

基于 YOLO-TM 的番茄采摘机器人感知算法

张文娟¹, 胡海州², 杨聪敏³

(1. 河南科技职业大学机电工程学院 河南周口 466000; 2. 郑州大学电气与信息工程学院 郑州 450001;
3. 河南农业大学信息与管理科学学院 郑州 450046)

摘要:为解决番茄采摘机器人作业中由类间和类内遮挡、光照变化复杂和尺度差异大等因素引起的检测精度不足的问题,提出一种基于 YOLO-TM(YOLO-transformer for tomatoes maturity detection)的番茄成熟度检测模型。首先,在骨干网络引入多头自注意力机制(multi-head self-attention, MHSA)以增强全局特征提取并抑制背景干扰;然后,构建双向自适应特征金字塔网络(bidirectional adaptive feature pyramid network, BAFPN)以提升多尺度特征融合能力;此外,设计 L_{box} 回归损失函数优化对小尺度番茄目标的定位精度。试验结果表明, YOLO-TM 在自制番茄成熟度检测数据集中平均精度均值(mAP)达到 95.3%,推理速度达到 94.6 帧·s⁻¹。与基线模型 YOLOv11 相比,其 mAP 提升 4.2 个百分点,且实地采摘成功率达到 94.0%。与 Faster R-CNN 及其他 YOLO 系列主流模型相比, YOLO-TM 在保持高实时性的同时明显提升了检测精度,展现出在复杂温室环境下的优越适应性,为番茄的自动化精准采摘及智能农业装备的视觉感知系统研发提供了坚实的理论与技术支持。

关键词: 番茄;成熟度;YOLO-TM;损失函数;小尺度目标

中图分类号:S641.2 文献标志码:A 文章编号:1673-2871(2026)04-060-12

Sensing algorithm for tomato harvesting robot based on YOLO-TM

ZHANG Wenjuan¹, HU Haizhou², YANG Congmin³

(1. College of Mechanical and Electrical Engineering, Henan Vocational University of Science and Technology, Zhoukou 466000, Henan, China; 2. School of Electrical and Information Engineering, Zhengzhou University, Zhengzhou 450001, Henan, China; 3. College of Information and Management Science, Henan Agricultural University, Zhengzhou 450046, Henan, China)

Abstract: To address the issue of insufficient detection accuracy in tomato harvesting robot operations caused by factors such as inter-class and intra-class occlusion, complex lighting variations, and larger scale differences, this paper proposes a tomato maturity detection model based on YOLO-TM (YOLO-transformer for tomatoes maturity detection). First, a multi-head self-attention (MHSA) mechanism is introduced into the backbone network to enhance global feature extraction and suppress background interference. Then, a bidirectional adaptive feature pyramid network (BAFPN) is constructed to improve multi-scale feature fusion capability. Furthermore, a L_{box} regression loss function is designed to optimize the localization accuracy of small-scale tomato targets. The experimental results show that YOLO-TM achieves a mean average precision (mAP) of 95.3% and an inference speed of 94.6 frames per second (FPS) on a self-collected tomato maturity detection dataset. Compared with the baseline model YOLOv11, YOLO-TM improves mAP by 4.2 percentage points, and achieves a field picking success rate of 94.0%. Compared with Faster R-CNN and other mainstream YOLO series models, YOLO-TM significantly improves detection accuracy while maintaining high real-time performance, demonstrating superior adaptability in complex greenhouse environments. This study provides a strong theoretical basis and technical support for the automated precise picking of tomato and the development of visual perception systems for intelligent agricultural equipment.

Key words: Tomato; Ripeness; YOLO-TM; Loss function; Small-scale target

番茄作为我国重要的经济作物,广泛应用于食品加工和鲜食领域,其种植面积和产量随着市场需求的增长持续扩大^[1]。然而,传统的番茄采摘方式

仍高度依赖人工操作,面临劳动力短缺、采摘成本高昂及效率低下等问题,严重制约了番茄产业规模的发展^[2]。因此,研究番茄采摘机器人替代人工进

收稿日期:2025-08-17;修回日期:2025-12-17

基金项目:国家自然科学基金(61903341)

作者简介:张文娟,女,讲师,主要从事智能控制及人工智能方向的研究。E-mail:wenjuan199007@163.com

行自动化采摘具有重要意义。采摘机器人通常由视觉感知系统与采摘系统组成^[1],采摘系统依赖视觉感知系统的引导完成果实的采摘任务,因而对果实进行快速、精准的识别与定位已成为实现自动采摘的关键。

随着深度学习技术的发展,基于卷积神经网络(convolutional neural network, CNN)的番茄检测算法因自适应性强及准确性高而被广泛应用。例如,龙洁花等^[4]提出了一种改进Mask R-CNN的不同成熟度番茄果实分割方法,采用跨阶段局部网络(cross stage partial network, CSPNet)与Mask R-CNN网络中的残差网络相融合,实现了番茄果实的检测与分割。Hu等^[5]提出了一种结合IFS和R-CNN的方法,用于检测重叠的成熟番茄,显著提高了检测精度,其水平和垂直中心位置偏移的平均相对误差分别为0.261%和1.179%。Liu等^[6]设计了一种YOLO-Tomato模型,引入密集架构以增强特征复用效能,提高模型的紧凑性和准确性,进而实现对番茄成熟度的检测。上述番茄成熟度检测算法虽取得了一定进展,但在面对光照变化、果实遮挡和重叠等复杂情况时,模型的特征提取能力、抑制背景噪声干扰和特征融合能力仍有待提高。

为提升模型的特征提取能力,研究人员通过优化模型的骨干网络结构对其进行改进。例如,Zu等^[7]提出了一种基于Mask R-CNN的绿色番茄检测与分割方法,采用ResNet作为骨干网络,并通过区域建议网络(RPN)生成感兴趣区域,实现了绿色番茄的检测与分割。周涛等^[8]提出一种基于YOLOv8的菠萝成熟度检测方法,通过将主干网络中的常规卷积层替换为深度可分离卷积(depthwise separable convolution, DSConv),以加强对菠萝特征的提取,从而提高了成熟度检测的准确性。针对背景干扰问题,通常采用引入注意力机制来增强模型对目标关键区域的聚焦能力。例如,田有文等^[9]通过引入卷积注意力机制模块(convolutional block attention module, CBAM),使模型能更好地聚焦于蓝莓果实区域,并将平均精度均值(mAP)提升了3.9%。梁敖等^[10]提出一种改进YOLOv5s算法,通过在骨干网络添加具有全局感受的SE(squeeze-and-excitation)注意力机制,以增强每个通道对草莓果实特征的关注程度,从而提高成熟度检测的准确性。虽然这些注意力机制能够有效增强模型的局部特征提取能力,但在处理全局上下文信息时仍存在一定的局

限。为此,基于Transformer^[11]的自注意力机制逐渐成为提升细粒度特征提取、改善全局依赖建模的有效方案。此外,为了提升多尺度目标的特征融合效果,研究人员通常通过优化颈部网络结构,加强不同尺度特征层之间的信息交流,从而增强语义特征与细节特征的表达能力。例如,苗荣慧等^[12]在特征融合结构中引入自适应空间特征融合(adaptively spatial feature fusion, ASFF)模块,有效利用图像的浅层特征和深层特征,加强特征尺度的不变性,从而提升苹果特征的表达能力,实现了92.5%的检测准确率。宋立航等^[13]将YOLOv5s模型中的特征金字塔结构替换为加权双向特征金字塔网络(bidirectional feature pyramid network, Bi-FPN),有效提升了油桃果实的检测性能。

基于现有的研究成果,针对温室环境中番茄成熟度检测问题,笔者提出一种基于YOLO-TM(YOLO-transformer for tomatoes maturity detection)的番茄成熟度检测模型。YOLO-TM引入多头自注意力机制(multi-head self-attention, MHSA)^[14]增强番茄特征的全局表示能力,构建双向自适应特征金字塔网络(bidirectional adaptive feature pyramid network, BAFPN)提升多尺度特征融合能力,并设计了一种回归损失函数 L_{box} 提高小尺度番茄的定位能力。同时,在番茄成熟度检测数据集上进行试验比较和测试,并将算法嵌入番茄采摘机器人中,结合实际番茄采摘场景进行采摘试验,验证模型在真实环境中的检测准确性和稳定性。

1 材料与方法

1.1 图像采集

试验研究区域选取辽宁省鞍山市台安县新开河镇金马村番茄种植基地,于2024年3—6月期间使用MV-EM202-10GM工业相机在温室环境下对不同番茄植株及其果实在不同成熟阶段进行拍摄。为获取不同光照条件下的图像数据,采集时间覆盖了全天的不同时段,包括08:00—10:00、11:30—13:30与15:00—17:00。最终共收集番茄图像6081幅,每幅图像的分辨率为1280×720像素。在图像采集过程中,部分图像采用跟踪拍摄方式,旨在记录番茄果实在成熟过程中的动态变化细节。图像种类涵盖了顺光、逆光、枝叶遮挡、无枝叶遮挡以及果实重叠等不同场景类型,以全面反映温室环境中番茄植株的多样性与复杂性,部分典型样本图像如图1所示。



注:A为果实顺光场景;B为逆光场景;C为枝叶遮挡场景;D为果实重叠场景。

Note: A represents fruit scene under frontlighting conditions; B represents fruit scene under backlighting conditions; C represents fruit scene partially occluded by branches and leaves; D represents overlapping fruits scene.

图1 番茄图像样本示例

Fig. 1 Tomatoes image sample example

1.2 数据集构建

本研究依据国家标准 GH/T 1193—2021 对番茄的成熟度进行等级划分,将其分为未成熟(raw)、半成熟(medium raw)和成熟(ripe)三类^[15]。为提升标注数据的准确性与一致性,采用开源标注工具 LabelImg 对图像中的番茄目标进行最小外接矩形标注,并将每幅图像的位置信息与类别标签以 XML 格式保存。完成标注后,数据集按照 8:1:1 的比例随机划分为训练集、验证集和测试集,分别包含 4865、608 和 608 幅图像。各类别标签的分布情况详见表 1。

表1 各类标签分布
Table 1 Label distribution

数据集 Data set	未成熟 Raw	半成熟 Medium raw	成熟 Ripe	总计 Total
训练集 Training set	23 254	14 524	32 121	69 899
验证集 Verification set	2 915	1 854	4 051	8 820
测试集 Test set	2 907	1 823	4 017	8 747

1.3 YOLO-TM 番茄成熟度检测算法

YOLOv11 作为 Ultralytics 公司推出的目标检

测模型^[16],在结构设计上较前代模型进行了多项关键创新。首先,该模型以 C3k2 模块替代原有的 C2f 模块,通过更高效的卷积操作实现特征分区,从而有效提升了特征表示能力。此外,在 SPPF 模块之后引入 C2PSA 模块,进一步增强了模型对复杂背景和多变场景的适应性。其颈部网络采用 PANet 结构,实现浅层与深层特征的双向融合,有助于提升对不同尺度目标的检测效果。为了增强上下文信息的建模能力,YOLOv11 引入深度可分离卷积操作,不仅提高了检测精度,同时显著降低了计算冗余。

得益于上述改进,YOLOv11 在目标检测任务中相较于前代模型表现出更优的精度与实时性,特别适用于对检测速度与准确度要求较高的农业场景。然而,在番茄成熟度识别任务中,由于受限于枝叶遮挡、果实重叠以及目标尺度多样性等复杂因素,YOLOv11 的检测性能仍有进一步提升的空间。为解决这一问题,提出一种改进的 YOLO-TM 模型,旨在实现更高效、精准的番茄成熟度检测。其整体网络结构如图 2 所示。

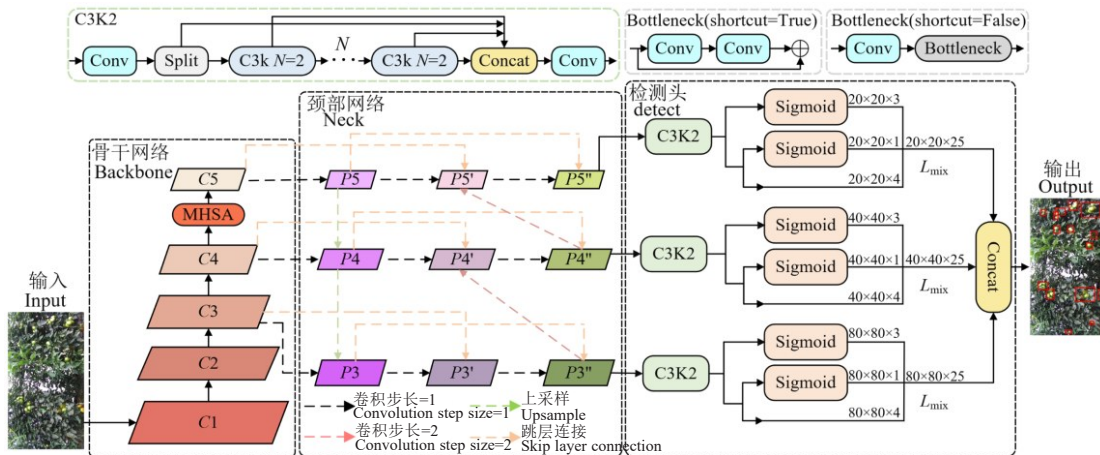


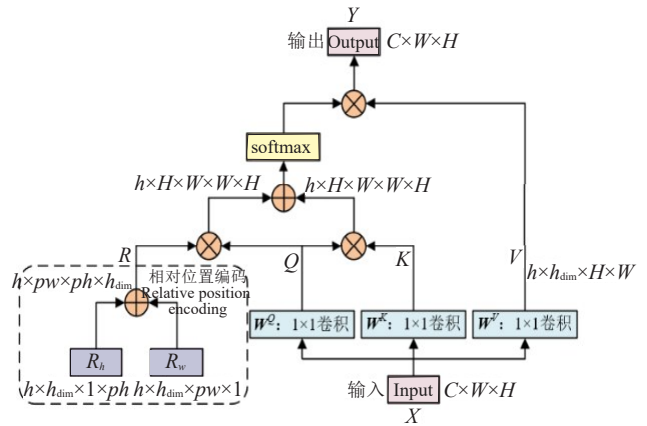
图2 YOLO-TM 网络结构

Fig. 2 YOLO-TM network structure

具体而言,在骨干网络引入多头自注意力机制MHSA,通过提取输入特征图的全局上下文信息,聚焦于图像中的关键区域,从而提升模型对番茄成熟度特征的识别能力;在颈部网络设计一种双向自适应特征金字塔网络BAFPN,使模型能够在多个尺度上提取和融合细粒度特征,从而避免小尺度番茄信息的丢失;此外,提出一种回归损失函数 L_{box} ,通过对小尺度目标预测框回归误差的加权调整,进一步提高小尺度番茄的定位精度。

1.3.1 多头自注意力机制 在番茄成熟度检测中,果实与叶片在颜色与纹理上相近,且光照变化显著,易导致特征提取阶段出现目标与背景混淆,从而影响识别与分类准确性。通道或空间注意机制如CBAM^[17]、BAM^[18]主要在单一维度上进行重标定,对跨区域的全局依赖建模能力有限,复杂背景下的上下文判别力仍然不足。为增强骨干网络对全局上下文与长程依赖的表征能力,在骨干网络末端引入多头自注意力机制MHSA。MHSA通过并行注意头在不同子空间内对特征进行全局交互,并结合相对位置编码刻画区域间的空间关系,从而更有效地突出番茄果实相关区域并抑制叶片与光照带来的干扰。MHSA的具体计算过程与实现细节参考自注意力机制相关工作^[19],网络结构如图3所示。

具体而言,MHSA通过构建多个注意力头来捕捉输入番茄图像中不同区域的特征信息,以便更准确地判断番茄的成熟度。首先,将输入特征图 $X \in R^{C \times W \times H}$ 划分为若干个图像区域(patch),并对每个区域提取相应的特征向量。每个图像区域的特征向量通过3个特征表示来进行编码:查询(Q)、键(K)和值(V)。其中,查询向量 Q 用于表示当前图像区域对其他区域的关注程度,键向量 K 表示当前区域与其他区域之间的关系强度,而值向量 V 则是对每个图像区域的成熟度进行编码后的特征表达。在MHSA模块中,每个图像区域的特征向量在经过全局特征融合之后,所输出的特征向量与其他所有区域的 V 向量密切相关。MHSA通过计算不同图像区域间特征向量的相似度,得到当前图像区域对其他区域的注意力分数,从而捕捉图像中各区域间的相对关系与上下文信息。MHSA不仅考虑不同区域特征内容之间的相似度,还引入每个区域的相对位置编码 R_w 和 R_h 分别表示图像区域在宽度和高度方向上的相对位置关系。通过在计算注意力分数时融合位置编码信息,MHSA进一步强化不同区域



注:⊕表示逐元素加法;⊗表示逐元素乘法; h 表示多头注意力模块中头的数量; h_{dim} 是每个头分到的特征维度; pw 和 ph 分别为相对位置编码中矩阵的宽度和高度; R 为联合相对位置矩阵; R_w 和 R_h 分别为宽高方向的相对位置编码矩阵; Q 、 K 和 V 分别表示查询向量、键向量和值向量。

Note: ⊕ denotes element-wise addition; ⊗ denotes element-wise multiplication; h represents the number of heads in the multi-head attention module; h_{dim} refers to the feature dimension allocated to each head; pw and ph indicate the width and height of the matrices in the relative position encoding, respectively; R_w and R_h represent the relative position encoding matrices for the width and height directions; R represents the joint relative position matrix; Q , K , and V denote the query, key, and value vectors.

图3 多头自注意力网络结构图
Fig. 3 Multi-head self-attention network structure diagram

之间的空间关系,以提升模型对番茄成熟度的识别能力。MHSA计算公式如(1~3)所示:

$$\begin{cases} Q = X \times W^Q \\ K = X \times W^K \\ V = X \times W^V \end{cases} \quad (1)$$

$$R = R_w \times R_h \quad (2)$$

$$\text{Attention}(Q, K, V, R) = \text{softmax}\left(\frac{Q \times K^T + Q \times R^T}{\sqrt{d_k}}\right)^T \times V \quad (3)$$

式中, W^Q 、 W^K 、 W^V 分别表示 Q 、 K 、 V 三个特征矩阵所用得到权重矩阵; $\text{Attention}(\cdot)$ 表示注意力输出结果,用于根据查询向量 Q 与键向量 K 的相关性分配权重,并对值向量 V 进行加权聚合,从而获得融合全局上下文信息的特征表示; R_w 和 R_h 分别为宽高方向的相对位置编码矩阵; R 为联合相对位置矩阵; d_k 为查询和键向量的维度; softmax 为归一化函数; T 表示矩阵转置操作。

通过构建多个注意力头,能够在多个特征空间内进行特征学习,从而获得更加丰富和多样化的特

征表示,使模型能够有效适应不同尺寸和复杂度的番茄图像数据。此外,利用MHSA模块提取特征时,充分考虑图像中的全局上下文信息,能够有效捕捉图像中番茄果实之间的长距离依赖关系,拓宽模型的全局视野。

1.3.2 双向自适应特征金字塔网络 番茄果实不同生长阶段存在明显尺度差异,小尺度果实浅层特征中易因细节信息弱与信息传递不足而被忽略,从而限制检测性能。YOLOv11 颈部采用PANet^[20]进行多尺度融合,但在高层语义与低层细节的权衡与传递方面仍存在不足,导致小目标的局部结构信息容易衰减。为提升多尺度表征能力,在颈部网络构建双向自适应特征金字塔网络BAFPN。BAFPN在自上而下与自下而上的双向路径上引入自适应加权融合与跨层连接,强化不同尺度特征之

间的语义交互与细节保真传递,从而增强小尺度番茄果实的可分性与可检测性。PANet与BAFPN的结构对比如图4所示。

BAFPN引入双向传递与密集连接机制,通过自下而上的特征金字塔实现高层与低层特征的有效聚合,并结合自上而下的信息流动强化特征图之间的语义交互。以图4-B中 C_4 特征层为例,其特征融合过程具体如下:首先,将当前层特征图 C_4 和经过上采样操作得到的第5层特征图 P_5 进行融合,并通过卷积操作生成特征图 P_4 ,以在保持语义信息的同时增强特征表达能力,如式(4)所示:

$$P_4 = \text{Conv} \left(\frac{w_4^1 \times C_4 + w_5^3 \times P_5 \times \text{Upsample}(P_5)}{w_4^1 + w_5^3 + e} \right) \quad (4)$$

式中, $w_j^i \leq 1$ 表示可学习权值,使用调节系数

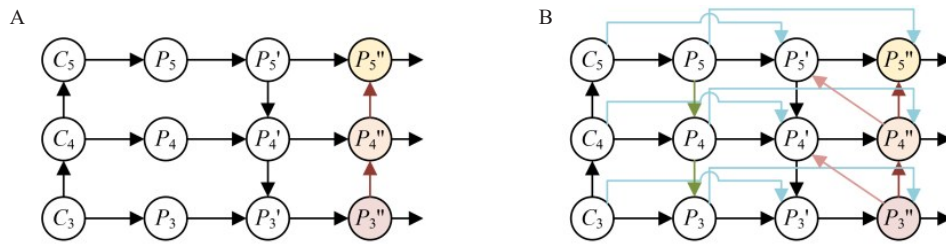


图4 PANet (A)和BAFPN (B)网络结构图

Fig. 4 The networks structure diagram of PANet (A) and BAFPN (B)

$\varepsilon=0.0001$ 来避免数值不稳定性,Conv表示卷积操作,Upsample表示上采样操作。

随后,通过跳层连接整合当前层特征图 C_4 、经过卷积处理的特征图 P_4 及第3层卷积处理后的 P_3' ,实现不同尺度特征之间的深度交互,如式(5)所示:

$$P_4' = \text{Conv} \left(\frac{w_4^1 \times C_4 + w_4^3 \times P_4 + w_3^1 \times \text{Conv}(P_3')}{w_4^1 + w_4^3 + w_3^1 + e} \right) \quad (5)$$

式中, P_4' 为融合后的第4层特征图, P_3' 为下层经过卷积后的特征图。

最后,结合 P_4 层和 P_4' 层特征图的多尺度信息,通过卷积操作生成特征图 P_4'' ,最大限度地提取多尺度特征图的细粒度信息,确保充分的特征融合与表达,如式(6)所示:

$$P_4'' = \text{Conv} \left(\frac{w_4^3 \times P_4 + w_4^4 \times P_4'}{w_4^3 + w_4^4 + e} \right) \quad (6)$$

式中, P_4'' 表示经过多次特征融合与卷积操作

后,生成的第4层最终融合特征图。

BAFPN通过扩展尺度层次并引入跳跃连接,构建一个多尺度特征的密集连接路径,进而提升对番茄果实不同尺度下目标的检测能力。尤其在小尺度目标的检测中,能够有效保留并强化局部细节信息,从而提升整体检测精度。

1.3.3 改进损失函数 YOLOv11采用完全交并比损失(complete intersection over union, CIoU)^[21]进行边界框回归优化,该损失在IoU基础上综合考虑中心距离与长宽比一致性,有助于提升定位精度。然而,CIoU对边界框位置偏差较为敏感,在小尺度目标出现较大偏移时,度量过程难以保持尺度不变,易在标签分配与回归优化阶段引入误差,进而影响小目标检测性能。为此,提出一种新的回归损失函数 L_{box} ,将归一化瓦瑟斯坦距离(normalized wasserstein distance, NWD)^[22]引入回归损失函数中。NWD损失通过将目标的边界框视为二维高斯分布,并根据真实框与预测框的分布相似度来刻画

它们之间的几何分布差异。该方法不依赖于目标框的重叠情况,能够更加精确地衡量小目标的几何差异,从而提升小尺度番茄检测任务中的定位精度。

NWD 损失函数计算过程如公式(7~8)所示:

$$L_{NWD} = \exp\left(-\frac{\sqrt{W_2^2(N_a, N_b)}}{t}\right); \quad (7)$$

$$W_2^2(N_a, N_b) = \left[[cx_a, cy_a, \frac{w_a}{2}, \frac{h_a}{2}]^T, [cx_b, cy_b, \frac{w_b}{2}, \frac{h_b}{2}]^T \right]^2 \quad (8)$$

式中, N_a 和 N_b 是由真实框 $A=(cx_a, cy_a, w_a, h_a)$ 和预测框 $B=(cx_b, cy_b, w_b, h_b)$ 建模的高斯分布, $W_2^2(\cdot)$ 表示 N_a 和 N_b 之间的瓦瑟斯坦距离,用于度量预测框与真实框的高斯分布距离, τ 为超参数。

结合 NWD 和 CIoU 损失,其回归损失函数 L_{box} 定义为:

$$L_{box} = \lambda L_{CIoU} + (1-\lambda)L_{NWD} \quad (9)$$

式中, λ 为 NWD 损失和 CIoU 损失之间的权重系数。

通过将 CIoU 和 NWD 相结合,并引入权重参数调控二者的贡献比例,能够更加精确地度量预测框与真实框之间的相似性,进而提升小尺度番茄的定位精度。

1.4 试验环境介绍

本文涉及的所有试验均在相同硬件平台上进行,具体的硬件配置和软件环境如表 2 所示。

在训练过程中将输入图像尺寸设置为 640×640 ,初始学习率设置为 0.01,最小学习率为 0.000 1,训练轮次为 200,批量大小为 8。动量参数和权重衰减分别设置为 0.937 和 0.000 5。优化器选用随机梯度下降优化器,并结合自适应梯度调整策略(adaptive gradient adjustment)动态调整学习率,以增强模型的收敛效果与训练稳定性。

表 2 试验硬件配置和软件环境

Table 2 Experiment hardware configuration and software environment

硬件 Hardware	配置参数 Parameters	软件 Software	配置参数 Parameters
CPU	Intel Core i9-10850K 10-core	操作系统 Operating system	Ubuntu 20.04
GPU	NVIDIA GeForce RTX 3080 10G	PyTorch	1.8.1
内存 Memory	32 GB	CUDA	11.1
硬盘 Hard disk	1 TB	CUDNN	8.0.5

1.5 评价指标

为全面评估模型在番茄成熟度检测任务中的性能及其效率,采用准确率(precision, P)、召回率(recall, R)、平均精度均值(mean average precision, mAP)对各模型性能进行定量评价。同时,利用检测速度(frames per second, FPS)作为衡量模型实时性能的重要指标。此外,还通过参数量(number of parameters, Params)和浮点运算次数(floating point operations, FLOPs)评估模型的计算复杂度和存储需求。 P 、 R 和 mAP 的计算公式如(10~13)所示:

$$P = \frac{TP}{TP + FN} \times 100\% \quad ; \quad (10)$$

$$R = \frac{TP}{TP + FN} \times 100\% \quad ; \quad (11)$$

$$AP = \int_0^1 P(R) dR \quad ; \quad (12)$$

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (13)$$

式中, P 表示精确率; TP 表示预测为正例的正样本, FN 表示预测为负例的正样本。 R 表示召回率; FN 表示预测为负例的正样本。 AP 表示每个类别的平均精度, mAP 表示所有类 AP 的平均值, N 表示类别总数, AP_i 表示第 i 个类别的平均精度。

2 结果与分析

2.1 损失函数对比

在 L_{box} 回归损失函数中, λ 作为 CIoU 和 NWD 损失之间的权重系数,其选取直接影响两者的相对贡献。为确定 λ 的最优值,进行了不同 λ 值下的对比试验。试验中 λ 的取值范围设置为 0.1~0.6,并对模型的 P 、 R 和 mAP 进行系统评估,试验结果如图 5 所示。

从图 5 可以看出,随着 λ 值的增大,模型在 P 、 R 和 mAP 上的表现逐渐提升,直至 $\lambda=0.4$ 达到最佳效

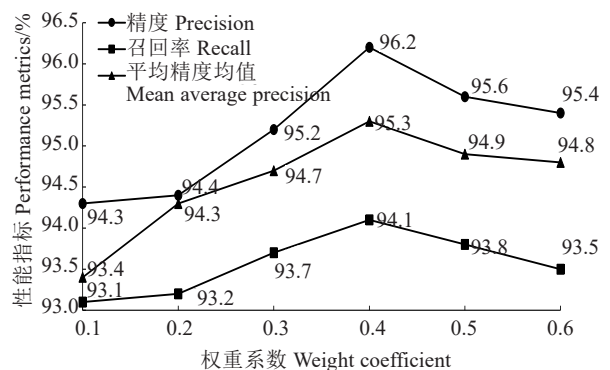


图 5 模型性能与 λ 值的关系

Fig. 5 Relationship between model performance and λ value

果。此时,模型的 P 为 96.2%, R 为 94.1%, mAP 为 95.3%。当 λ 值进一步增大时,模型的精度和 mAP 均有所下降,表明 $\lambda=0.4$ 时 CIoU 与 NWD 之间的平衡已达到了最优状态。基于此,在后续的试验中固定 $\lambda=0.4$ 作为 CIoU 和 NWD 损失的权重系数。

此外,为验证 CIoU+NWD 损失函数的有效性,针对目标检测模型的回归损失部分进行系统的对比试验。首先,选取 GIoU (generalized intersection over union)^[23]、DIoU (distance intersection over union)^[24]、SIoU (segmented intersection over union)^[25] 与原始 CIoU 损失进行对比,以评估它们在番茄成熟度检测任务中的性能差异。在上述基础上,进一步将归一化瓦瑟斯坦距离(NWD)引入各回归损失

函数,并对这些损失函数的检测性能进行全面对比分析,结果如表 3 所示。

由表 3 可知,SIoU 在所有单一回归损失函数中表现最佳,其 mAP 达到 94.8%。当归一化瓦瑟斯坦距离(NWD)与各回归损失函数结合后,所有损失函数的性能均实现明显提升。其中,CIoU+NWD 组合在各项指标上整体表现最佳,其精度 P 和召回率 R 分别达到了 96.2%和 94.1%,均为最高值,同时 mAP 达到 95.3%,展现出更优的综合性能。相比之下,SIoU+NWD 的 mAP 为 95.1%,虽表现优异,但仍略低于 CIoU+NWD。综上,CIoU 与 NWD 的结合更契合回归任务需求,能够准确地刻画预测框与真实框之间的几何关系,适用于番茄成熟度检测。

表 3 不同损失函数对比试验
Table 3 Comparative experiment of different loss functions

损失函数 Loss function	准确率 $P/\%$	召回率 $R/\%$	平均精度均值 $mAP/\%$	损失函数 Loss function	准确率 $P/\%$	召回率 $R/\%$	平均精度均值 $mAP/\%$
GIoU	92.3	90.6	91.7	GIoU+NWD	93.3	91.4	92.5
DIoU	94.5	92.3	93.8	DIoU+NWD	95.6	93.4	94.7
SIoU	95.7	93.6	94.8	SIoU+NWD	95.8	93.7	95.1
CIoU	95.6	93.7	94.7	CIoU+NWD	96.2	94.1	95.3

2.2 模型训练收敛性分析

为深入评估 YOLO-TM 算法在特征空间中的优化效率及稳定性,对基准模型 YOLOv11 与改进模型 YOLO-TM 在训练全周期内的总损失值演变轨迹进行了对比分析。训练集与验证集的损失响应曲线如图 6 所示。

由图 6 可知,改进模型 YOLO-TM 在收敛速率与最终拟合精度方面均明显优于基准模型 YOLOv11。在迭代初始阶段,YOLO-TM 的损失曲线呈现出陡峭的下降趋势,迅速收敛至低损失区间,证明 L_{box} 回归损失函数有效解决了微小目标的梯度弥散问题,提升了参数寻优效率。当迭代轮次超过 50 次进入平稳期后,基准模型的验证损失在降至

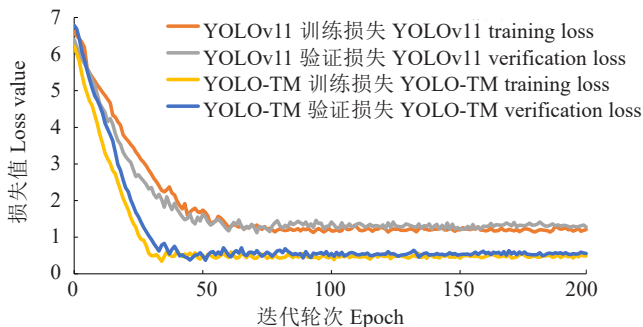


图 6 YOLOv11 与 YOLO-TM 训练及验证损失曲线对比
Fig. 6 Comparison of training and validation loss curves between YOLOv11 and YOLO-TM

1.1 附近时出现收敛停滞与局部震荡,反映出其在处理复杂背景特征时存在提取瓶颈。相比之下,YOLO-TM 的验证损失进一步下降并稳定在 0.3 左右,这一明显优势证实了 MHSA 全局注意力机制与 BAFPN 特征金字塔有效增强了模型对光照变化及枝叶遮挡等困难样本的判别能力。此外,训练与验证损失差值始终维持在 0.1 以内,表明模型具备良好的泛化性能且未出现过拟合。综上所述,训练过程的动态演变数据充分验证了 YOLO-TM 算法改进策略的有效性,为高精度番茄成熟度检测提供了坚实的试验依据。

2.3 消融试验

本文通过消融试验分析 MHSA 注意力机制、BAFPN 模块和 L_{box} 损失函数对检测性能的影响。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。SSPENet 模型的消融试验结果如表 4 所示,其中“√”表示启用该模块,“×”表示未启用该模块。

由表 4 可以看出,各改进模块对模型性能方面均具有提升作用。单独引入 MHSA 注意力机制后,模型的 mAP 提升至 92.4%,参数量略微增加至 2.7 M, FLOPs 增至 6.8 G,表现出明显的精度提升效果。仅嵌入 BAFPN 模块后, mAP 达到 92.1%,并在计算效率方面保持较优表现,展现出良好的适用性。采

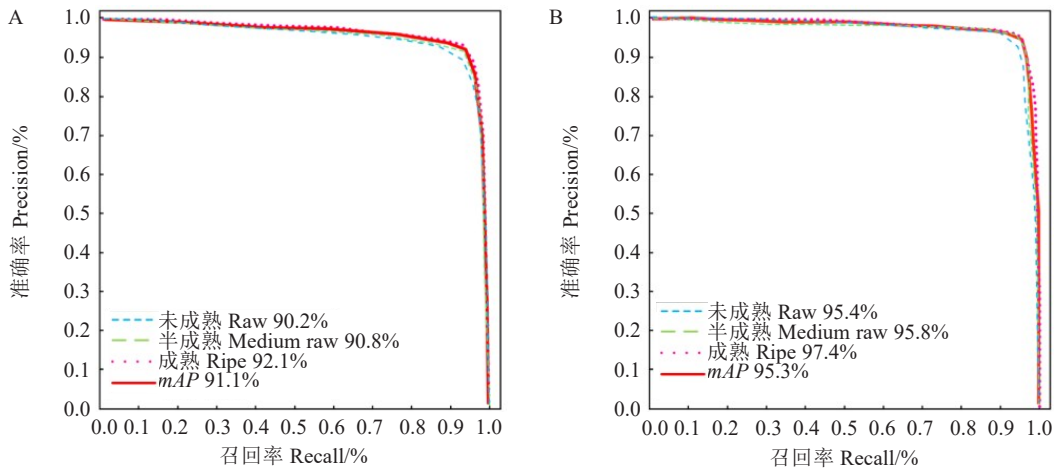
表4 不同损失函数消融试验
Table 4 Comparative experiment of different loss functions

YOLOv11	MHSA	BAFPN	L_{box}	平均精度 $AP/\%$			召回率 $R/\%$	平均精度均值 $mAP/\%$	检测速度 $FPS/(帧 \cdot s^{-1})$	参数量 $Params/M$	浮点运算次数 $FLOPs/G$
				未成熟 Raw	半成熟 Medium raw	成熟 Ripe					
√	×	×	×	90.2	90.8	92.1	89.8	91.1	99.7	2.6	6.5
√	√	×	×	92.0	92.2	92.9	91.6	92.4	98.6	2.7	6.8
√	×	√	×	91.6	91.7	92.3	91.2	92.1	97.9	2.8	7.4
√	×	×	√	91.5	91.6	92.6	90.9	91.9	99.4	2.6	6.5
√	√	√	×	94.9	94.8	96.4	93.2	94.8	96.6	2.9	8.0
√	√	√	√	95.4	95.8	97.4	94.1	95.3	94.6	2.9	8.0

用 L_{box} 损失函数后, mAP 提升至 91.9%, 进一步验证了其在优化边界框回归任务中的有效性。当 MHSA 和 BAFPN 模块联合使用时, mAP 进一步提升至 94.8%, 展现了注意力机制与局部金字塔模块在特征提取和表达上的协同作用。最终, 结合三种改进策略后模型的 mAP 提升了 4.2 个百分点, 而参数量仅增加 0.3 M, 表明联合使用多种改进模块能够在显著提升检测精度的同时保持良好的实时性能, 从

而实现检测效率与准确性的平衡。图 7 展示了改进前后模型各类别及整体 mAP 的 $P-R$ 曲线。

通常情况下, $P-R$ 曲线越接近右上角, 则表明识别方法的性能越好。由图 7-A 与图 7-B 的面积对比可以看出, 所提出的 YOLO-TM 在各类别上的检测性能均有所提升, 其 $P-R$ 曲线的包围面积较 YOLOv11 明显增大, 进一步验证了 YOLO-TM 在检测精度和整体性能上的优越性。



注: A 为 YOLOv11 模型 $P-R$ 曲线; B 为 YOLO-TM 模型 $P-R$ 曲线。

Note: A. Precision-Recall curve of the YOLOv11 model; B. Precision-Recall curve of the YOLO-TM model.

图 7 改进前后模型在番茄果实成熟度数据集上的 $P-R$ 曲线图

Fig. 7 $P-R$ curves of the model before and after improvement on tomatoes fruit maturity data set

2.4 不同模型对比

为全面评估改进算法的优越性, 选取当前主流目标检测算法, 包括 Faster R-CNN、RT-DETR^[26]、Deformable-DETR^[27]、YOLOv9、YOLOv10 和 YOLOv11, 在番茄成熟度数据集上进行统一标准下的横向对比试验。对比试验结果如表 5 所示。

通过对比试验结果可知, Faster R-CNN 在番茄成熟度检测任务中表现出较明显的不足, 其 mAP 仅为 85.8%, FPS 为 6.8 帧·s⁻¹, 难以满足实时性和精度

要求。RT-DETR 和 Deformable-DETR 在检测性能上优于 Faster R-CNN, mAP 为 90.6% 和 91.2%, 但二者计算资源需求较大。YOLOv9 以参数量 2.0 M 和 7.7 G 的 FLOPs 实现了 90.2% mAP 和 100.8 帧·s⁻¹ 的检测速度, 展现出良好的实时性; YOLOv10 将 mAP 提升至 90.3%, 但检测速率略有下降。相比之下, YOLO-TM 模型达到了 95.3% 的 mAP , 相较于上述算法分别提升 9.5、4.7、4.1、5.1、5.0 和 4.2 个百分点, 尽管 YOLO-TM 的检测速率低于其他 YOLO 系

表5 对比试验结果
Table 5 Results of contrast experiment

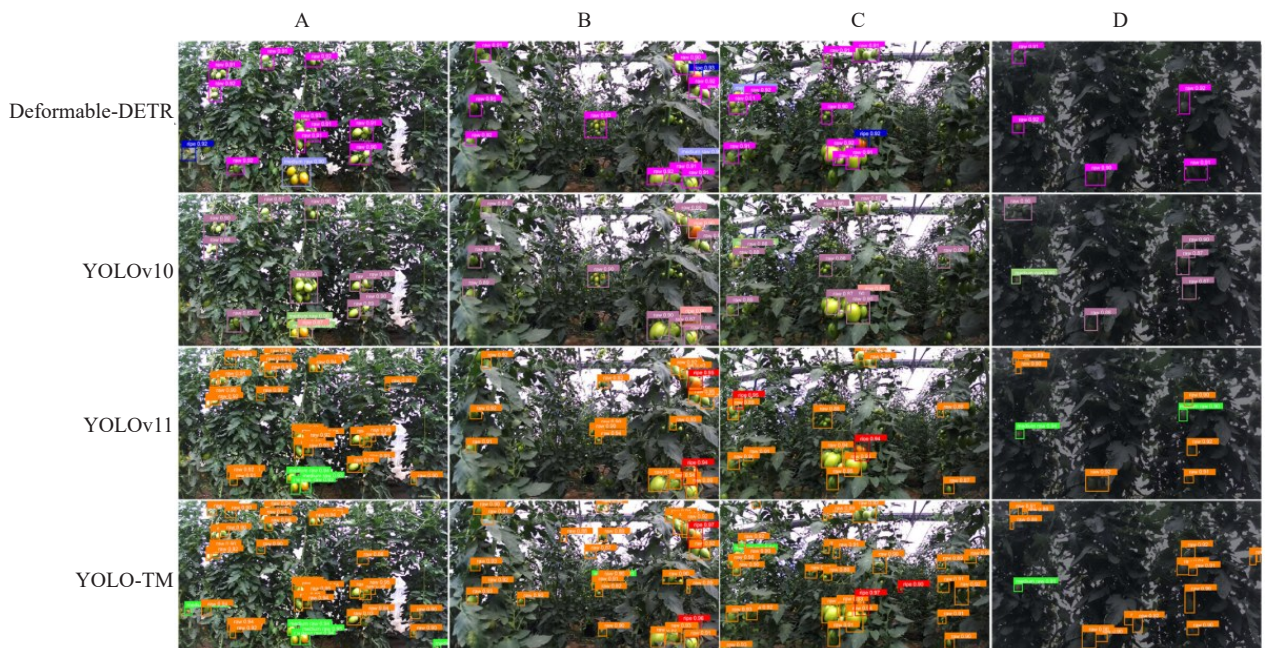
算法 Algorithm	准确率 Precision/%	召回率 Recall/%	平均精度均值 <i>mAP</i> /%	检测速度 FPS/(帧·s ⁻¹)	参数量 Params/M	浮点运算次数 FLOPs/G
Faster R-CNN	86.4	85.1	85.8	6.8	102.7	168.4
RT-DETR	89.8	90.2	90.6	94.2	20.0	99.1
Deformable-DETR	90.9	90.8	91.2	68.6	40.0	173.0
YOLOv9	90.3	89.6	90.2	100.8	2.0	7.7
YOLOv10	90.8	89.4	90.3	100.4	2.3	6.7
YOLOv11	91.0	89.9	91.1	99.7	2.6	6.5
YOLO-TM	96.2	94.1	95.3	94.6	2.9	8.0

列算法,但在保持高实时性的同时,实现了最高的检测精度,展现出优越的综合性能。

为直观展示算法改进前后的检测效果,选取检测精度较高的4种模型对不同场景下的番茄成熟度图像进行对比测试,检测效果如图8所示。

由图8可知,Deformable-DETR和YOLOv10模型在检测任务中误检现象较为严重。而在对比本研究的基准模型YOLOv11与改进后的YOLO-TM时,差异尤为明显。首先,在图8-A和8-C所示的复杂背景与果实重叠场景中,YOLOv11虽然能够有效检测出特征明显的完整目标,但对于受枝叶严重遮挡的小尺度番茄出现了明显的漏检。

这是因为YOLOv11的卷积结构在处理深层语义时容易丢失微弱的边缘特征信息。相比之下,YOLO-TM得益于BAFPN结构,通过双向特征融合有效增强了对弱特征目标的表达能力,成功捕获了这些被基准模型遗漏的高遮挡率果实。其次,针对图8-D的逆光场景,虽然YOLOv11能够识别目标,但受光照分布不均的影响,其检测框的边界回归精度不足,难以紧密贴合果实轮廓。而YOLO-TM引入的MHSA多头自注意力机制能够从全局角度捕捉上下文信息,降低了模型对局部光照强度变化的敏感度。从可视化结果看,YOLO-TM在逆光环境下不仅避免了漏检,且生成的边界框拟合度更高,置



注:A为顺光及果实重叠场景;B为顺光及枝叶遮挡场景;C为果实重叠及枝叶遮挡场景;D为逆光及枝叶遮挡场景。

Note: A represents a front-lit scene with overlapping fruits; B represents a front-lit scene with occlusion from branches and leaves; C represents a scene with both overlapping fruits and occlusion from branches and leaves; D represents a backlit scene with occlusion from branches and leaves.

图8 不同算法检测结果的可视化对比
Fig. 8 Visual comparison of detection results of different algorithms

信度也普遍优于 YOLOv11。综上所述, YOLO-TM 模型通过充分提取并融合多尺度特征, 精准聚焦于番茄的关键区域, 在复杂光照和遮挡条件下均表现出了优于主流算法的鲁棒性。

2.5 混淆矩阵对比

为评估不同模型在番茄成熟度检测任务中对各成熟度类别的分类性能, 采用混淆矩阵对模型的误检情况进行分析。具体而言, 将训练后的 Faster R-CNN、RT-DETR、Deformable-DETR、YOLOv9、YOLOv10、YOLOv11 和 YOLO-TM 模型应用于番茄成熟度数据集, 对不同成熟度样本进行预测, 并生成对应的混淆矩阵, 如图 9 所示。在混淆矩阵中, 横坐标表示模型预测的成熟度类别标签, 纵坐标表示数据的真实标签。

由图 9 可知, YOLO-TM 在 Ripe 类别上的准确率达到 98.0%, 相较于其他算法分别提高了 9.0、7.0、5.0、6.0、3.0 和 2.0 个百分点。表明 YOLO-TM

对 Ripe 类别样本的检测更加精准, 有效减少了漏检与误检。在 raw 和 medium raw 类别上, YOLO-TM 的检测准确率均达到 100%, 展现出卓越的分类均衡性。Faster R-CNN 在 Ripe 类别中有 9.0% 的样本被误分类为 raw, 而 RT-DETR 和 Deformable-DETR 在 Medium raw 类别上的误检率分别为 7.0% 和 4.0%。综上, YOLO-TM 在番茄成熟度检测任务中表现优异, 以更高的检测精度和稳定性有效减少漏检与误检, 在各类别上的分类性能均衡且优于其他模型。这一结果充分证明了 YOLO-TM 在番茄成熟度检测任务中的优势及其较大的应用潜力。

2.6 采摘试验

为验证 YOLO-TM 算法在采摘机器人上的可行性, 将 YOLO-TM 嵌入采摘机器人并进行采摘试验。采摘机器人系统由移动平台、机械臂、末端执行器、深度相机和控制器组成。机械臂选用 KUKA Robotics 的 LBR iiwa 四轴机械臂, 具有最大负

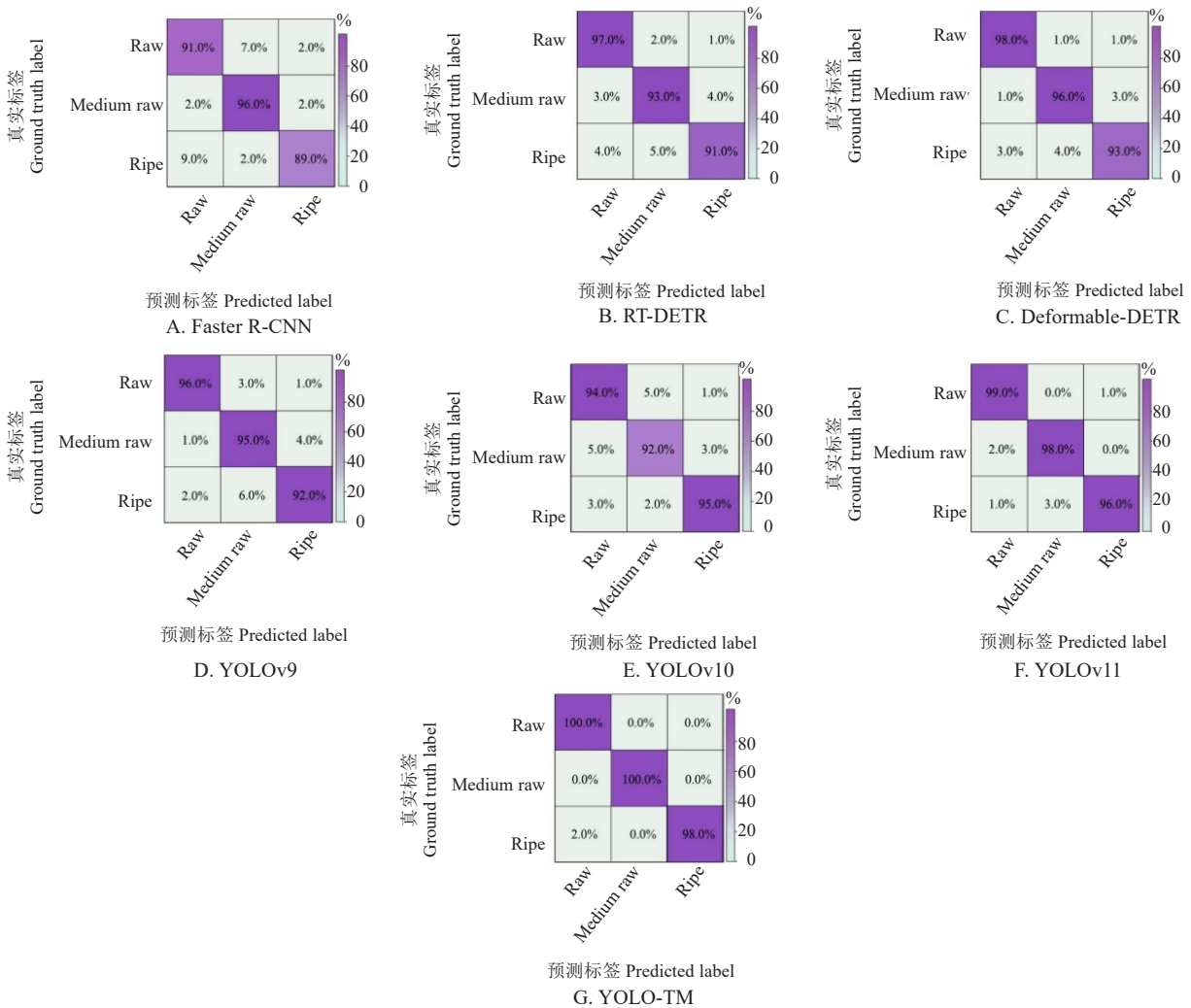


图 9 混淆矩阵对比示意图

Fig. 9 Confusion matrix contrast diagram

载 7.0 kg 和重复定位精度±0.1 mm,相较于六轴机械臂,其操作更为简便。末端执行器采用负压吸入式采摘方式,具有较好的误差容忍性,有助于减小视觉系统误差的影响,同时可在采摘过程中实现果实的收集。深度相机选用 Azure Kinect DK 相机,该相机具有紧凑的体积和完备的配套资源,能够提供在 20~60 cm 的工作范围内场景的 RGB 图像和深度信息。控制器安装在移动平台上,搭载 NVIDIA Jetson AGX Xavier,并配备显示器,以便实时监控检测结果。

由于 YOLO-TM 基于二维图像的数据进行训练,因此在实际的机器人采摘过程中,需要将二维图像中的目标信息转换为三维空间中的实际位置。为此,首先在相机坐标系下获取目标的二维像素位置,并计算该位置对应的像素宽度及目标质心位置 $P(u, v)$ 。利用深度相机提供的深度信息,系统能够获得目标在相机坐标系中的三维坐标 $P_{\text{camera}}=(x, y, z)$ 。考虑到番茄果实通常为椭圆球形状,并且存在从二维成像平面到三维空间的深度信息丢失问题,为补偿由于投影导致的深度信息丢失,系统沿 z 轴进行补偿,得到目标的近似三维位置 P'_{camera} 。随后,利用奇异值分解(singular value decomposition, SVD)方法解算出从相机坐标系到机械臂基座坐标系的齐次变换矩阵 $T_{\text{robot-camera}}$,并通过该变换矩阵将目标在相机坐标系下的三维位置 P'_{camera} 转换为机械臂基座坐标系中的三维位置 P_{robot} ,进而完成从二维到三维的转换,其计算过程如公式 17 所示:

$$\begin{bmatrix} P_{\text{robot-}x} \\ P_{\text{robot-}y} \\ P_{\text{robot-}z} \\ 1 \end{bmatrix} = T_{\text{robot-camera}} \begin{bmatrix} P'_{\text{camera-}x} \\ P'_{\text{camera-}y} \\ P'_{\text{camera-}z} \\ 1 \end{bmatrix} \quad (17)$$

采摘路径规划为优先抓取检测模型检测出的番茄成熟置信度较大的目标,并在完成采摘后返回初始位置。如果置信度相同,则优先抓取距离相机较近的物体。最后,随机选取 6 个采摘点对番茄采摘机器人进行实地验证,共计 52 个成熟番茄、23 个半成熟番茄和 9 个不成熟番茄,结果如表 6 所示。

表 6 采摘机器人检测性能对比
Table 6 Comparison of detection performance of picking robots

模型 Model	正确检测数 Correct detections	漏检数 Missed detections	误检数 False detections	检测速度 FPS/ (帧·s ⁻¹)
YOLOv11	76	8	7	9.2
YOLO-TM	79	5	5	8.9

试验结果表明, YOLOv11 和 YOLO-TM 的总体识别成功率分别为 90.5%和 94.0%,验证了视觉系统在给定识别框时的准确性,以及三维坐标转换的正确性。即使在计算资源受限的番茄采摘机器人场景下, YOLO-TM 在各项评估指标上均优于 YOLOv11,表明 YOLO-TM 在实现高检测精度的同时,能够满足番茄采摘机器人对高实时性检测的需求,具有更高的实际应用价值。

为进一步评估模型在实际温室作业中的鲁棒性,本文结合现场采摘试验,重点分析了 YOLO-TM 算法在 NVIDIA Jetson AGX Xavier 嵌入式计算平台上的环境适应表现。针对温室环境中最为棘手的光照剧变与枝叶遮挡问题,改进算法展现出了明显的工程优势。首先,在应对复杂光照条件方面,温室环境常伴随早晨的强逆光及正午的直射光斑干扰,这往往导致基准模型因局部特征过曝或过暗而丢失目标。YOLO-TM 模型通过引入 MHSA 多头注意力机制,能够从全局视角聚合图像的上下文信息,有效降低了模型对局部亮度波动的敏感度。现场测试显示,即使在逆光导致果实边缘模糊的极端条件下,该算法仍能保持稳定的边界框回归,确保了机械臂末端执行器的精准定位。其次,在应对高密度的枝叶遮挡方面,番茄植株生长中后期的枝叶交错常造成果实被大面积覆盖。试验观察发现,依靠 BAFPN 结构的双向特征融合能力,系统能够最大限度地保留和传递小尺度目标的细粒度特征。在果实被遮挡超过 50%的情况下,改进模型依然能够利用残留的纹理与形状信息,精准推断出果实的完整位置,有效减少了因漏检或误判导致的机械臂无效抓取动作。综上所述, YOLO-TM 在保证实时性的前提下,明显提升了采摘机器人在复杂农艺场景下的感知可靠性。

3 讨论与结论

在温室番茄采摘作业中,光照条件的剧烈变化、枝叶的严重遮挡以及果实尺度的多样性,长期制约着视觉感知系统的检测精度与实时性。现有的通用目标检测算法难以在这些复杂干扰下兼顾准确率与推理速度。针对上述问题,笔者提出的 YOLO-TM 模型通过引入 MHSA 机制,有效利用全局上下文信息抑制了逆光及阴影带来的背景噪声;同时,BAFPN 结构与 L_{box} 损失函数的结合,明显增强了模型对小尺度果实边缘几何特征的回归能力。试验结果表明, YOLO-TM 在自制番茄成熟度

检测数据集中 mAP 达到 95.3%, 检测速度达到 94.6 帧·s⁻¹, 在检测精度与效率之间取得了优于主流算法的平衡, 更在实地采摘试验中实现了 94.0% 的高成功率, 验证了其在工程应用中的鲁棒性。

尽管 YOLO-TM 在综合性能上表现优异, 但在部分极端场景下仍存在局限性, 需在未来工作中进一步优化: (1) 当果实呈簇状紧密堆叠且遮挡率较高时, 仅依靠矩形检测框难以精准表征果实轮廓, 可能导致机械臂在抓取时定位偏差。后续研究拟引入基于边缘感知的辅助监督模块或实例分割技术, 以进一步锐化重叠目标的边界特征。(2) 当前模型参数主要基于特定时段采集的数据集训练, 面对跨生长周期或不同温室的光照与植株形态差异时, 泛化性能可能出现波动。未来的工作将重点结合在线学习技术, 赋予系统在作业过程中持续更新模型权重的能力, 从而实现对动态农业环境的自适应感知。

综上所述, YOLO-TM 模型在检测精度、推理效率以及复杂温室环境适应性方面均表现优异, 能够为番茄自动化采摘、农业机器人视觉感知等关键环节提供可靠的图像识别技术支持。该研究成果为智能农业装备在现代设施农业中的落地应用奠定了坚实的理论与技术基础, 具有良好的推广前景和较高的工程实用价值。

参考文献

- [1] 孙亚楠, 段琳博, 钟华昱, 等. 温室气体排放与番茄产量对水肥气耦合的响应机制研究[J]. 农业机械学报, 2024, 55(5): 312-322.
- [2] 王亚薇, 何津立, 林熙淼, 等. 温室番茄采摘机器人伸缩式机械臂设计与试验[J]. 农业机械学报, 2024, 55(增刊 1): 18-28.
- [3] 刘建航, 何鉴恒, 陈海华, 等. 基于改进 YOLO v4 和 ICNet 的番茄串检测模型[J]. 农业机械学报, 2023, 54(10): 216-224.
- [4] 龙洁花, 赵春江, 林森, 等. 改进 Mask R-CNN 的温室环境下不同成熟度番茄果实分割方法[J]. 农业工程学报, 2021, 37(18): 100-108.
- [5] HU C H, LIU X, PAN Z, et al. Automatic detection of single ripe tomato on plant combining faster R-CNN and intuitionistic fuzzy set[J]. IEEE Access, 2019, 7: 154683-154696.
- [6] LIU G X, NOUAZE J C, TOUKO MBOUEMPE P L, et al. YOLO-tomato: A robust algorithm for tomato detection based on YOLOv3[J]. Sensors, 2020, 20(7): 2145.
- [7] ZU L L, ZHAO Y P, LIU J Q, et al. Detection and segmentation of mature green tomatoes based on mask R-CNN with automatic image acquisition approach[J]. Sensors, 2021, 21(23): 7842.
- [8] 周涛, 王骥, 麦仁贵. 基于改进 YOLOv8 的实时菠萝成熟度目标检测方法[J]. 华中农业大学学报, 2024, 43(5): 10-20.
- [9] 田有文, 覃上声, 闫玉博, 等. 基于改进 YOLOv8 的田间复杂环境下蓝莓成熟度检测[J]. 农业工程学报, 2024, 40(16): 153-162.
- [10] 梁放, 代东南, 牛思琪, 等. 基于改进 YOLOv5s 的草莓成熟度实时检测算法[J]. 山东农业科学, 2024, 56(11): 156-163.
- [11] 贾伟宽, 孟虎, 马晓慧, 等. 基于优化 Transformer 网络的绿色目标果实高效检测模型[J]. 农业工程学报, 2021, 37(14): 163-170.
- [12] 苗荣慧, 李港澳, 黄宗宝, 等. 基于 YOLO v7-ST-ASFF 的复杂果园环境下苹果成熟度检测方法[J]. 农业机械学报, 2024, 55(6): 219-228.
- [13] 宋立航, 张屹, 时寅豪. 基于改进 YOLOv5s 的自然环境下油桃成熟度检测方法[J]. 中国农机化学报, 2024, 45(9): 250-257.
- [14] TAN H C, LIU X P, YIN B C, et al. MHA-Net: Multihead self-attention network for occluded person re-identification[J]. IEEE Transactions on Neural Networks and Learning Systems, 2023, 34(11): 8210-8224.
- [15] 沙品洁, 郝建强, 郭林宇, 等. 高品质鲜食番茄全产业链标准体系构建及示范应用[J]. 蔬菜, 2025(5): 9-16.
- [16] SHARMA A, KUMAR V, LONGCHAMPS L. Comparative performance of YOLOv8, YOLOv9, YOLOv10, YOLOv11 and Faster R-CNN models for detection of multiple weed species[J]. Smart Agricultural Technology, 2024, 9: 100648.
- [17] 雷雨, 陈旭, 阮超, 等. 基于改进 DeblurGANv2 模型的小麦条锈菌夏孢子离焦模糊显微图像复原方法[J]. 农业机械学报, 2025, 56(1): 366-376.
- [18] 倪纪鹏, 朱立成, 董力中, 等. 基于 SwinS-YOLACT 的番茄采摘机器人实时实例分割算法研究[J]. 农业机械学报, 2024, 55(10): 18-30.
- [19] 余琼, 张瑞, 李德豪, 等. 基于残差块与注意力机制的果蔬自动识别方法[J]. 农业机械学报, 2023, 54(S2): 214-222.
- [20] 马钦, 万传峰, 卫建, 等. 基于 U-Net 和特征金字塔网络的秸秆覆盖率计算方法[J]. 农业机械学报, 2023, 54(1): 224-234.
- [21] 岳凯, 张鹏超, 王磊, 等. 基于改进 YOLOv8n 的复杂环境下柑橘识别[J]. 农业工程学报, 2024, 40(8): 152-158.
- [22] ZHAO W, KANG Y, CHEN H, et al. Adaptively attentional feature fusion oriented to multiscale object detection in remote sensing images[J]. IEEE Transactions on Instrumentation and Measurement, 2023, 72(99): 1-11.
- [23] 汤子旋, 张伟, 李俊麟, 等. FSG-YOLO: 基于 YOLOv8 改进的轻量化月面障碍物检测算法[J]. 载人航天, 2024, 30(5): 666-675.
- [24] 庄建军, 叶振兴. 基于改进 YOLOv5m 的电动车骑行者头盔与车牌检测方法[J]. 南京信息工程大学学报, 2024, 16(1): 1-10.
- [25] 金忠文, 葛动元, 姚锡凡. 基于 SWimAM 设计的 YOLOv5 轻量化交通标志检测方法[J]. 科学技术与工程, 2024, 24(31): 13475-13483.
- [26] 许光宇, 林浩杰. 基于改进 RT-DETR 的农作物害虫检测算法[J]. 齐鲁工业大学学报, 2026, 40(1): 26-37.
- [27] 刘慧, 王防修, 王意, 等. 轻量级改进 RT-DETR 的葡萄叶片病害检测算法[J]. 浙江大学学报(工学版), 2026, 60(3): 604-613.