

基于 SSPENet 的圣女果成熟度检测算法

段新娥¹, 张志旺², 周青星³

(1. 山西工程技术学院 山西阳泉 045000; 2. 南京财经大学信息工程学院 南京 210023;

3. 山西农业大学信息科学与工程学院 山西晋中 030800)

摘要: 为实现温室环境中圣女果果实成熟度的快速精确识别, 提出一种基于立体空间金字塔注意力网络(stereoscopic spatial pyramid attention network, SSPENet)的圣女果成熟度检测算法。首先, 构建空间立体注意力机制(spatial stereoscopic attention mechanism, SSAM), 通过自适应聚焦关键区域, 提升对果实特征的感知能力; 其次, 在颈部网络引入局部注意金字塔模块(local attention pyramid module, LAPM), 增强小尺度圣女果的特征融合, 提高小尺度目标的检测精度; 最后, 提出高效的几何回归损失函数 L_{EGR} , 优化边界框的几何特性, 进一步提升小尺度圣女果的定位精度。试验结果表明, SSPENet 在圣女果成熟度数据集上取得了 96.1% 的 mAP, 相比基线模型提升 5.1 百分点, 且推理速度可达 94.7 帧·s⁻¹, 在检测精度与计算效率之间实现了良好平衡。该研究为温室环境下圣女果成熟度检测提供一种高效且可扩展的技术方案, 具有广阔的应用前景。

关键词: 圣女果; 成熟度; SSPENet; 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(2025)12-052-11

Detection method of cherry tomato fruit ripeness based on SSPENet

DUAN Xin'e¹, ZHANG Zhiwang², ZHOU Qingxing³

(1. Shanxi Institute of Technology, Yangquan 045000, Shanxi, China; 2. College of Information Engineering, Nanjing University of Finance & Economics, Nanjing 210023, Jiangsu, China; 3. College of Information Science and Engineering, Shanxi Agricultural University, Jinzhong 030800, Shanxi, China)

Abstract: To achieve rapid and accurate identification of cherry tomato fruit ripeness in greenhouse environment, this study proposed a ripeness detection algorithm based on the stereoscopic spatial pyramid attention network (SSPENet). First, a spatial stereoscopic attention mechanism (SSAM) was constructed to enhance the perception of fruit features by adaptively focusing on key regions. Second, a local attention pyramid module (LAPM) was incorporated into the neck network to strengthen the feature fusion of small-scale cherry tomato, thereby improving detection accuracy for small-scale targets. Finally, an efficient geometric regression loss function (L_{EGR}) was proposed to optimize the geometric properties of bounding boxes, further improving the localization accuracy for small-scale cherry tomatoes. The experimental results showed that SSPENet achieved 96.1% mAP on the cherry tomato ripeness dataset, representing a 5.1 percentage point improvement over the baseline model, with an inference speed of 94.7 frames per second. This achieved a good balance between detection accuracy and computational efficiency. This study provides an efficient and scalable technical solution for cherry tomato ripeness detection in greenhouse environment, with broad application prospects.

Key words: Cherry tomato; Maturity; SSPENet; Loss function; Small scale target

圣女果作为一种具有高营养价值的小型番茄品种, 近年来在全球范围内的种植面积和市场需求持续增长^[1]。果实成熟度的准确识别与分级采摘对于其贮藏稳定性、加工适应性和商品价值具有至关重要的影响。然而, 在温室种植环境中, 受限于种植密度大及温湿度波动明显等因素, 圣女果的成熟

度检测、分拣与采摘作业仍高度依赖人工, 导致整体作业效率低下, 难以满足现代设施农业对高效与智能化生产的要求^[2]。因此, 开发一种实时检测圣女果成熟度的检测方法, 不仅有助于提高采摘效率、降低人力成本, 还能有效推动温室果蔬管理的精细化和智能化, 具有重要的工程应用价值与现实意义。

收稿日期: 2025-03-19; 修回日期: 2025-06-07

基金项目: 国家自然科学基金(12001266)

作者简介: 段新娥, 女, 副教授, 主要从事软件工程、模式识别及算法方向的研究。E-mail: duanxine7310@126.com

通信作者: 张志旺, 男, 教授, 主要从事最优化与机器学习、人工智能和自然语言处理方向的研究。E-mail: zhangzhiwang73@163.com

传统番茄成熟度检测方法主要依赖计算机视觉技术,通过颜色或形状等低层特征进行图像分析与分类^[3]。例如,Yin等^[4]利用CIELAB颜色空间和K-means聚类算法对番茄图像进行分割,并结合形态学处理实现了重叠与遮挡区域的剔除,检测准确率达到95%。Liu等^[5]提出基于番茄轮廓特征的分级方法,利用傅里叶描述子提取形状信息并结合线性回归模型进行分类,结果表明该方法的平均分级准确率为90.7%。梁喜凤等^[6]采用最大类间方差法(Otsu)对R-G(Red-Green)直方图进行分析,实现了番茄成熟度的初步划分。然而,上述方法高度依赖手工提取的单一低层特征,难以全面表征果实在复杂背景、光照变化及遮挡重叠等实际场景中的多样化表现,易导致检测结果不稳定,难以满足温室环境下对果实成熟度检测的准确性与适应性要求。

随着深度学习算法和计算机硬件性能的提升,基于卷积神经网络(convolutional neural network, CNN)的目标检测方法为果实成熟度检测提供了全新的技术路径。CNN系列算法可细分为两阶段算法与单阶段算法。其中,两阶段算法如Mask R-CNN^[7]和Cascade R-CNN^[8],通过生成候选边界框实现高精度检测。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。岳有军等^[8]将Cascade RCNN网络中的非极大值抑制算法替换为Soft-NMS(soft non-maximum suppression)算法,采用适合番茄形状的锚框,从而增强模型对重叠果实的识别能力。Hu等^[9]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,其预测框在水平方向与垂直方向的中心偏移平均相对误差分别为0.261%和1.179%。以上基于两阶段的算法通过生成候选区域并进行精确的目标定位与分类,能实现较高的检测精度,但其计算复杂度较高,限制了实时应用的普适性。

另一类单阶段方法如SSD^[10]和YOLO系列^[11-13],通过端到端的目标分类与定位,显著降低模型复杂度,进而提高检测速度,因而更适合实时检测需求。张磊等^[14]提出了一种改进的YOLOv4算法,旨在应对复杂环境下番茄果实的检测问题。该方法能够在6种不同环境下有效识别番茄,并对成熟度为成熟、半成熟和未成熟的番茄进行分类,其平均检测精度分别达到84%、77%和85%。Liu等^[15]设计了一种YOLO-Tomato模型,引入密集架

构以增强特征复用功能,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。尽管这些方法在检测精度上取得了一定进展,但由于目标定位精度不足,特别是在复杂场景中难以精确拟合番茄的真实轮廓,仍存在一定的局限性。为此,刘芳等^[16]引入多尺度检测结构,通过更精细地提取番茄果实特征信息,并优化类别标签与预测框的生成过程,从而显著提高了番茄果实的定位精度。Zeng等^[17]通过遗传算法对模型超参数进行优化,进一步平衡了分类损失与定位损失的权重分配,显著提升了模型对番茄果实的边界框回归能力,从而提高了定位精度。然而,此类改进仍存在一定局限,尤其在检测被叶片遮挡的番茄目标时,易出现误检现象且对番茄的重复框选问题仍未得到有效解决。为应对这一挑战,一些学者尝试引入注意力机制来进一步提升模型的性能。苗荣慧等^[18]在特征融合网络中加入全局注意力机制(global attention mechanism, GAM),以加强对圣女果关键特征的关注,从而提高了成熟度检测的准确性。Wu等^[19]则提出上下文锚点注意力机制(context anchor attention, CAA),旨在通过强化模型对番茄果实的识别能力,精确地进行成熟度和采摘位置的检测。Wei等^[20]基于YOLOv11模型,结合通道注意力机制和空间注意力机制,进一步增强了模型对圣女果在复杂田间环境中的特征表达能力,提升了6.2%的mAP。但在处理小尺度番茄目标时,仍存在漏检问题且检测精度未能完全满足实际应用的需求。

尽管现有番茄成熟度检测研究已取得一定进展,但仍存在若干亟待解决的难点:

1)在温室种植环境中,圣女果果实常遭受叶片遮挡且果实间存在重叠,使得模型难以准确区分相邻果实的成熟度特征。

2)圣女果果实体型较小,模型在小尺度目标检测时容易发生漏检并降低整体检测性能。针对上述挑战,提出一种基于立体空间金字塔注意力网络(stereoscopic spatial pyramid attention network, SSPENet)的圣女果成熟度检测算法,旨在提升模型对细粒度纹理、小目标特征及复杂场景的适应能力。

SSPENet的核心改进主要体现在以下三方面:

1)在骨干网络设计一种全新的空间立体注意力机制(spatial stereoscopic attention mechanism, SSAM),通过融合通道间关联特征与空间显著性信息,以增强模型在复杂环境中对关键视觉区域的聚焦能力,提升遮挡与重叠条件下的检测准确性。

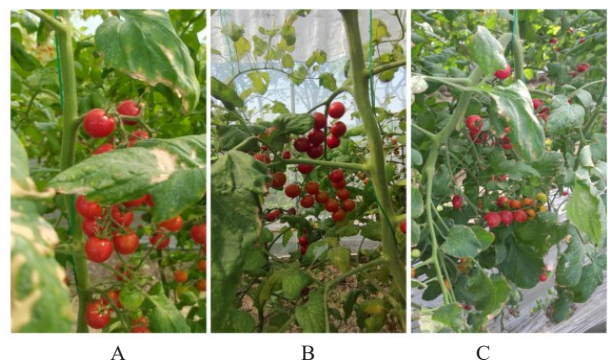
2)引入局部注意金字塔模块(local attention pyramid module, LAPM)^[21],通过分层提取与加权增强局部显著区域的特征响应,引导模型多尺度整合局部特征,从而提升对小尺度果实的识别能力。

3)构建损失函数(efficient geometric regression loss, L_{EGR}),优化边界框定位过程,进一步提高模型对小目标位置与形状的拟合精度。

1 材料与方法

1.1 图像采集

试验研究区域选取广东省茂名市电白区那霍镇水石村圣女果种植基地,于 2024 年 3—5 月使用 DF500-1944P 工业相机在温室环境下拍摄不同植株和不同成熟度的圣女果图像。采集时间涵盖上午、中午和下午,以获取在不同光照条件下的图像数据,共获取分辨率为 1080×1920 像素的圣女果图像 7061 幅。部分图像采用跟踪拍摄方式,记录圣女果在成熟过程中的动态变化。采集的图像类型包括顺光图像、逆光图像、枝叶遮挡图像、无枝叶遮挡图像以及果实重叠图像等。部分样本图像如图 1 所示。



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

Note: A. A scene featuring fruit overlapping and occlusion caused by branches and leaves; B. A backlit scene characterized by fruit overlapping and occlusion from branches and leaves; C. A front-lit scene with occlusion due to branches and leaves.

图 1 圣女果图像样本示例

Fig. 1 Cherry tomato image sample sample

1.2 数据集构建

依照 GH/T 1193—2021 国家标准对圣女果成熟度进行划分,将圣女果分为未成熟(raw)、半成熟(medium raw)和成熟(ripe)三类。为确保数据的准确性和标注的一致性,使用 Labellmg 图像标注软件对圣女果图像进行人工标注,并将每个图像的位置信息及类别标签保存为 XML 格式的标签文件,各

类别标签的分布情况如表 1 所示。具体而言,未成熟类包含未熟期和绿熟期的圣女果,半成熟类包括变色期及红熟前期的圣女果,而成熟类则包括红熟中期和红熟后期的圣女果。为提高模型的泛化能力,防止模型在训练数据中过拟合特定特征,采用随机镜像、翻转、高斯模糊和高斯噪声等数据增强方法对数据集进行扩建,该手段有效地模拟了圣女果在真实场景中的生长姿势、光照条件和拍摄角度的变化,从而增加了训练样本的多样性。最终构建的圣女果数据集图像共 9645 张,图像均以 JPG 格式保存。为满足模型训练需求,按照 8:1:1 的比例将数据集划分为训练集(7716 张)、验证集(965 张)和测试集(964 张)。

表 1 各类标签分布
Table 1 Label distribution

数据集 Dataset	未成熟 Raw	半成熟 Medium raw	成熟 Ripe	总计 Total
训练集 Training set	27 584	15 254	34 281	77 119
验证集 Verification set	3 725	1 902	4 451	10 078
测试集 Test set	3 448	2 017	4 319	9 784

1.3 SSPENet 圣女果成熟度检测算法

YOLOv11 是 Ultralytics 公司推出的最新一代目标检测模型^[20],其主要改进包括采用 C3k2 模块替代 C2f 模块,并在 SPPF 模块后引入 C2PSA 模块。通过深度可分离卷积方法,YOLOv11 增强了对上下文信息的捕捉能力,并有效减少冗余计算,从而提升目标检测精度。这些创新使得 YOLOv11 在检测性能上较前代模型取得显著提升,尤其适用于对实时性要求较高的应用场景。

尽管 YOLOv11 在检测效率和精度方面表现优异,但在圣女果成熟度检测中,受限于枝叶遮挡、果实间重叠及尺度变化等因素,其检测效果仍存在进一步优化的空间。为此,提出一种 SSPENet 算法,旨在实现高效精准的圣女果成熟度检测,其网络结构如图 2 所示。

具体而言,在骨干网络构建一种空间立体注意力机制 SSAM,通过对输入特征图进行加权,聚焦于图像中的关键区域,提升模型对圣女果成熟度特征的识别能力;在颈部网络引入局部注意金字塔模块 LAPM,使模型能够在不同尺度上提取和融合细粒度特征,避免小尺度圣女果信息丢失;此外,提出了一种 L_{EGR} 损失函数,旨在精确度量预测框与真实框之间的差异,从而提高小尺度圣女果的定位精度。

1.3.1 空间立体注意力机制 在圣女果成熟度检

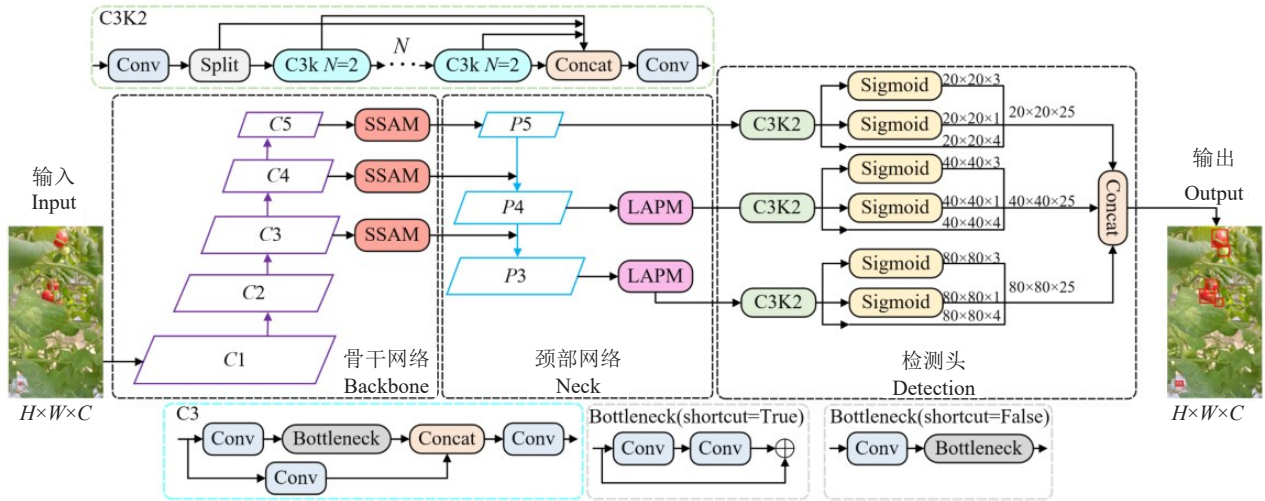
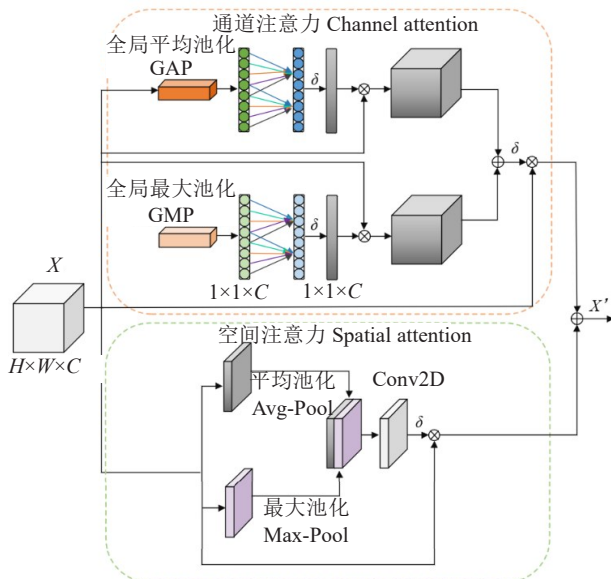


图 2 SSPENet 网络结构图

Fig. 2 SSPENet network structure

测任务中,由于果实常受到叶片遮挡且果实之间存在重叠,使类间与类内特征的分变得更为困难。此外,光照变化对图像质量的影响也会导致 YOLOv11 在特征提取时难以有效聚焦关键区域,进而引发误检和漏检。为此,笔者构建一种空间立体注意力机制 SSAM,旨在增强模型对关键区域的聚焦能力,进而提高检测精度。SSAM 注意力机制分为通道注意力和空间注意力 2 个分支,其网络结构如图 3 所示。

具体而言,在通道注意力分支中,首先对输入



注:⊕表示逐元素加法,⊗表示逐元素乘法。

Note:⊕ denotes element-wise addition, and ⊗ denotes element-wise multiplication.

图 3 SSAM 网络结构图

Fig. 3 SSAM network structure

特征图 $X \in R^{H \times W \times C}$ 分别应用全局平均池化(global average pooling, GAP)和全局最大池化(global max pooling, GMP)操作,提取全局通道特征以实现降维处理。通过 GAP 和 GMP 分别从全局视角捕获通道的平均信息和显著信息,增强通道的特征表达能力^[18],随后,为避免降维操作影响通道注意力的学习,采用 1D 卷积实现局部跨通道交互,即在不同通道之间进行信息交互,而不改变输入特征图的维度。局部跨通道交互的覆盖范围通过 1D 卷积核大小 k 的自适应调整来控制。卷积核大小的计算如公式(1)所示:

$$k = \Psi(D) = \left\lfloor \frac{\log_2(D)}{\gamma} + \frac{b}{\gamma} \right\rfloor_{\text{odd}} \quad (1)$$

式中, $\lfloor t \rfloor_{\text{odd}}$ 表示取 t 最邻近的奇数, $\gamma=2$ 和 $b=1$ 为超参数, D 为输入特征层的通道数,通过映射函数 $\Psi(D)$ 实现高维通道间远距离的交互与低维通道间近距离的交互。

将通过 1D 卷积得到的结果输入 Hard-Sigmoid (δ) 激活函数,以对卷积输出进行线性转换,从而生成每个通道的注意力权重,激活过程如公式(2)所示:

$$\delta(x) = \begin{cases} 0, & x < -3 \\ x/6 + 1/2, & -3 \leq x \leq 3 \\ x, & x > 3 \end{cases} \quad (2)$$

式中, x 是通过 1D 卷积得到的特征值。Hard-Sigmoid 激活函数有效避免了传统 Sigmoid 函数中指数运算的复杂度,同时确保输出值限制在 $[0, 1]$ 。

将计算得到的注意力权重图与输入特征图逐通道进行逐元素相乘,生成最终的通道注意力特征

图 $M_c(X)$, 如公式(3)所示:

$$M_c(X) = X \otimes \delta [W_1 \times \text{GAP}(X) \oplus W_2 \times \text{GMP}(X)]. \quad (3)$$

式中, W_1 和 W_2 表示 1D 卷积层的权重。

在空间注意力分支中, 沿通道轴分别应用平均池化(average pooling)和最大池化(max pooling)操作, 并将池化结果在通道维度上拼接为一个 $H \times W \times 2$ 的特征描述符。随后, 通过 Conv2D 卷积将通道维度进行压缩, 输出一个 $H \times W \times 1$ 的空间注意力特征图。接着, 利用 Hard-Sigmoid 激活函数对卷积结果进行线性变换, 并与输入特征图逐元素相乘, 生成空间注意力加权后的特征图 $M_s(X)$, 如公式(4)所示:

$$M_s(X) = X \otimes \delta (\text{Conv2D}([\text{Max-Pool}; \text{Avg-Pool}])). \quad (4)$$

将通道加权后的特征图与空间注意力权重图逐元素相乘, 得到最终的输出特征图 X' , 如公式(5)所示:

$$X' = M_c(X) \otimes M_s(X). \quad (5)$$

在骨干网络中, SSAM 模块分别嵌入于 C3、C4 和 C5 层的输出之后, 通过通道注意力的无降维跨通道交互和空间注意力的显著性编码, 增强模型对遮挡、重叠和光照变化等复杂场景中关键区域的聚焦能力, 从而有效提升圣女果成熟度检测的精度。

1.3.2 局部注意力金字塔模块 由于圣女果的果实体积较小, 其特征在浅层特征图上分辨率低、局部信息不明显, 导致小尺度目标难以被有效捕获, 现有模型的检测性能受到显著限制。尽管 YOLOv11 通过 FPN+PAN 结构实现了多尺度特征融合, 但对于小尺度目标特征表达的不一致性仍存在局限性。为解决这一问题, 笔者引入一种局部注意力金字塔模块 (LAPM), 通过逐层增强小目标的显著特征而提升检测精度。LAPM 的流程示意图如图 4 所示。

具体而言, LAPM 的输入为浅层特征图 $F \in \mathbb{R}^{N \times H \times W \times C}$ 。其中, N 表示批量大小, H 和 W 分别表示

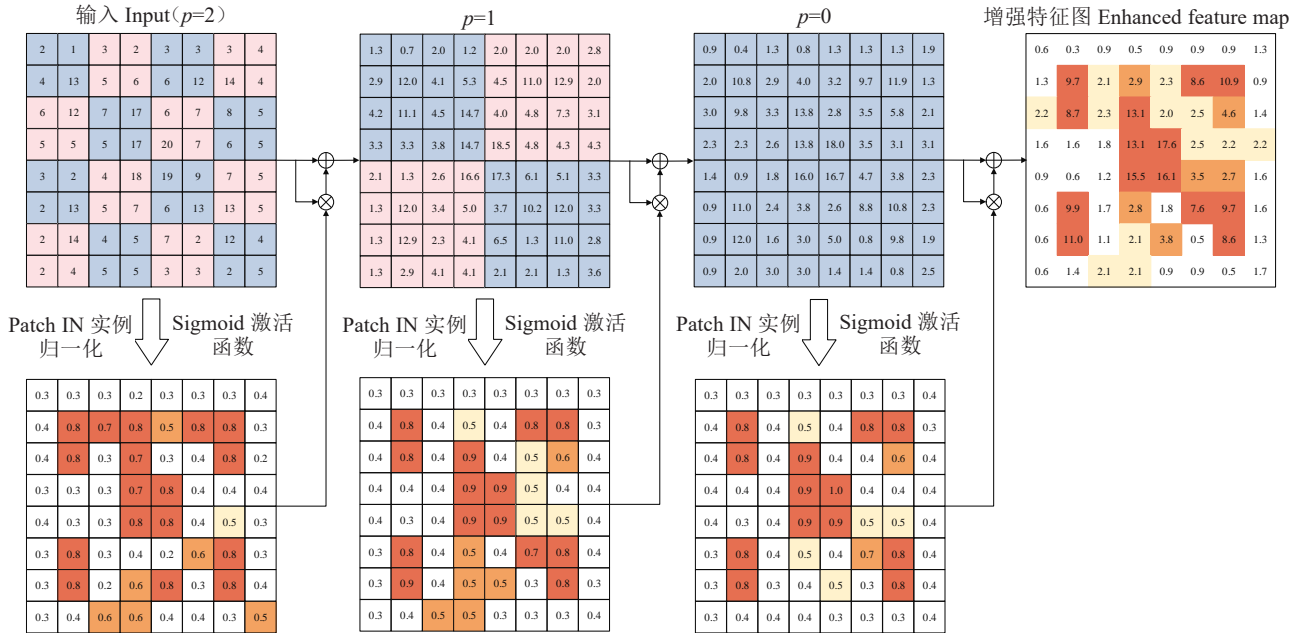


图 4 LAPM 流程示意图

Fig. 4 LAPM process diagram

特征图的高度和宽度, C 表示通道数。输入特征图 F 被引入到金字塔结构的最顶层 $p=n$ 。逐层向下进行分割与处理。在金字塔的第 p 层, 输入特征图 F_p 被划分成尺寸大小为 $H/2^p \times W/2^p$ 的多个局部区域 (patches), 随着层级的降低, 局部分块的尺度逐步增大。在每个局部区域内执行实例归一化 (instance normalization, Patch IN), 实现局

部特征的标准化的, 具体计算公式如下:

$$\text{Patch IN}(F_p, p) = \frac{F_p - \mu_{\text{patch}}}{\rho_{\text{patch}} + \varepsilon}. \quad (6)$$

式中, μ_{patch} 和 ρ_{patch} 分别表示当前局部区域的均值和标准差, ε 为数值稳定项, 防止分母为 0。

对归一化后的特征图应用 Sigmoid 激活函数, 生成局部注意力图 w_p , 如公式(7)所示:

$$w_p = M(F_p, p) = \sigma(\text{Patch IN}(F_p, p))。 \quad (7)$$

式中, $M(\cdot, p)$ 表示第 p 层的注意力图计算函数。 σ 为 Sigmoid 激活函数,用于将归一化后的特征映射到 $[0,1]$ 范围,突出局部高响应区域的显著特征。

随后,将生成的局部注意力图 w_p 与输入特征图 F_p 进行逐元素相乘,实现显著特征的加权增强。同时,引入残差连接机制融合原始特征与注意力加权特征,输出第 $p-1$ 层的特征图 F_{p-1} ,如公式(8)所示:

$$F_{p-1} = \theta \times F_p + (1 - \theta) \times w_p \otimes F_p。 \quad (8)$$

式中, θ 为残差连接的权重系数, \otimes 表示逐元素相乘操作。

第 $p-1$ 层的输出特征图 F_{p-1} 作为下一层的输入,重复“局部分割→归一化→激活→加权增强”的过程,逐层聚合多尺度注意力特征。最终,在最底层 $p=0$ 输出增强后的特征图 F_{-1} ,其计算过程如公式(9)所示:

$$F_{-1} = \theta \times F_0 + (1 - \theta) \times w_0 \otimes F_0。 \quad (9)$$

如图2所示,为进一步提升小尺度圣女果的检测性能,将LAPM模块应用于浅层特征图 $P3$ 和 $P4$ 。由于 $P3$ 层特征图尺寸较大,设置金字塔层级 $n=3$,局部分块尺寸为 $\{10 \times 10, 20 \times 20, 40 \times 40, 80 \times 80\}$; $P4$ 层特征图尺寸较大,设置金字塔层级 $n=2$,局部分块尺寸为 $\{10 \times 10, 20 \times 20, 40 \times 40\}$,处理后的特征图 $P3'$ 和 $P4'$ 的计算过程如公式(10~11)所示:

$$P4' = \text{LAPM}(P4, n=2); \quad (10)$$

$$P3' = \text{LAPM}(P3, n=3)。 \quad (11)$$

式中, n 表示金字塔的最高层级。

LAPM 模块通过逐层局部分割、实例归一化和注意力加权的方式,实现对小尺度目标显著特征的逐级增强。通过金字塔结构在不同尺度上对局部区域进行特征提取和融合,捕获多尺度下的小目标信息,为圣女果成熟度检测任务中小目标特征的有效建模提供支持。

1.3.3 改进损失函数 在目标检测任务中,损失函数设计对模型性能具有关键影响。YOLOv11 的损失函数由分类损失和回归损失两部分组成。其中,分类损失采用可变焦点损失(varifocal loss, VFL)。

回归损失由分布式焦点损失(distribution focal loss, DFL)和完全交并比损失(complete intersection over union, CIOU)构成。

CIOU 损失函数通过结合预测框与真实框的重叠面积、中心点距离以及长宽比等几何特性,引导模型对边界框进行精确优化。

虽然 CIOU 损失函数在优化目标几何特性方面

具有优势,但其长宽比相似性度量指标 v 仅衡量预测框与真实框整体的长宽比差异,未能解耦宽度和高度的独立变化。这一局限性导致模型对尺度变化敏感的小目标优化效果较差,从而影响检测精度。

为此,ElIoU(efficient intersection over union)损失函数^[22]被提出。ElIoU 在 CIOU 的基础上引入宽度和高度的独立优化,通过分解宽高几何差异,增强边界框几何特征的表达能力。

尽管 ElIoU 在几何特征表达方面取得一定进展,但其在预测框与真实框距离较远时收敛速度较慢,同时宽度和高度的优化计算较为复杂,仍存在训练效率低下的问题。为解决上述问题,提出一种新的损失函数(efficient geometric regression loss, L_{EGR})。 L_{EGR} 在保留 ElIoU 宽高解耦思想的基础上,结合 CIOU 的长宽比调整方法,对宽度、高度和中心点距离进行联合优化。通过提升初期收敛速度和优化计算效率,能够更加精确地度量预测框与真实框之间的几何相似性,并提升边界框回归性能。 L_{EGR} 损失函数计算公式(12)如下:

$$L_{\text{EGR}} = 1 - \text{IoU} + \alpha v + \frac{\rho^2(b, b^{\text{gt}})}{c^2} + \frac{\rho^2(w, w^{\text{gt}})}{c_w^2} + \frac{\rho^2(h, h^{\text{gt}})}{c_h^2} \quad (12)$$

式中, αv 用于初期快速调整预测框比例。

综上,笔者提出的 SSPENet 圣女果成熟度检测算法的总损失 L_{sum} ,如公式(13)所示:

$$L_{\text{sum}} = L_{\text{VFL}} + L_{\text{DFL}} + L_{\text{EGR}}。 \quad (13)$$

通过引入 L_{EGR} ,模型能够更高效地优化边界框的几何特性,进一步提升小尺度目标的定位精度,从而增强圣女果成熟度检测中小目标的检测性能。

1.4 试验环境配置与网络参数设置

为确保圣女果成熟度检测试验的公平性,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 300,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 Ranger(融合 RAdam 和 Lookahead 方法的优化器),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。

1.5 评价指标

对于圣女果果实成熟度检测任务而言,模型的

表 2 试验硬件配置和软件环境
Table 2 Experiment hardware configuration and software environment

硬件	配置参数	软件	配置参数
Hardware	Parameters	Software	Parameters
CPU	Intel Core i9-10900K CPU	操作系统	Ubuntu 18.04
		Operating system	
GPU	NVIDIA GeForce RTX 2080 Ti 11 G	PyTorch	1.8.0
		CUDA	11.2
内存	64 G	CUDNN	8.1.0
Memory			
硬盘	2 T		
Hard disk			

评估需同时兼顾检测精度与处理效率。为此,采用准确率(precision, P)、召回率(recall, R)、平均精度均值(mean average precision, mAP)对各模型在圣女果成熟度检测中的性能进行定量评价。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。

2 结果与分析

2.1 不同注意力机制算法对比

为进一步验证所构建的 SSAM 注意力机制在圣女果果实成熟度检测中的有效性,在骨干网络中分别嵌入 CBAM^[22]、SE^[23]和 ECA^[24]注意力机制,在圣女果果实成熟度数据集上进行对比,试验结果如表 3 所示。

由表 3 可知,嵌入不同注意力机制后,模型的检测精度、召回率及平均精度均值均有所提升,表明注意力机制的引入能够有效增强网络对特征信息的捕获能力,从而提升整体检测性能。ECA 机制通过避免全连接操作,达到了 93.2%的 mAP,略优于 SE 注意力机制。CBAM 机制结合空间与通道注意力,使 mAP 提升至 94.7%,但帧率下降至 91.5 FPS。其中,SSAM 实现了 96.6%的准确率,相

表 3 不同注意力机制性能对比
Table 3 Comparison of performance of different attention mechanisms

注意力机制	准确率	召回率	平均精度均值	检测速度
Attention mechanism	P/%	R/%	mAP/%	FPS
无 None	93.1	88.1	90.2	96.1
SE	93.8	90.4	91.7	92.3
ECA	94.2	92.6	93.2	95.7
CBAM	95.8	93.3	94.7	91.5
SSAM	96.6	93.9	96.1	94.7

比未使用注意力机制,提升了 3.5 百分点。此外,召回率和 mAP 分别达到了 93.9%和 96.1%,均优于其他注意力机制。SSAM 在保持较高精度和召回率的同时,实现了 94.7 FPS 的检测速度,表明其能够兼顾检测性能与实时性,适用于圣女果果实成熟度检测任务中的复杂场景。

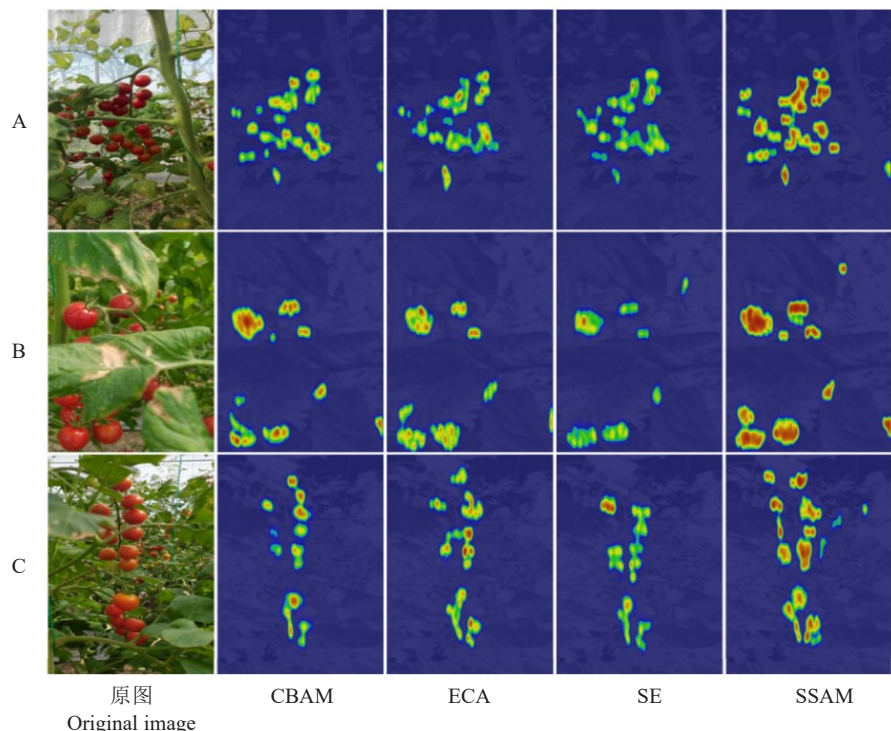
为进一步直观展示不同注意力机制对圣女果果实成熟度检测特征的影响,采用 Grad-CAM^[25]类激活热力图对模型的 3 个输出层进行可视化分析。通过热力图可明确模型预测时的注意区域,揭示其对目标特征的关注分布,从而提高模型的可解释性。试验对比结果如图 5 所示,其中红色区域越鲜亮、面积越大,表示模型的预测关注度越高。

由图 5 可知,当骨干网络分别引入 CBAM、ECA 和 SE 注意力机制时,网络对密集分布的圣女果果实信息的捕获能力仍存在一定不足,尤其是 SE 注意力机制,其关注区域较为分散且覆盖范围有限,目标特征捕获不充分。相比之下,SSAM 注意力机制对圣女果成熟度特征的关注程度最高,能够更加精准地捕获目标的空间位置信息,同时在复杂场景中快速聚焦关键区域,展现出更强的特征表达能力与定位能力。

2.2 不同损失函数对比

为验证 L_{EGR} 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。选取 CIoU、EIoU、GIoU (generalized intersection over union)^[26]、DIoU (distance intersection over union)^[27]、SIoU (segmented intersection over union)^[26]进行对比,以评估它们在圣女果果实成熟度检测任务中的性能差异。试验结果如表 4 所示。

由表 4 可知,不同损失函数对圣女果成熟度检测任务中的性能差异明显。GIoU 和 DIoU 分别实现 92.6%和 94.8%的 mAP,但在召回率和准确率上表现一般,表明其在边界框几何优化中的能力存在局限性。EIoU 通过改进边界框的宽高比调整和中心点定位策略,将 mAP 提升至 95.0%,明显优于 GIoU 和 DIoU。SIoU 进一步改进宽高比例匹配,mAP 达到 95.5%,准确率和召回率也有所提升,展现了其在几何特性优化中的优势。相比之下, L_{EGR} 通过提升初期收敛速度和优化计算效率,更精确地度量了预测框与真实框的几何相似性,从而全面提升模型检测性能。其准确率、召回率和 mAP 分别达到 96.6%、93.9%和



注:A 为果实重叠场景;B 为枝叶遮挡严重场景;C 为背景叶片干扰场景。
Note: A. A scene characterized by overlapping fruits; B. A scene with severe occlusion caused by branches and leaves; C. A scene affected by background interference from foliage.

图 5 不同注意力机制在各种复杂环境的类激活热力图
Fig. 5 Activation heat maps of different attention mechanisms in various complex environments

表 4 不同损失函数对比试验

Table 4 Comparative experiment of different loss functions

损失函数	准确率	召回率	平均精度均值
Loss function	P	R	mAP
CIoU	96.0	92.8	95.2
EIoU	95.9	93.4	95.0
GIoU	94.3	91.2	92.6
DIoU	95.4	92.7	94.8
SIoU	96.2	93.7	95.5
L_{EGR}	96.6	93.9	96.1

96.1%,进一步证明了其在小目标检测任务中的高效性。

2.3 消融试验

通过消融试验分析 SSAM 注意力机制、LAPM 模块和 L_{EGR} 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 5 所示,其中“√”表示启用该模块,“×”表示未启用该模块。

由表 5 可以看出,单独引入 SSAM 注意力机制后,模型的 mAP 提升至 92.3%,同时参数量略微增加至 2.7 M,FLOPs 增至 6.9 G,表现出明显的精度提升效果。仅嵌入 LAPM 模块后,mAP 达到 92.0%,并在计算效率方面保持较优表现,展现出良好的适用性。采用 L_{EGR} 损失函数后,mAP 提升至

表 5 消融试验

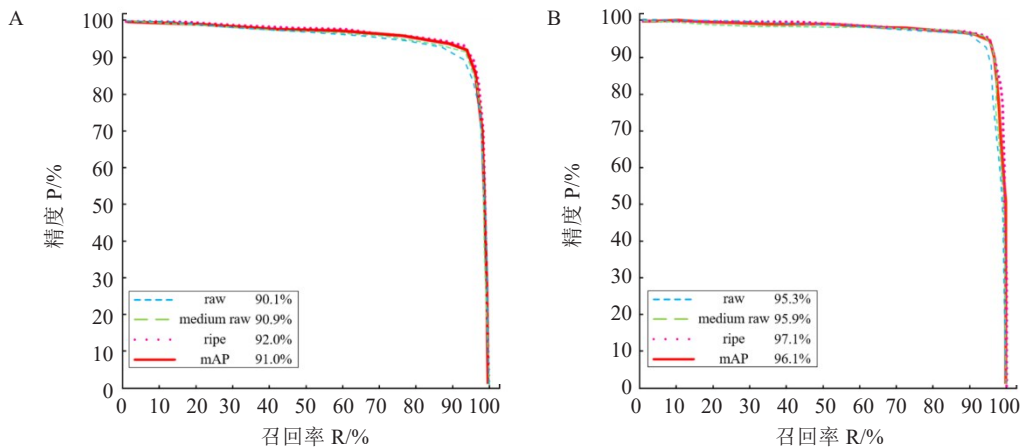
Table 5 Comparative experiment of different loss functions

YOLOv11	SSAM	LAPM	L_{EGR}	平均精度 AP/%			召回率 R/%	平均精度均值 mAP/%	检测速度 FPS	参数量 Params/M	浮点运算次数 FLOPs/G
				未成熟	半成熟	成熟					
				Raw	Medium raw	Ripe					
√	×	×	×	90.1	90.9	92.0	89.9	91.0	99.8	2.6	6.5
√	√	×	×	91.9	92.1	92.9	91.5	92.3	98.5	2.7	6.9
√	×	√	×	91.5	91.7	92.8	91.1	92.0	97.8	2.8	7.5
√	×	×	√	91.4	91.7	92.6	90.9	91.9	99.3	2.6	6.5
√	√	√	×	94.8	94.7	96.1	92.8	95.2	96.3	2.9	8.1
√	√	√	√	95.3	95.9	97.1	93.9	96.1	94.7	2.9	8.1

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 SSAM 和 LAPM 模块联合使用时, mAP 进一步提升至 95.2%,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 3 种改进策略后模型的 mAP 提升至 96.1%,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明显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 6 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* 曲线所围成的面积越大,模型的综合性能越优。由图 6-A 和图 6-B 的面积对比可以看出,所提出的 SSPENet 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 YO-



注:A 为 YOLOv11 模型 *P-R* 曲线;B 为 MSASNet 模型 *P-R* 曲线。

Note: A. *P-R* curve of the YOLOv11 model; B. *P-R* curve of the MSASNet model.

图 6 改进前后模型在圣女果果实成熟度数据集上的 *P-R* 曲线图

Fig. 6 *P-R* curves of the model before and after improvement on cherry tomatoes fruit maturity dataset

LOv11 明显增大,进一步验证了 SSPENet 在检测精度和整体性能上的优越性。

2.4 不同模型对比

为验证 SSPENet 模型在圣女果成熟度检测中的有效性,在自建数据集上将其与 9 种主流检测方法进行比较。包括 Faster R-CNN、SSD、Deformable-DETR^[28]、DINO^[28]、RT-DETR^[28]、YOLOX、YOLOv9、YOLOv10 和 YOLOv11。对比结果如表 6 所示。

由表 6 可知,Faster R-CNN 作为双阶段检测算法,在 mAP 方面达到了 85.7%,展现出较高的检测精度,但由于计算复杂度较高,检测速度仅为 6.9 FPS,难以满足实时检测需求。DINO 和 RT-DETR 在 mAP 方面分别达到 88.9%和 89.0%,虽优于 YOLOX 模型,但计算开销较大,未能实现精度与效率的良好平衡。YOLOv9 以最少的参数量和 7.7 G 的 FLOPs 实现了 89.3%的 mAP,表现出较高的检测效率;YOLOv10 将 mAP 提升至 90.2%,但检测速度略有下降;YOLOv11 进一步优化了精度与速度的平衡,实现了 91.0%的 mAP 和 99.8 FPS 的检测速度,在 YOLO 系列算法中表现最佳。而 SSPENet 模型在 mAP 上达到 96.1%,相较于上述算法分别提

表 6 对比试验
Table 6 Contrast experiment

算法 Algorithm	准确率 P/%	召回率 R/%	平均精度 均值 mAP/%	检测速度 FPS	参数量 Params/ M	浮点运 算次数 FLOPs/G
Faster R-CNN	86.3	85.0	85.7	6.9	102.7	168.4
SSD	85.5	83.8	84.8	17.6	29.4	99.1
Deformable-DETR	89.9	88.8	89.2	58.6	40.0	173.0
DINO	89.2	88.4	88.9	55.4	47.5	178.5
RT-DETR	89.7	88.7	89.0	62.4	20.0	99.1
YOLOX	88.4	86.9	87.9	68.4	9.0	26.8
YOLOv9	90.0	88.7	89.3	101.2	2.0	7.7
YOLOv10	90.7	89.3	90.2	100.3	2.3	6.7
YOLOv11	91.2	89.9	91.0	99.8	2.6	6.5
SSPENet	96.6	93.9	96.1	94.7	2.9	8.1

升 10.4、11.3、6.9、7.2、7.1、8.2、6.8、5.9 和 5.1 百分点,尽管 SSPENet 的帧率略低于 YOLOv9 和 YOLOv10,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圣女果的果实成熟度图像进行对比测试,检测效果如图 7 所示。

通过对图 7-A、图 7-B 和图 7-C 中不同模型检测结果的对比分析可知,在遮挡严重、果实重叠及光照变化显著等复杂场景下,YOLOv9 与 YOLOv10 模型普遍存在较高的误检率,表现出较弱的环境适应性。尽管 YOLOv11 在常规场景下具备较强的目标检测能力,但在逆光或强遮挡条件下仍存在漏检现象。造成上述问题的原因在于现有模型在多目标密集分布和局部特征模糊的情形下,特征提取与空间关注能力有限,难以有效区分关键目标区域。相较而言,SSPENet 在各类复杂场景中均展现出更优越的检测性能:在强遮挡区域,所引入的 SSAM 注意力机制能够聚焦于果实边缘及显著特征区域,提升对局部结构的敏感性;在果实密集重叠场景中,LAPM 模块通过金字塔式多尺度建模,有效缓解了检测框重叠及目标混淆问题;而在光照变化剧烈的条件下,模型仍能保持较高的检测稳定性与一致性。此外,SSPENet 模型在所有检测目标上的置信度分布均明显高于其他对比算法,表明其

具备更强的特征判别能力和更好的环境适应性,进一步验证了其在圣女果成熟度检测任务中的稳定性与实用性。

3 讨论与结论

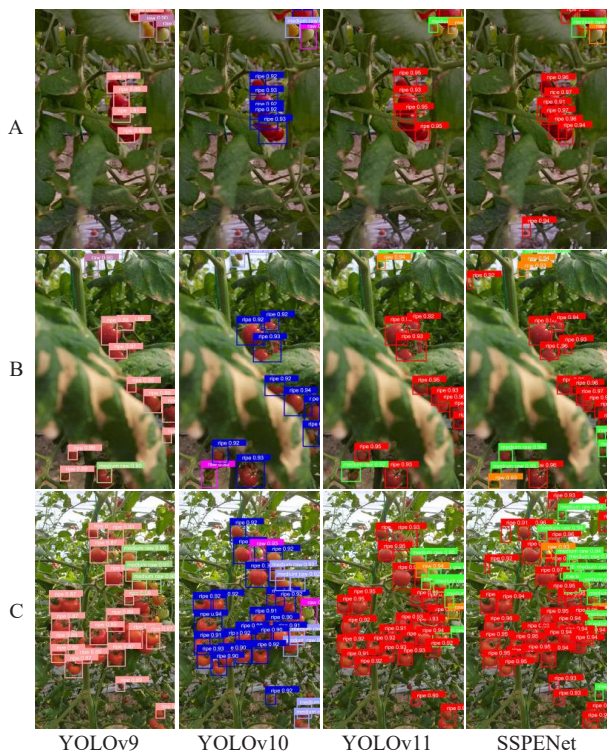
目前,果蔬图像识别研究多集中于目标检测模型的整体精度提升与计算效率优化,但在果实成熟度识别这一细分方向中,针对不同成熟阶段果实间细粒度颜色与形态特征的表达能力仍显不足。尤其在自然采集条件下,由于光照变化、枝叶遮挡、果实堆叠及背景复杂等因素干扰,现有模型在区分不同成熟度阶段果实时仍易出现误判,制约了模型在真实生产环境中的识别稳定性与通用性。为解决上述问题,笔者提出了一种 SSPENet 圣女果成熟度识别模型,在整体网络结构中引入空间注意力机制与多尺度特征融合模块,以提升模型对果实表面颜色分布、边缘特征及细粒度形态差异的表达能力。试验结果表明,SSPENet 模型在保持低计算复杂度的同时,平均识别精度达到 96.1%,较其他主流模型在成熟度分类精度方面更具优势。

从部署角度而言,SSPENet 推理速度为 94.7 FPS,可满足设施农业中对果实成熟状态实时识别的响应要求,适用于自动巡检机器人、移动采摘辅助终端等智能农机系统的集成部署。同时,模型在不同光照条件下依然表现出良好的鲁棒性,具备一定的推广应用潜力。尽管模型整体表现良好,但在复杂实际应用环境中仍存在以下几个关键问题需进一步解决:

(1)在果实密集分布、遮挡严重或背景干扰显著的场景中,模型易出现漏检或误检现象,影响整体识别准确性。后续将通过引入图像增强与边缘注意力机制,提升模型在复杂背景下的目标区域分离与边界表达能力,从而进一步提高目标定位精度。

(2)半成熟阶段果实的颜色呈连续过渡特征,模型在中间态分类判断上存在混淆。未来工作将通过增加该类样本数量并引入渐变性标签标注策略,构建多标签或弱监督分类结构,以提升模型对果实成熟度连续状态的建模能力。

综上所述,SSPENet 模型在识别精度、推理效率以及复杂环境适应性方面均表现良好,能够为圣女果采前成熟度分级、果实发育监测等环节提供可靠的图像识别技术支撑。研究结果为果蔬智能识别系统在设施农业中的落地应用提供了技术基础,



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

Note: A. A scene under backlighting conditions with overlapping fruits; B. A front-lit scene with occlusion caused by branches and leaves; C. A front-lit scene featuring overlapping fruits.

图 7 不同算法检测结果的可视化对比

Fig. 7 Visual comparison of detection results of different algorithms

具有较好的推广前景和较高的实用价值。

参考文献

- [1] 康晨, 吴珏, 姜安泽, 等. 贮藏温度对黄妃樱桃番茄果实品质的影响[J]. 中国食品学报, 2023, 23(4): 295-304.
- [2] 杨智超, 曹阳, 沈超怡, 等. 基于高压静电场处理的樱桃番茄果实贮藏期生理品质及其代谢[J]. 食品科学, 2021, 42(21): 168-176.
- [3] 王俊平, 徐刚. 机器视觉和电子鼻融合的番茄成熟度检测方法[J]. 食品与机械, 2022, 38(2): 148-152.
- [4] YIN H P, CHAI Y, YANG S X, et al. Ripe tomato extraction for a harvesting robotic system[C]//2009 IEEE International Conference on Systems, Man and Cybernetics, 2009: 2984-2989.
- [5] LIU L, LI Z K, LAN Y F, et al. Design of a tomato classifier based on machine vision[J]. PloS One, 2019, 14(7): e0219803.
- [6] 梁喜凤, 金超杞, 倪梅娣, 等. 番茄果实串采摘点位置信息获取与试验[J]. 农业工程学报, 2018, 34(16): 163-169.
- [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] 岳有军, 孙碧玉, 王红君, 等. 基于级联卷积神经网络的番茄果实目标检测[J]. 科学技术与工程, 2021, 21(6): 2387-2391.
- [9] 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.
- [10] 陈新, 伍萍辉, 祖绍颖, 等. 基于改进 SSD 轻量化神经网络的番茄疏花疏果农事识别方法[J]. 中国瓜菜, 2021, 34(9): 38-44.
- [11] WANG G, CHEN Y F, AN P, et al. UAV-YOLOv8: A small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios[J]. Sensors, 2023, 23(16): 7190.
- [12] QIU X Y, CHEN Y J, CAI W H, et al. LD-YOLOv10: A lightweight target detection algorithm for drone scenarios based on YOLOv10[J]. Electronics, 2024, 13(16): 3269.
- [13] 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.
- [14] 张磊, 刘琪芳, 聂红玫, 等. 基于改进 YOLOv4 网络模型的番茄果实检测[J]. 中国农机化学报, 2022, 43(12): 162-169.
- [15] LIU G X, NOUAZE J C, MBOUEMBE P L T, et al. YOLO-to-mato: A robust algorithm for tomato detection based on YOLOv3[J]. Sensors, 2020, 20(7): 2145.
- [16] 刘芳, 刘玉坤, 林森, 等. 基于改进型 YOLO 的复杂环境下番茄果实快速识别方法[J]. 农业机械学报, 2020, 51(6): 229-237.
- [17] ZENG T H, LI S Y, SONG Q M, et al. Lightweight tomato real-time detection method based on improved YOLO and mobile deployment[J]. Computers and Electronics in Agriculture, 2023, 205: 107625.
- [18] 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.
- [19] WU M N, LIN H R, SHI X R, et al. MTS-YOLO: A multi-task lightweight and efficient model for tomato fruit bunch maturity and stem detection[J]. Horticulturae, 2024, 10(9): 1006.
- [20] 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.
- [21] SHIM S H, HYUN S, BAE D, et al. Local attention pyramid for scene image generation[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 7774-7782.
- [22] 陈金荣, 许燕, 周建平, 等. 深度学习方法在红花采摘机器人中的应用[J]. 农机化研究, 2025, 47(4): 186-191.
- [23] 毛致颖, 刘宇航, 杨春勇, 等. 融合注意力及多重知识迁移的茶叶病害轻量化检测方法[J]. 中国农机化学报, 2024, 45(12): 140-147.
- [24] 谢忠坚, 廖珩宇, 文春明, 等. 基于改进 YOLOv5s 的蔗节检测方法[J]. 中国农机化学报, 2024, 45(12): 224-229.
- [25] 莫恒辉, 魏霖静. 基于改进 YOLOv7 的复杂环境下苹果目标检测[J]. 浙江大学学报(工学版), 2024, 58(12): 2447-2458.
- [26] 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.
- [27] 王鑫怡, 刘旭腾, 郑纪业, 等. 基于改进 YOLOv8 的轻量级鱼类检测方法[J]. 渔业现代化, 2024, 51(6): 91-99.
- [28] LIANG H, LI Y, ZHANG Q, et al. Do-DETR: Enhancing DETR training convergence with integrated denoising and RoI mechanism[J]. Multimedia Systems, 2025, 31(2): 171.