有标准卷积。根据相关研究^[22],若在网络的所有阶段应用 GSConv,可能导致网络深度增加、数据流阻力加大,从而延长推理时间,影响实时性。因此,本文仅在颈部网络中引入 GSConv,但保留骨干网络中的标准卷积结构。这样做的优势在于:首先,颈部网络主要负责多尺度特征的融合与传递,引入 GSConv 可以有效减少计算量并提升特征融合效率;其次,颈部网络处理的特征图在通道数和空间尺寸上较小,使用 GSConv 能明显降低模型复杂度;最后,保留骨干网络的标准卷积结构,有助于保持特征提取能力的高效性,避免对性能产生负面影响。

1.2.3 ASFF 模块 受种植环境和成熟度差异的影响,番茄果实在图像中的尺度差异较大。尽管 YOLOv11-n 通过特征金字塔(feature pyramid network, FPN)设计应对了多尺度检测问题,但由于不

同尺度下番茄果实特征的不一致性,特征金字塔在 YOLOv11-n 中的有效性仍受到一定限制。具体而言,YOLOv11-n 中的高层特征图主要负责较大果实的检测,而低层特征图则侧重于较小果实的特征提取。这种特征分配方式可能导致某一尺度的目标在其他尺度的特征图中被误判为背景。此外,特征拼接或加权融合过程中,不同层次的特征可能存在冲突,这不仅影响了特征的有效融合,还可能对训练过程造成负面影响,从而削弱了 YOLOv11-n 在多尺度目标检测中的性能。为解决上述问题,本试验引入了自适应空间特征融合(adaptive spatial feature fusion, ASFF)模块。该模块通过特征重缩放和自适应融合两阶段操作,有助于提高特征的尺度不变性,并几乎不增加推理开销。ASFF 模块的结构如图 4 所示,其能够在不同尺度的特征之间实现更有效的对齐和融合,从而增强了模型在复杂温室环境中的鲁棒性。

在特征重缩放阶段,ASFF 模块针对 YOLOv11-n 中特征金字塔网络的 3 个尺度(20×20、40×40 和 80×80)调整了上采样和下采样策略。对于上采样,使用尺寸为 1×1 的卷积将番茄果实图像特征的通道数压缩至对应层的通道数,并通过插值上采样提升分辨率。对于 1/2 下采样,使用尺寸为 3×3、步长为 2 的卷积,同时调整通道数和分辨率以匹配对应层。对于 1/4 下采样,在步长为 2×2 的卷积前添加步长为 2×2 的最大池化层,以进一步降低分辨率。

在自适应融合阶段,ASFF 模块自适应地学习不同尺度特征图的空间权重,实现加权融合,如公式(6)所示:

$$y_{ij}^l = \alpha_{ij}^l \times x_{ij}^{1 \rightarrow l} + \beta_{ij}^l \times x_{ij}^{2 \rightarrow l} + \gamma_{ij}^l \times x_{ij}^{3 \rightarrow l}; \quad (6)$$

$$\alpha_{ij}^l + \beta_{ij}^l + \gamma_{ij}^l = 1。 \quad (7)$$

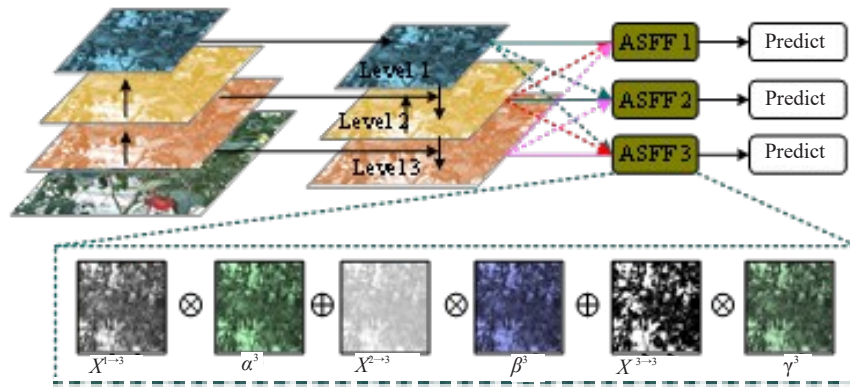


图 4 ASFF 的网络结构

Fig. 4 The network structure of ASFF

式中: x^l 表示在第 l 层分辨率下番茄果实特征, y_{ij}^l 表示融合后特征图。 $x_{ij}^{n \rightarrow l}$ 表示从第 n 层调整到第 l 层后, 在位置 (i, j) 处的番茄果实特征向量; $\alpha_{ij}^l, \beta_{ij}^l$ 和 γ_{ij}^l 表示从不同层次调整到第 1 层特征图自适应学习的空间重要性权重。以 α_{ij}^l 为例, 空间重要性权重可以表示为:

$$\alpha_{ij}^l = \frac{e^{\lambda_{\alpha_{ij}}^l}}{e^{\lambda_{\alpha_{ij}}^l} + e^{\lambda_{\beta_{ij}}^l} + e^{\lambda_{\gamma_{ij}}^l}} \quad (8)$$

式中: $\alpha_{ij}^l, \beta_{ij}^l$ 和 $\gamma_{ij}^l \in [0, 1]$ 是通过将 $\lambda_{\alpha_{ij}}^l, \lambda_{\beta_{ij}}^l, \lambda_{\gamma_{ij}}^l$ 应用 softmax 函数得到的。此外, $x_{ij}^{1 \rightarrow l}, x_{ij}^{2 \rightarrow l}, x_{ij}^{3 \rightarrow l}$ 的权重尺度图 $\lambda_{\alpha_{ij}}^l, \lambda_{\beta_{ij}}^l, \lambda_{\gamma_{ij}}^l$ 是通过一个尺寸为 1×1 的卷积层计算的, 并可通过反向传播进行学习。

通过引入 ASFF 模块, 模型能够自适应地融合不同尺度的特征, 缓解多尺度特征间的不一致性, 提升番茄果实检测精度。

1.3 试验配置及参数设置

为确保温室番茄果实成熟度检测试验的公平性, 本研究在相同硬件平台和统一硬件配置下开展所有试验。在试验中, 输入图像尺寸设置为 640×640 , 初始学习率为 0.01, 批量大小为 16, 训练轮次为 300, 动量为 0.937, 权重衰减为 0.000 5。为增加训练集和验证集的多样性, 采用了多种数据增强技术, 包括随机亮度调整、高斯噪声添加、运动模糊、随机翻转和尺度缩放等。此外, 模型训练过程中使用了 AdamW 优化器以提高优化效率。试验于郑州信息科技职业学院研教楼 211 实验室完成, 试验具体硬件配置和软件环境详见表 2。

1.4 评价指标

为全面衡量各模型的性能, 采用目标检测领域常用的准确性、效率和复杂性评价指标。

准确性评估指标: 精度 (precision, P)、召回率

表 2 试验硬件配置和软件环境

Table 2 Experimental hardware configuration and software environment

硬件	配置参数	软件	配置参数
Hardware	Configuration parameters	Software	Configuration
CPU	Intel Xeon Gold 5220 CPU	系统	Ubuntu 18.04 System
GPU	NVIDIA GeForce RTX 3090 24 G	Pytorch	1.9.1
内存	64 G	CUDA	11.4
Memory			
硬盘	8 T	CUDNN	8.2.4
Hard disk			

(recall, R) 和平均精度均值 (mean average precision, mAP), 其计算公式如下:

$$P = \frac{T_p}{T_p + F_p} \cdot 100\%; \quad (9)$$

$$R = \frac{T_p}{T_p + F_N} \cdot 100\%; \quad (10)$$

$$mAP = \frac{\int_0^1 P \times R \times dR}{M} \quad (11)$$

式中: T_p 为真阳性数, F_N 为假阴性数, F_p 为假阳性数, M 为成熟度级别总数。

效率评估指标: 平均检测时间 (detection time, DT), 评估了模型完成 1 次检测所需的平均时间, 以衡量其实时检测能力, 计算公式为:

$$DT = \frac{1}{N} \sum_{i=1}^N t_i \quad (12)$$

式中: N 为测试集中的图像总数。

复杂性评估指标: 参数量 (parameters, Par), 即模型中可训练参数的数量; 每秒浮点运算数 (FLOPs), 即模型在推理过程中所需的计算资源; 模型大小 (model size, MS), 即模型在存储设备上所占用的空间, 影响其在资源受限环境中的部署能力。

2 结果与分析

2.1 消融试验

本研究通过消融试验分析 CCA 模块、GSConv 模块和 ASFF 模块对检测性能的影响。所有试验均在相同的环境和超参数设置下进行,以确保结果的准确性。YOLO-LTD 模型的消融试验结果如表 3 所示,其中“√”表示启用该模块,“-”表示未启用该模块。

通过消融试验可以看出,引入 CCA、GSConv

和 ASFF 模块均对温室番茄成熟度检测性能产生了积极影响。仅引入 CCA 模块时,精度 P、召回率 R 和 mAP 分别为 91.34%、94.11%和 93.14%,较基线模型提升 0.87、1.47 和 1.41 百分点;与此同时,检测时间和参数量等复杂度指标几乎保持不变。这表明 CCA 模块捕获的全局上下文信息有助于提高模型的检测精度,同时不会增加模型的计算负担。仅引入 GSConv 模块时,虽然检测精度 P 和召回率 R 变化不明显,但推理时间减少 1.13 ms,参数量、FLOPs 和模型大小分别降低 0.33 M、1.16 G 和 0.32 Mb,凸

表 3 消融试验结果
Table 3 Results of ablation experiments

Yolov11-n	CCA	GSConv	ASFF	P/%	R/%	mAP/%	DT/ms	Par/M	FLOPs/G	MS/Mb
√				90.47	92.64	91.73	7.81	2.60	6.54	5.43
√	√			91.34	94.11	93.14	7.81	2.60	6.58	5.43
√		√		90.32	92.71	91.69	6.68	2.27	5.42	5.11
√			√	91.08	93.76	92.75	8.13	2.76	7.01	5.39
√	√	√	√	92.07	95.44	94.23	7.21	2.52	6.14	5.18

显了 GSConv 在减轻模型复杂度和加速推理方面的优势。仅引入 ASFF 模块时,模型的精度 P、召回率 R 和 mAP 相较于原模型分别提升 0.61、1.12 和 1.02 百分点,但检测时间和模型复杂度增加,说明 ASFF 模块在提升特征融合效果的同时,也带来了额外的计算负担。

综合引入 CCA、GSConv 和 ASFF 模块时,性能达到最佳:精度 P 达到 92.07%,召回率 R 达到 95.44%,mAP 达到 94.23%,分别比基线模型提升 1.60、2.80 和 2.50 百分点;检测时间缩短至 7.21 ms,参数量和 FLOPs 降至 2.52 M 和 6.14 G,模型大小减小至 5.18 Mb。该结果表明,通过合理组合 CCA、GSConv 和 ASFF 模块,不仅能明显提升温室番茄成熟度检测的准确性和鲁棒性,还能有效控制模型的计算与存储开销,从而在实际应用中实施高精度、低复杂度的优化方案。

为直观展示 YOLOv11-n 和 YOLO-LTD 模型在温室番茄成熟度检测任务中的表现差异,使用 Grad-CAM 梯度加权激活热图进行可视化结果分析,如图 5 所示。图中的红色区域代表了模型对目标的强烈关注,而蓝色区域则代表较低的关注度。

从热图中可以看出,YOLOv11-n 在处理较为明显的目标(如成熟番茄)时,能够集中关注并准确检测。然而,对于未成熟番茄、小尺度番茄以及遮挡的番茄,其热图中的关注区域较为分散,导致漏检现象较为明显。这是因为 YOLOv11-n 在面对多目

标遮挡和颜色相似的情况时,难以精确区分目标与背景,进而影响检测精度。相比之下,YOLO-LTD 通过 CCA 和 ASFF 模块,在复杂背景和小尺度目标的情况下展现出明显优势。特别是对于绿色番茄或与叶片颜色相似的果实,YOLO-LTD 能够精准聚焦于番茄的关键特征,减少了因颜色相似和遮挡带来的误检和漏检。此外,YOLO-LTD 通过 ASFF 模块的多尺度特征融合,确保即使在小果实或部分果实遮挡的情况下,也能提供更精确的定位。

2.2 不同模型的对比试验

为验证 YOLO-LTD 的有效性和优越性能,将其与主流目标检测模型进行了比较,包括 Mask R-CNN、Faster R-CNN、YOLOv8-n、YOLOX-s、YOLOv9-t^[20]、YOLOv10-n^[23]和 YOLOv11-n。图 6 展示了不同模型对番茄成熟度类别的分类结果混淆矩阵,横坐标为预测标签,纵坐标为真实标签。

图 6 显示,YOLO-LTD 在各个成熟度类别上的分类性能优于其他模型。具体来说,YOLO-LTD 在 immature 和 medium 类别上,分别达到了 98.39%和 99.18%的准确率,展示了其卓越的分类均衡性。相比之下,其他模型在这 2 个类别的分类中存在明显的混淆现象。例如,Mask R-CNN 在 immature 类别上的分类准确率仅为 89.12%,明显低于 YOLO-LTD。出现这种情况的原因主要是未成熟和半成熟的番茄果实通常呈现与背景叶片和枝干相似的颜色,增加了模型的辨识难度。此外,这些果实

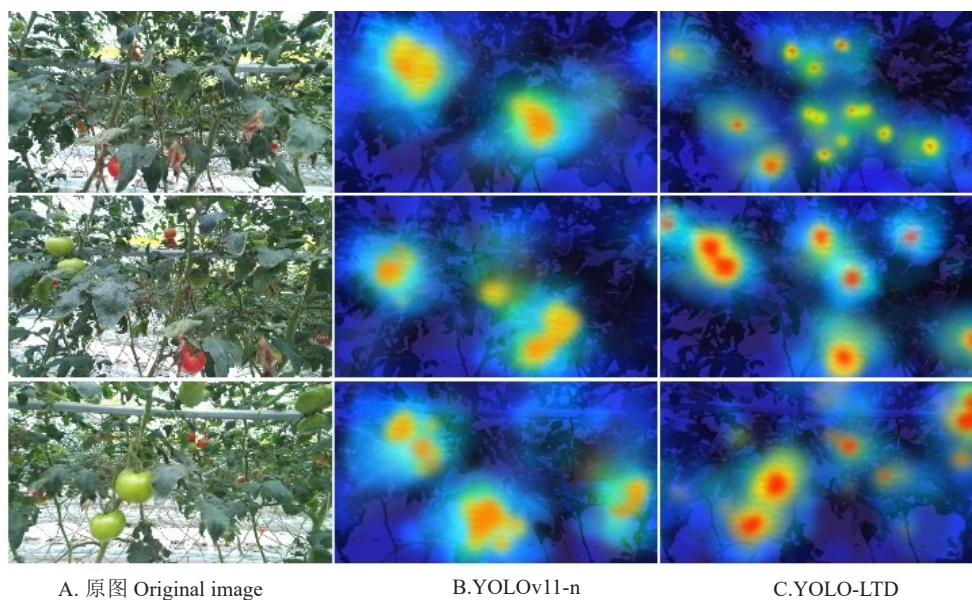


图 5 YOLOv11-n 和 YOLO-LTD 激活热图对比

Fig. 5 Comparison of YOLOv11-n and YOLO-LTD activation heatmaps

在外观上缺乏明显的色彩对比或形态特征,致使难以准确区分,从而影响分类准确性。相比之下,YOLO-LTD 通过引入改进机制有效解决了这些问题,明显提高了分类的精度和鲁棒性。此外,在 mature 类别上,YOLO-LTD 的分类准确率达到 100.00%,相比其他模型分别提高了 7.15、8.00、6.90、2.05、6.09、5.25 和 4.02 百分点。以上结果表明,YOLO-LTD 在番茄成熟度分类中具有更高的准确性,有效减少了误检现象。

为进一步多维评估 YOLO-LTD 的性能,对各模型在精度(P)、召回率(R)和 mAP 方面进行了对

比,如图 7 所示;同时,对各模型的效率和复杂度进行了比较,涵盖每秒处理时间(DT)、参数量(Par)、计算量(FLOP)和模型尺寸(MS),如图 8 所示。

由图 7~8 可见,YOLO-LTD 在精度指标上超越所有对比模型,同时相较于传统的两阶段检测模型 Mask R-CNN 和 Faster R-CNN,YOLO-LTD 在计算效率和模型复杂度上具有明显优势,明显提升了计算效率,并有效减少了存储和计算需求。这一优势源于 YOLO-LTD 作为一阶段模型在实时性和轻量化上的突出表现,而二阶段模型由于需要生成候选区域,计算量和推理时间相对较多。此外,与同

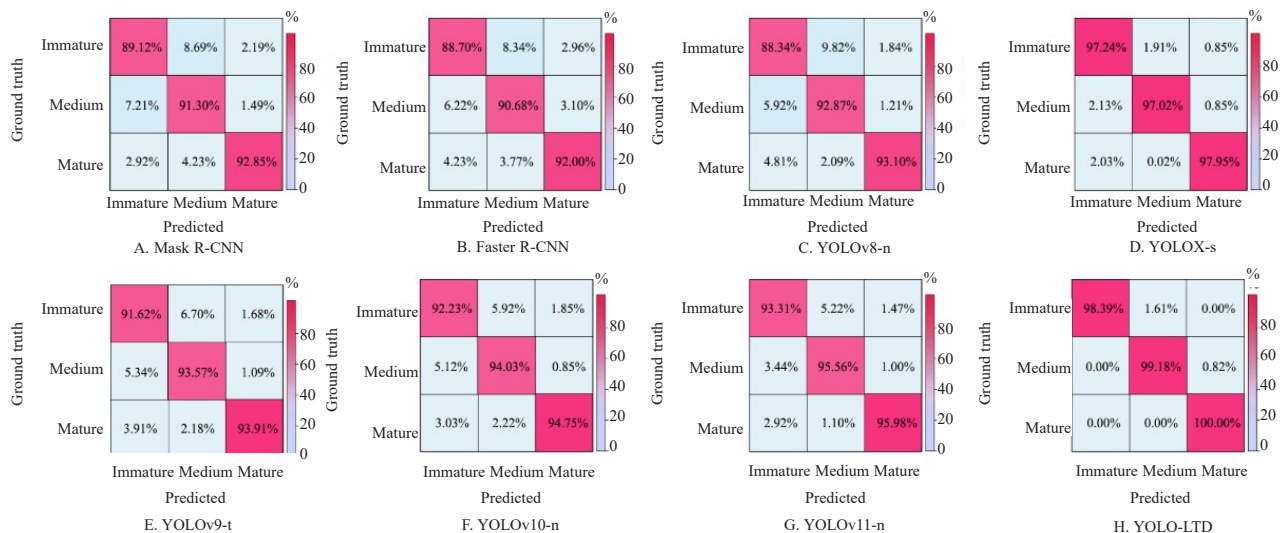


图 6 混淆矩阵对比示意图

Fig. 6 Comparison diagram of the confusion matrix

类的 YOLO 系列一阶段模型相比, YOLO-LTD 在检测精度上进一步提升, 展现了更强的整体性能。综上, YOLO-LTD 整体性能在番茄果实成熟度检测中明显提高, 其在精度、效率和模型复杂度方面的

优势充分验证了本文模型的有效性和实用性。

图 9 进一步展示了几种先进方法的检测结果对比, 直观地体现了 YOLO-LTD 在复杂温室环境下的优势。

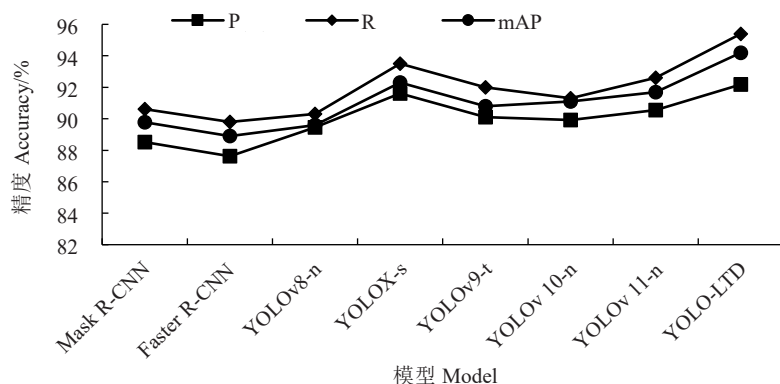


图 7 YOLO-LTD 模型与不同模型的精度对比

Fig. 7 Comparison of accuracy between YOLO-LTD model and different models

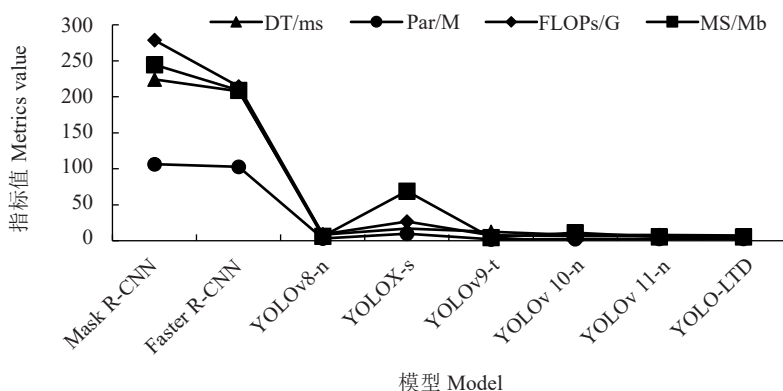


图 8 YOLO-LTD 模型与不同模型的效率和复杂度对比

Fig. 8 Comparison of efficiency and complexity between YOLO-LTD model and different models

由图 9 可以看出, 对于大多数明显的目标, 所有方法均能有效进行检测。然而, 在面对遮挡和小尺度目标时, YOLOX-s、YOLOv10-n 和 YOLOv11-n 出现一定的漏检情况, 并且存在误分类的情况, 尤其是将叶片误识别为番茄果实。相比之下, YOLO-LTD 在处理这些有挑战性目标时展现出更高的鲁棒性, 能够准确识别并分类番茄果实, 特别是在小尺度和遮挡情况下的表现优于其他方法。此结果突出了 YOLO-LTD 在复杂环境下的卓越检测能力, 进一步证明了所提的优化策略在提升番茄成熟度检测精度方面的有效性。

2.3 模型鲁棒性分析

2.3.1 不同尺度下检测结果对比 在番茄检测中, 由于果实大小和成像距离的变化, 呈现出明显的尺度差异。为评估模型对不同尺度目标的鲁棒性, 根据目标在图像中的像素点数量定义了 3 种尺度: 大

尺度目标(占图像总像素的 1%以上), 中尺度目标(占图像总像素的 0.3%~1%), 小尺度目标(占图像总像素的 0.3%以下)。不同尺度下本文模型与对比模型的精度对比如图 10 所示。由图 10 可见, 在小尺度目标检测中 YOLO-LTD 表现最佳, 优于其他 3 种模型, 尤其在召回率(R)上表现突出, 显示出其小目标检测中的独特优势。对于中尺度目标, YOLO-LTD 继续保持领先, 精度为 92.39%, 在召回率 R 和 mAP 方面也表现优越。在大尺度目标检测中, YOLO-LTD 仍然保持较高精度, 在 mAP (98.43%) 方面的表现最为突出, 进一步验证了其在目标检测中的强大能力。综上, YOLO-LTD 在不同尺度的目标检测中都表现出了较强的鲁棒性。

2.3.2 不同光照下检测结果对比 在温室环境中, 番茄果实面临多样的光照条件, 这些条件会影响图像质量、颜色特征及目标检测模型的表现。为评估

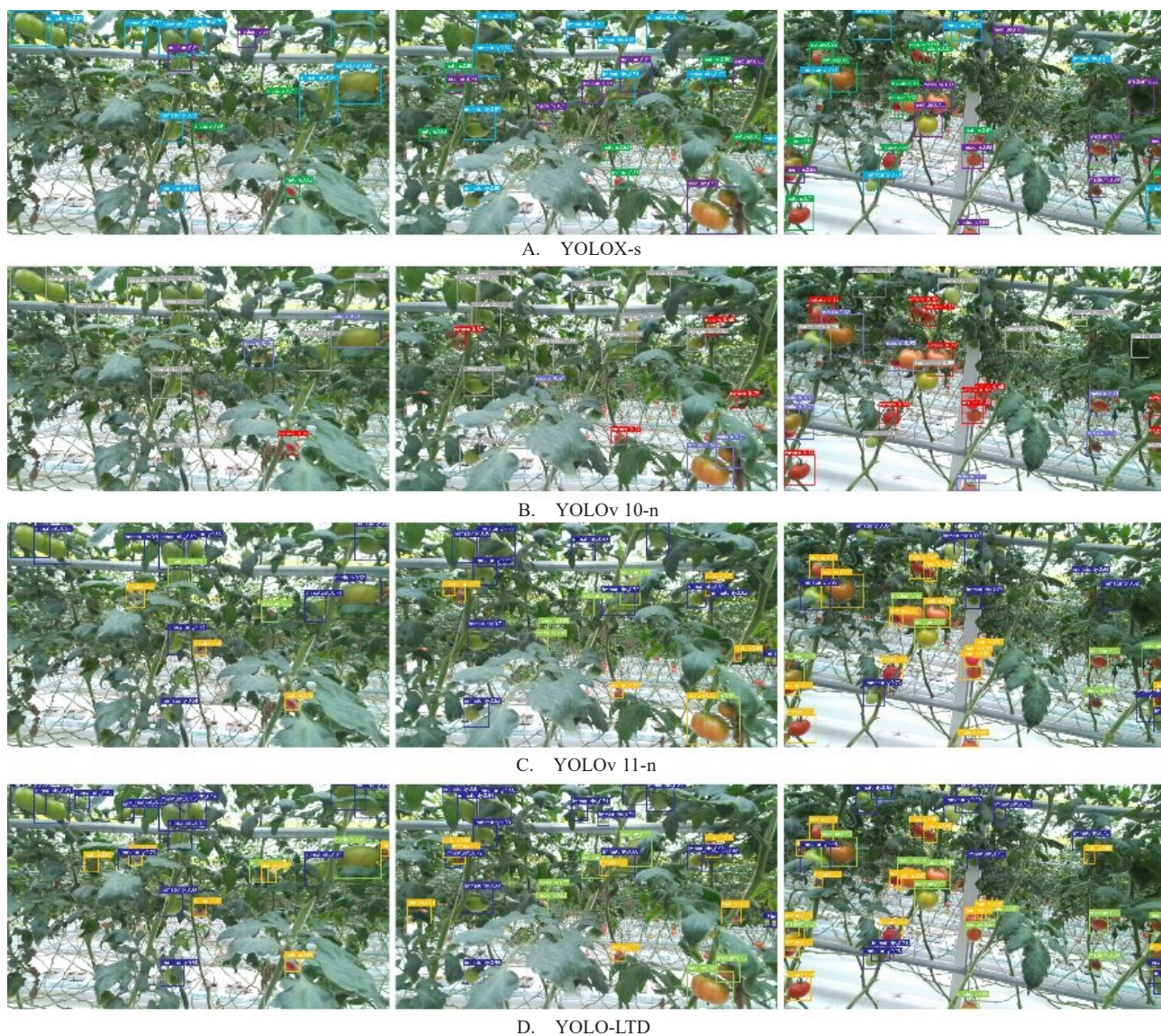


图9 检测结果对比

Fig. 9 Comparison of detection results

YOLO-LTD 模型在不同光照下的鲁棒性,在自然光、人工光、钠黄光和微弱光环境中,对比本文方法与对比模型的检测精度,试验结果如图 11 所示。由图 11 可知,本文提出的 YOLO-LTD 模型在不同光照条件下展现了优越的性能,尤其在光照较差或复杂的环境中表现更为突出。具体来说,在微弱光条件下,YOLO-LTD 的 mAP 达到了 93.70%,相比 YOLOX-s 的 91.04% 和 YOLOv11-n 的 89.94%,分别提高了 2.66 和 3.76 百分点,显示了其在低光照环境中的出色鲁棒性。即使在钠黄光这一具有单一色温和较低对比度的特殊光源下,YOLO-LTD 也能够有效提高检测精度,mAP 为 94.32%,高于 YOLOv11-n 的 92.32%,进一步证明了该模型在挑战性光照条件下的优势。在人工光和自然光条件下,

YOLO-LTD 依然表现优异,mAP 分别为 94.03% 和 94.77%,相较于 YOLOv11-n 和 YOLOX-s 表现均有明显提升。特别是在自然光下,YOLO-LTD 的检测精度仍然保持领先,证明了其在常规光照条件下的稳定性和高效性。该模型通过优化的多尺度特征融合、改进注意力机制以及高效的背景抑制能力,有效应对了光照变化带来的影响,明显提高了目标检测的精度和召回率。综上所述,YOLO-LTD 在不同光照条件下的试验结果表明,其相较于其他对比模型,具备更强的适应性和稳定性,尤其在复杂光照环境下表现尤为突出。

图 12 展示了 YOLOv11-n 与 YOLO-LTD 在人工光、微弱光和钠黄光条件下的检测结果。从图中可以观察到,YOLOv11-n 在人工光和钠黄光条件

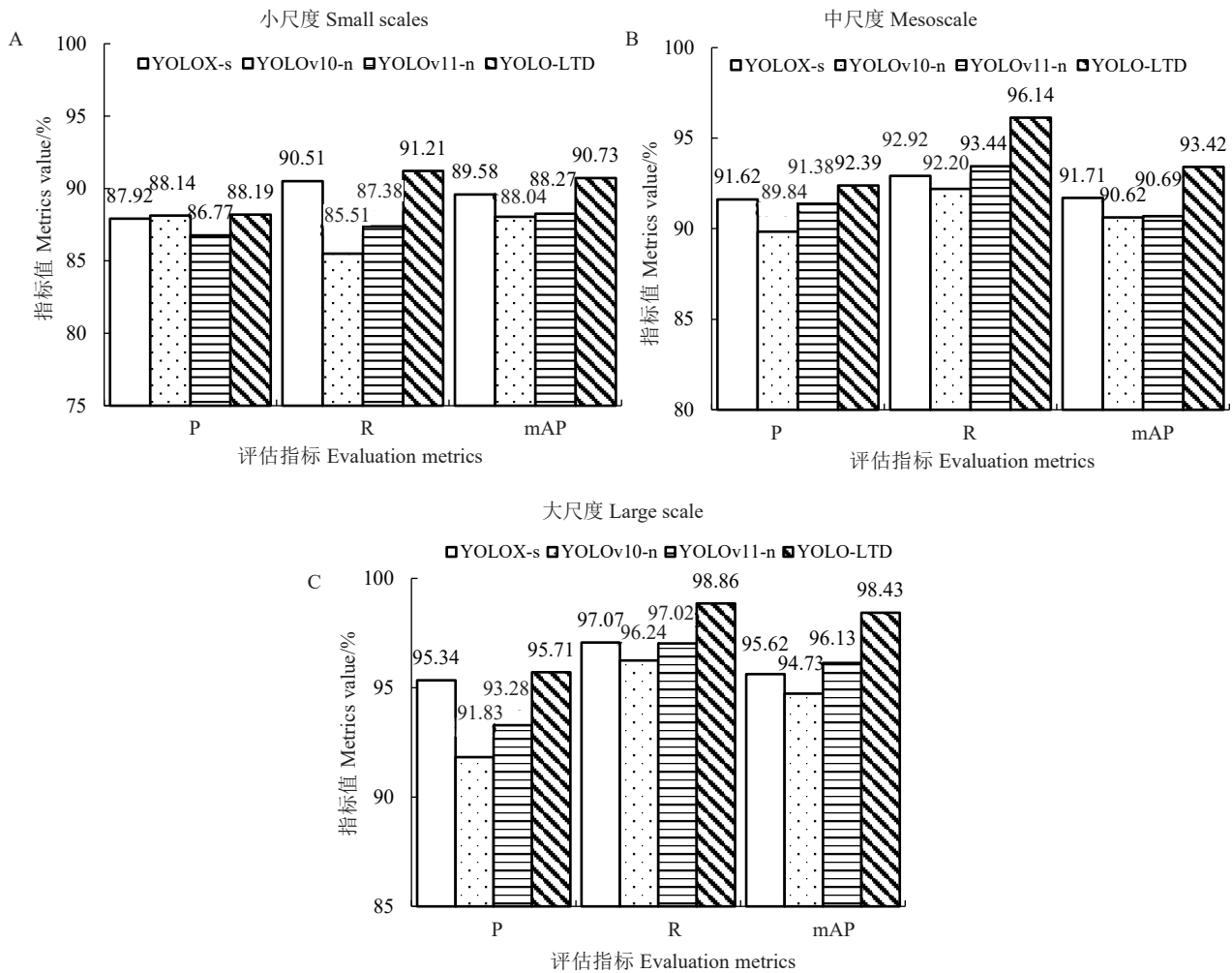


图 10 不同尺度下的对比结果

Fig. 10 Comparison results at different scales

下存在漏检,尤其是在遮挡目标和小尺度果实的识别精度较低时,主要是光照变化导致的图像特征提取不足。相比之下,YOLO-LTD 在相同条件下表现出更高的鲁棒性,成功减少了漏检并提高了对小尺度和遮挡目标的检测精度。这归因于 YOLO-LTD 优化了特征提取和多尺度目标处理,能够在复杂光照环境下更准确地识别番茄果实,展现出更强的适应能力。

3 讨论与结论

在温室番茄果实成熟度检测中,面临着枝叶遮挡、果实重叠、光照变化等环境干扰,以及计算量过大导致检测速度较慢的挑战。针对这些问题,笔者提出了一种改进的 YOLO-LTD 算法。该算法通过引入 criss-cross attention (CCA) 机制,增强了模型对番茄果实图像全局上下文信息的感知能力;在颈部网络中采用 GSConv 替换传统卷积层,实现了模型

的轻量化;同时,通过引入自适应空间特征融合(ASFF)模块,有效缓解了不同尺度特征之间的不一致性。

试验结果表明,在自行采集并标注的番茄数据集上,通过引入 CCA(交叉注意力模块)、GSConv(轻量化卷积模块)和 ASFF(自适应空间特征融合模块)3 种策略,YOLO-LTD 模型的检测精度 P、召回率 R 和平均精度均值(mAP)分别达到 92.07%、95.44%和 94.23%,相比基线模型 YOLOv11 分别提升了 1.60、2.80 和 2.50 百分点。同时,模型的检测时间缩短至 7.21 ms,参数量、计算量(FLOPs)和模型大小分别降至 2.52 M、6.14 G 和 5.18 Mb。在明显提升检测精度的同时,YOLO-LTD 保持了较高的检测效率和较低的模型复杂度,使其能够高效部署于计算资源有限的移动设备,为实际应用提供了更高的可行性。

在不同光照条件下,YOLO-LTD 模型均明显优于对比方法,表现出高度的鲁棒性,表明其在复杂

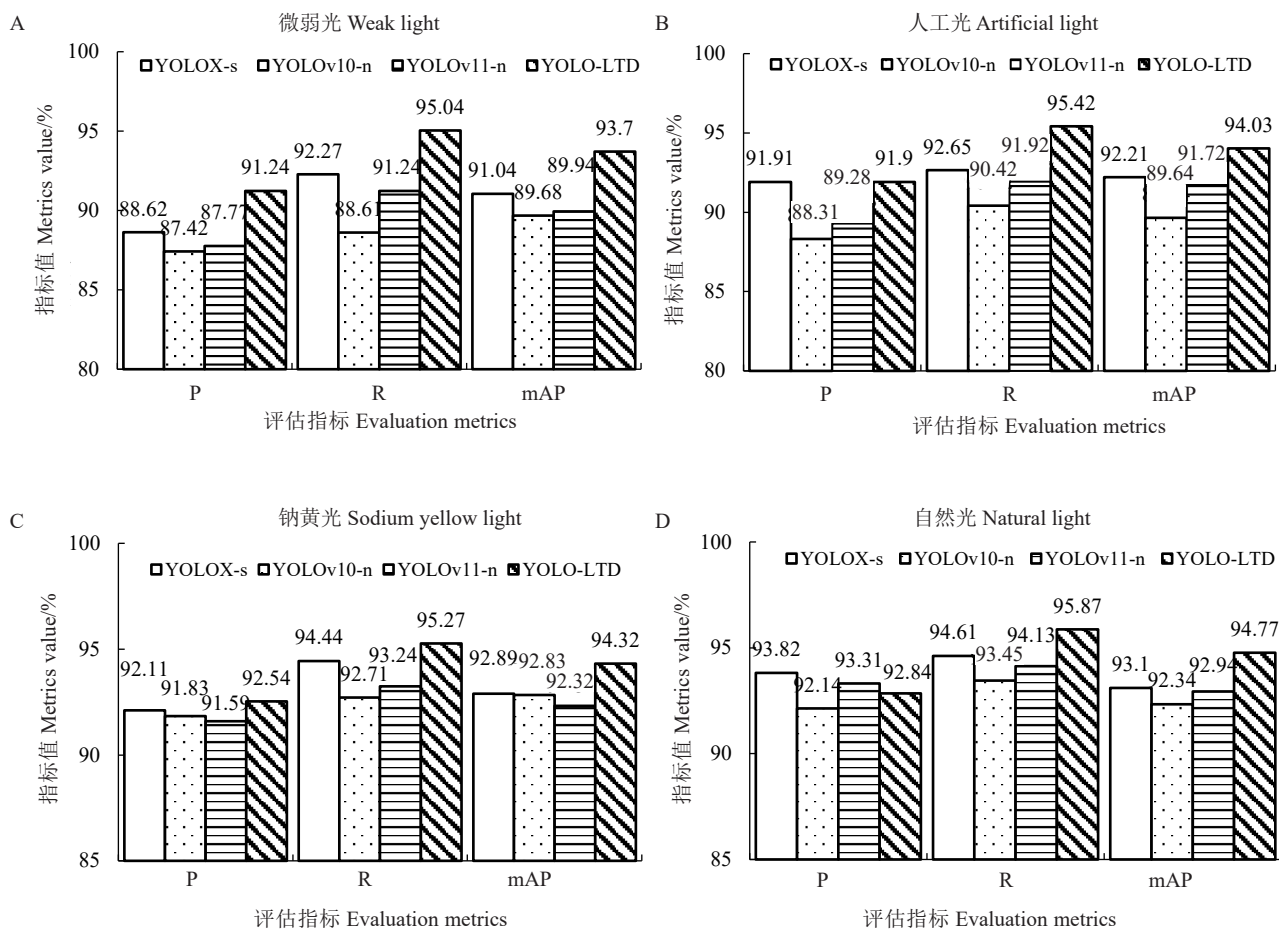


图 11 不同光照下的对比结果

Fig. 11 Comparison results under different illumination conditions

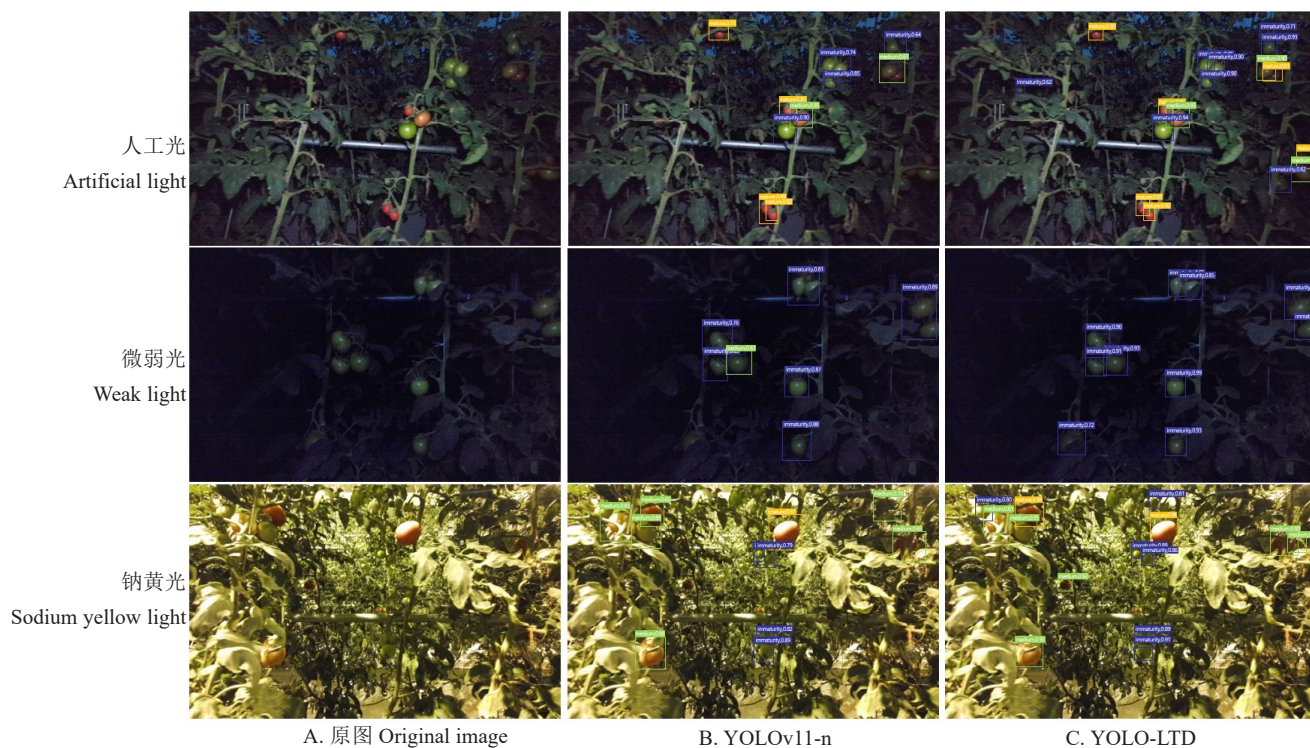


图 12 不同光照下检测结果对比

Fig. 12 Comparison of detection results under different illumination conditions

环境下具有良好的适应能力,具有较高的实际应用价值和广阔的应用前景。然而,尽管 YOLO-LTD 模型在检测未成熟、半成熟和成熟番茄方面表现良好,但对于被叶片遮挡的果实,尤其是未成熟果实,由于其颜色与叶片相似,检测置信度仍需进一步提升。在未来的研究中,计划采用更先进的标注工具或方法,结合遮挡检测和识别技术,以更全面地评估模型在遮挡情况下的表现,提升模型对遮挡场景的鲁棒性,进一步增强模型的性能和提高其实用价值,推动智能农业的发展。

参考文献

- [1] 卢一铭,徐龙水,徐卫红.不同供硼水平对番茄营养和风味品质的影响[J].西南大学学报(自然科学版),2023,45(7):107-122.
- [2] 高倩.基于计算机视觉的番茄识别和成熟度检测研究[D].江苏镇江:江苏大学,2023.
- [3] 于丰华,周传琦,杨鑫,等.日光温室番茄采摘机器人设计与试验[J].农业机械学报,2022,53(1):41-49.
- [4] 孙建桐,孙意凡,赵然,等.基于几何形态学与迭代随机圆的番茄识别方法[J].农业机械学报,2019,50(S1):22-26.
- [5] 冯青春,程伟,杨庆华,等.基于线结构光视觉的番茄重叠果实识别定位方法研究[J].中国农业大学学报,2015,20(4):100-106.
- [6] AYUNINGTYAS D, SURYANI E, WIHARTO W. Identification of tomato maturity based on his color space using the K-nearest neighbour method[C]//2021 International Conference on Artificial Intelligence and Computer Science Technology (ICAIC-ST), 2021: 73-78.
- [7] WANG Z, LING Y M, WANG X L, et al. An improved faster R-CNN model for multi-object tomato maturity detection in complex scenarios[J]. Ecological Informatics, 2022, 72: 101886.
- [8] XU P H, FANG N, LIU N, et al. Visual recognition of cherry tomatoes in plant factory based on improved deep instance segmentation[J]. Computers and Electronics in Agriculture, 2022, 197: 106991.
- [9] LÓPEZ-CORREA J M, MORENO H, RIBEIRO A, et al. Intelligent weed management based on object detection neural networks in tomato crops[J]. Agronomy-Basel, 2022, 12(12):2953.
- [10] 徐婷婷,宋亮,卢学鹤,等.基于 YOLO v7-RA 的火龙果品质与成熟度双指标检测方法[J].农业机械学报,2024,55(7):405-414.
- [11] 刘天真,苑迎春,滕桂法,等.基于改进 YOLO v4 的自然场景下冬枣果实分类识别[J].江苏农业科学,2024,52(1):163-172.
- [12] CHEN W B, LIU M C, ZHAO C J, et al. MTD- YOLO: Multi-task deep convolutional neural network for cherry tomato fruit bunch maturity detection[J]. Computers and Electronics in Agriculture, 2024, 216: 108533.
- [13] 常文龙,谭钰,周立峰,等.基于改进 YOLOv5s 的自然环境下番茄成熟度检测方法[J].江西农业大学学报,2024,46(4):1025-1036.
- [14] YANG G L, WANG J X, NIE Z L, et al. A lightweight YOLOv8 tomato detection algorithm combining feature enhancement and attention[J]. Agronomy, 2023, 13(7):1824.
- [15] GAO J, ZHANG J X, ZHANG F, et al. LACTA: A lightweight and accurate algorithm for cherry tomato detection in unstructured environments[J]. Expert Systems with Applications, 2024, 238: 122073.
- [16] 杨国亮,盛杨杨,洪鑫芳,等.改进 YOLOv8n 的果园番茄目标检测算法[J].计算机工程与应用,2024,60(23):238-248.
- [17] 张俊宁,毕泽洋,闫英,等.基于注意力机制与改进 YOLO 的温室番茄快速识别[J].农业机械学报,2023,54(5):236-243.
- [18] WEI J F, NI L Y, LUO L, et al. GFS-YOLO11: A maturity detection model for multi- variety tomato[J]. Agronomy, 2024, 14(11):2644.
- [19] ZHANG J M, HUANG H T, JIN X K, et al. Siamese visual tracking based on criss-cross attention and improved head network[J]. Multimedia Tools and Applications, 2023, 83(1):1-27.
- [20] 黄英来,牛达伟,侯畅,等.改进 YOLOv8n 的花生品质检测方法[J].计算机工程与应用,2024,60(23):257-267.
- [21] PHIPHITPHATPHAISIT S, SURINTA O. Multi-layer adaptive spatial-temporal feature fusion network for efficient food image recognition[J]. Expert Systems with Applications, 2024, 255: 124834.
- [22] LI Z Y, JIANG X Q, SHUAI L Y, et al. A real-time detection algorithm for sweet cherry fruit maturity based on YOLOX in the natural environment[J]. Agronomy, 2022, 12(10):2482.
- [23] GUAN S T, LIN Y M, LIN G Y, et al. Real-time detection and counting of wheat spikes based on improved YOLOv10[J]. Agronomy, 2024, 14(9):1936.