

张永宏,宋先鲁,李宇超,等. 基于 YOLO v8-DIS 算法的番茄果实成熟度检测[J]. 江苏农业科学,2025,53(5):46-54.

doi:10.15889/j.issn.1002-1302.2025.05.006

# 基于 YOLO v8-DIS 算法的番茄果实成熟度检测

张永宏<sup>1</sup>, 宋先鲁<sup>1</sup>, 李宇超<sup>1</sup>, 秦夏洋<sup>1</sup>, 董天天<sup>2</sup>, 郭明艳<sup>3</sup>, 季雁琛<sup>3</sup>

(1. 南京信息工程大学自动化学院, 江苏南京 210044; 2. 南京信息工程大学电子与信息工程学院, 江苏南京 210044;

3. 昆山市乐佳农业发展有限公司, 江苏苏州 215000)

**摘要:**针对 YOLO v8 算法模型在番茄果实成熟度检测过程中存在的漏检和误检问题,通过改进算法模型,使其能够精准分辨番茄果实成熟状态。首先采用可变形卷积(DCN)增加卷积核几何变换的内部机制,使骨干网络中的 C2f 结构专注于果实成熟度信息,其次采用带有缩放比例因子的辅助边框加速计算损失收敛,提升预测框实际尺寸精度。SiLU 函数引入了回归之间的向量角度,定义了新的惩罚因子。另外将注意力机制模块 LSKA 引入 Backbone 和 Neck 网络中间,调整不同成熟度细节信息的权重。由实验室采集的番茄果实数据集,使用数据增强技术对原始图像进行旋转、亮度增强、模糊等操作,避免训练过拟合。试验结果表明,改进后的算法模型对比原始算法,显著提高番茄果实成熟度的检测精度,算法精确率提升 2.4 百分点,平均精准度均值提升 1.2 百分点,同等试验条件下损失值降低,收敛速度更快。本研究的 YOLO v8-DIS 算法可以更好地分辨番茄果实的成熟状态,果实成熟度分辨性能相较于主流算法模型有较大提升,以期为目标检测在农业果实采摘的应用中提供技术支撑。

**关键词:**目标检测;YOLO;可变形卷积;番茄成熟度;注意力机制

**中图分类号:**S126;TP391.41 **文献标志码:**A **文章编号:**1002-1302(2025)05-0046-08

我国是世界上最大的番茄生产和消费国家之一<sup>[1]</sup>。2023 年番茄种植面积约为 111.3 万  $\text{hm}^2$ ,设施番茄种植面积占比近 60%,番茄产量占全球产量的 1/3 以上,其中番茄出口数量为 16.25 万  $\text{t}$ <sup>[2]</sup>。目前番茄果实的采摘仍然以人工采摘为主,成熟的番茄果实需要尽快采摘,劳动力不足和人工采摘劳动强度较大,使得机械化采摘作业正成为现代农业的必然趋势。番茄采摘机器人在采摘过程中由于番茄果实之间的相互遮挡、成熟度不同等作物特性构成了复杂的采摘环境因素,导致其难以精准识别定位。

智慧农业的兴起给现代果业带来新的发展机遇,以番茄、苹果、甜橙、西瓜等日常果蔬为对象的目标检测算法被大量应用<sup>[3]</sup>。在计算机视觉领域,单阶段识别算法和两阶段识别算法是 2 种不同的目标检测方法,两阶段算法的代表性算法有 R-CNN (region-based convolutional neural network)、Faster

R-CNN、Mask R-CNN 等<sup>[4-6]</sup>。其识别过程中,第 1 阶段通常是生成 1 组候选区域,第 2 阶段对这些候选区域进行分类和回归问题,然后得到最终的目标框和类别。两阶段算法在检测精度上具有优势。单阶段算法的代表性算法主要有 YOLO (you only look once)、SSD (single shot multibox detector) 等,它将目标检测任务视为直接从图像中提取目标框和类别的问题,所以在速度上相较于两阶段的识别算法有着较大的优势<sup>[7-8]</sup>。随着深度卷积神经网络目标检测技术的发展以及 Faster R-CNN、YOLO 系列等一些经典算法不断的更新开源,在农业检测应用中,由于单阶段检测算法的检测速度要优于两阶段算法,所以在农业检测中单阶段算法得到更多的实际应用<sup>[9-10]</sup>。冀汶莉等在 YOLO v5 的网络基础上将骨干网络替换成 PP-LCNet 轻量级网络,减少了模型参数量;在特征融合网络的末端添加基于标准化的注意力模块 (normalization-based attention module, NAM),增强模型对杂草和玉米幼苗的特征提取能力<sup>[11]</sup>。Li 等通过 ConvNext 对主干网络进行改进,提升检测棉花外来纤维的准确率,但检测速度相对较低<sup>[12]</sup>。Jia 等改进 Mask R-CNN 提高苹果在被遮挡环境下的识别精度,但是两阶段模型识别效率较低<sup>[13]</sup>。王小荣等在 YOLO v7 的基础上,

收稿日期:2024-03-22

基金项目:江苏省现代农业机械装备与技术示范推广项目(编号: NJ2022-02)。

作者简介:张永宏(1974—),男,山东莒南人,博士,教授,研究方向为方向模式识别与智能系统。E-mail:zyh@nuist.edu.cn。

通信作者:宋先鲁,硕士,研究方向为机器视觉、机器人。E-mail: songxianlu23@163.com。

增加 Swin Transformer 注意力机制提高模型对各分类样本的检测精准率,改进 Focal Loss 损失函数提升多分类任务下不均衡样本的识别率<sup>[14]</sup>。司永胜等在基于 YOLO v5s 模型的基础上,在 Backbone 中添加改进 RFB 结构,扩大特征提取感受野并对分支贡献度进行加权,更好地利用了不同尺度特征,采用 VariFocal Loss 损失函数,提高了模型对遮挡等场景下难样本的检测能力<sup>[15]</sup>。杨国亮等针对番茄检测精度低问题,在 YOLO v5s 主干网络末端添加 GAM 注意力模块,提升模型检测精度并改善鲁棒性;使用加权双向特征金字塔网络(BIFPN)修改原有结构,改善了番茄采摘时检测精度不高的情况,但是检测速率有所下降<sup>[16]</sup>。

因此,本研究在设施大棚环境下拍摄数据集,在 YOLO v8 模型基础上进行改进和优化,通过更换可变形卷积提升信息提取能力,更换不同的损失函

数,在合适的通道输出位置增加注意力机制提升对图像的特征提取能力,使模型能够更快更准确的识别不同成熟度的番茄果实,以期为番茄成熟度检测提供参考。

## 1 试验数据集

### 1.1 数据集采集

采集现代化种植基地的番茄图像数据,用于模型的训练和测试。本研究所使用的图像数据于 2023 年 10 月 7 日在江苏省昆山市陆家未来智慧果园 A + 温室工厂,分别在 1 d 内的不同时间段进行拍摄,拍摄角度模拟采摘机器人的视角,为了使模型在训练过程中能够更好地提取果实的表观特征,设备距离果实的距离为 30 ~ 40 cm,剔除不符合试验条件的图像,使用数据相机等移动设备共采集图像数据 804 幅,图 1 为在设施大棚内采集的图像示例。



图1 不同成熟度番茄果实

### 1.2 数据集处理

为保证模型的训练准确,提高模型的鲁棒性,加强模型对不同环境下番茄果实的识别能力,避免因数

据量不足导致模型过拟合,对图片数据分别进行亮度提高和降低、添加高斯噪声以及水平翻转等操作模拟不同光照环境和噪声环境<sup>[17]</sup>,增强效果见图 2。

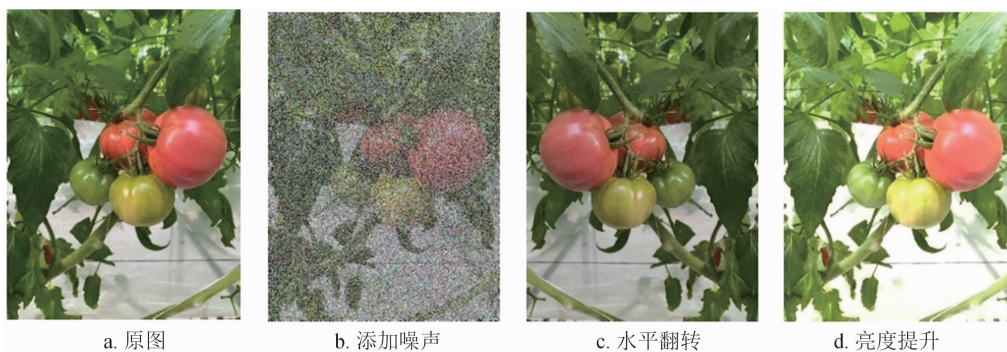


图2 番茄果实数据增强效果

增强之后的数据集为 1 608 幅图片,使用开源软件 LabelImg 进行标注,在专业种植人员指导下以及成熟度色卡的参考下,将标签类型分为成熟、灌浆期、完全绿色果实,生成对应的 YOLO 数据格式 xml 文件。数据集按照 7 : 2 : 1 的比例划分为训练

集、验证集、测试集。

## 2 网络模型的改进

### 2.1 YOLO v8 网络模型

YOLO v8 作为最新的 YOLO 模型,网络结构可

以分为骨干网络 (Backbone)、颈部网络 (Neck)、头部网络 (Head), 如图 3 所示。YOLO v8 在 YOLO v5 的基础上进行设计, 在原来的 C3 模块基础上进行重新设计, 结合 ELAN (efficient layer aggregation network) 思想得到新的 C2f 模块, 新的 C2f 模块让 YOLO v8 可以在特征提取阶段获得更丰富的梯度流信息。相较于 YOLO v5 的 SPP 特征金字塔池化结构, 基于 SPP 结构的 SPPF 被提出用于 YOLO v8 网

络, 在拥有更快速度的情况下, 有效避免了对图像区域裁剪、缩放而导致的失真问题。根据网络深度和宽度的不同, YOLO v8 可以分为如下几个版本, 分别为 YOLO v8n、YOLO v8s、YOLO v8m、YOLO v8l、YOLO v8x, 根据其不同的层数、计算量和模型体积, 分别有不同的应用场景。本研究的最终目标是将该算法部署到边缘计算的设备上, 考虑到算力的局限性, 因此采用模型体积较小的 YOLO v8n 为基线模型。

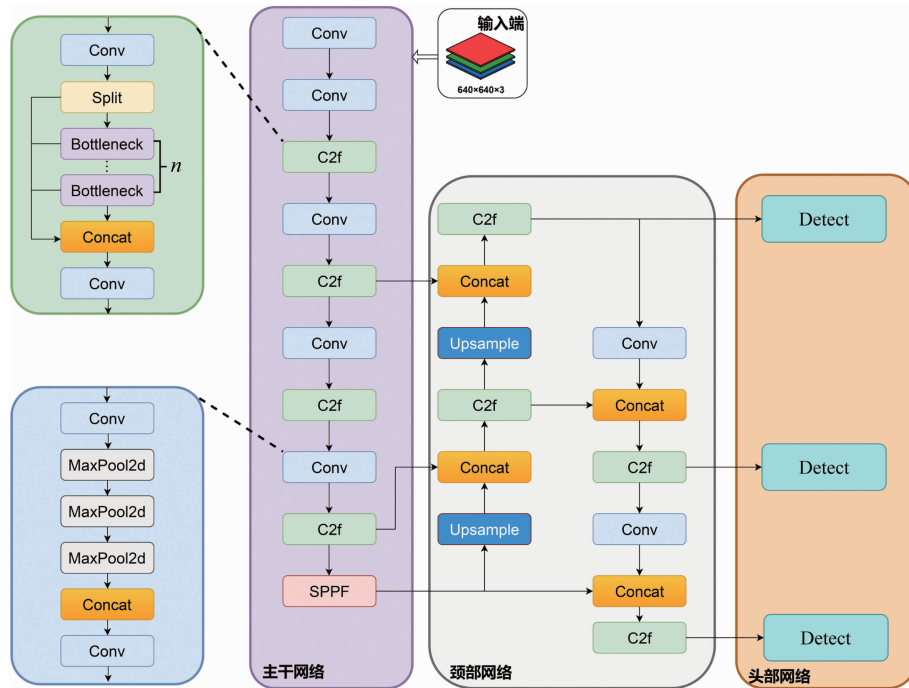


图3 YOLO v8 网络结构

## 2.2 深度可变形卷积

不同成熟度的番茄外观区别较大<sup>[18]</sup>。通过优化特征提取的方式提高识别的精准度。常规的卷积核因为固定的几何特征, 缺乏几何变换的内部机制。在其他不同的目标检测领域中, 固定不变的卷积核同样在应用过程中对于精度等指标有较大的影响。可变形卷积 (deformable convolution network, DCN) 系列算法的提出, 极大地提高了模型学习复杂目标的能力<sup>[19]</sup>。相较于传统卷积核固定不变的矩形, 卷积核在不同的场景和不同的阶段, 通过不规则的卷积核拟合目标的外观特征。最终学习出最优的卷积核形态, 能够极大地解决传统卷积特征提取能力不足的缺点, 使模型能够更好地学习数据集的特性。

可变形卷积在原标准卷积的基础上, 给每一个卷积采集点增加 1 个 2D 偏移量 ( $x, y$ ), 可变形卷积原理如图 4 所示。此卷积核的采集点有 9 个, 可变形卷积为这 9 个点都赋予 1 个 2D 偏移量, 所以卷积

核的位置和尺寸特征可以根据检测目标对象的表征进行相应的偏移, 使卷积层能够提取更多的图像特征。可变形卷积能够自适应学习感受野, 从而优化模型的结果。

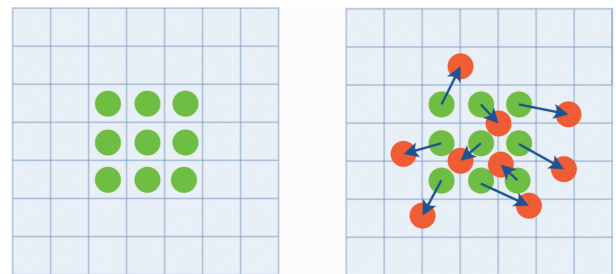


图4 深度可变形卷积原理

对于标准的卷积操作, 输出特征图  $y$  上的每个位置  $p_0$ , 其计算过程如公式 (1) 所示:  $w(p_n)$  为在  $p_n$  位置处的卷积核权重值大小;  $x$  为待检测的特征图;  $p_n$  代表卷积核采样区 ( $R$ ) 中的位置元素;  $(p_0 + p_n)$  代表待检测的采样位置。

$$y(p_0) = \sum_{p_n \in R} w(p_n) \cdot x(p_0 + p_n)。 \quad (1)$$

可变形卷积在传统卷积计算基础上对采样结果进行修改,增加的  $\Delta p_n$  为  $p_n$  的扩充量,为避免将无关信息进行计算,引入权重系数  $\Delta m_n$  进行无关信息抑制。输出特征值  $y(p_0)$  的计算如公式(2)所示:

$$y(p_0) = \sum_{p_n \in R} w(p_n) \cdot x(p_0 + p_n + \Delta p_n) \Delta m_n。 \quad (2)$$

YOLO v8 中 C2f 模块主要负责网络中的特征提

取工作,由于番茄果实不同成熟期的颜色和深度各不相同,使用传统矩形卷积在特征提取阶段容易遗漏,难以精确的识别细节特征。本研究对 C2f 模块进行改进, C2f 内部结构见图 5, 将其中 Bottleneck 模块中的传统卷积改进为可变形卷积 DCNv2, 来增强检测特征图中对番茄识别有用的信息,改进后的算法可以更好拟合不同生长期的番茄果实表征。

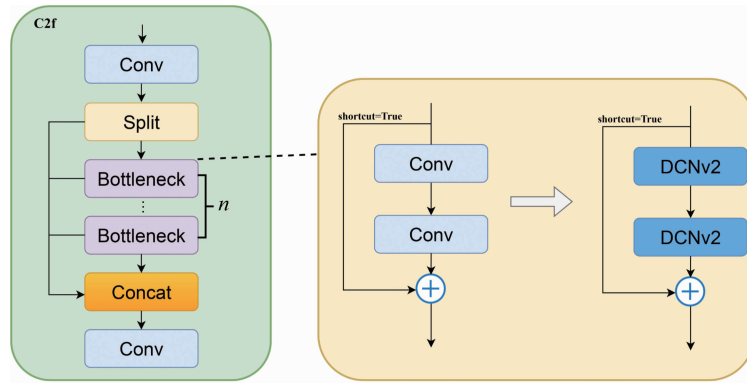


图5 改进 C2f 模块卷积方法

### 2.3 损失函数改进

对于 YOLO v8 的回归损失函数进行改进,将 CIoU 替换为 Inner (辅助边框生成器) - SIoU, Inner - SIoU 包括 SIoU 损失函数和辅助回归边框, YOLO v8 中 box - IoU 使用的函数为 CIoU (complete IoU loss), CIoU 在 DIoU 的基础上增加了形状损失, CIoU 损失计算如公式(3)所示:

$$L_{\text{CIoU}} = 1 - \text{IoU} + \frac{\rho^2(b, b^{\text{gt}})}{c^2} + \alpha v。 \quad (3)$$

式中:  $\text{IoU}$  为预测框和真实框的重合度,值越高说明重合程度越高;  $b$  和  $b^{\text{gt}}$  为预测边界框和真实边界框的中心点;  $\rho^2$  为  $b$  和  $b^{\text{gt}}$  2 个中心之间的欧氏距离;  $c$  为能同时覆盖预测边界框和真实边界框的最小矩形的对角线;  $\alpha$  为权衡参数,用于动态调整形状一致性项  $v$  的权重。虽然 CIoU 在 DIoU 的基础上做了相应的改进,但是都没有考虑到预测框与 GT 框之间不匹配的方向性回归问题,针对这个问题,本研究引入 SIoU 代替 CIoU,相比原 CIoU, SIoU 的损失函数值计算见公式(4), SIoU 引入了回归之间的向量角度,定义了新的惩罚因子包括  $\Delta$  和  $\Omega$  距离损失,改进后的 YOLO v8 在精准度方面有明显提升。

$$L_{\text{SIoU}} = 1 - \text{IoU} + \frac{(\Delta + \Omega)}{2}。 \quad (4)$$

Inner - IoU 是一种基于辅助边框的回归损失,

通过尺度因子  $ratio$  能够控制辅助边框尺寸从而加速计算损失收敛,原理见图 6, 其中红色边框为 Inner - IoU 生成的辅助边框,针对不同的数据集和检测器,调整不同尺度的辅助边框大小,从而增强模型的泛化性。Inner - IoU 实现公式如下:

$$\begin{aligned} g_l^{\text{gt}} &= x_c^{\text{gt}} - \frac{w^{\text{gt}} \times ratio}{2}, b_r^{\text{gt}} = x_r^{\text{gt}} + \frac{w^{\text{gt}} \times ratio}{2}, \\ b_t^{\text{gt}} &= y_c^{\text{gt}} - \frac{h^{\text{gt}} \times ratio}{2}, b_b^{\text{gt}} = y_b^{\text{gt}} + \frac{h^{\text{gt}} \times ratio}{2}, \\ b_l &= x_c - \frac{w \times ratio}{2}, b_r = x_c + \frac{w \times ratio}{2}, \\ b_t &= y_c - \frac{h \times ratio}{2}, b_b = y_c + \frac{h \times ratio}{2}, \\ inter &= [\min(b_r^{\text{gt}}, b_r) - \max(b_l^{\text{gt}}, b_l)] \times \\ &[\min(b_b^{\text{gt}}, b_b) - \max(b_t^{\text{gt}}, b_t)]; \\ union &= (w^{\text{gt}} \times h^{\text{gt}}) \times (ratio)^2 + (w \times h) \times \\ &(ratio)^2 - inter; \end{aligned}$$

$$\text{IoU}_{\text{inner}} = \frac{inter}{union}。 \quad (5)$$

式中:  $b_l, b_r, b_t, b_b$  分别代表预测框辅助边界的左、右、上、下边界坐标;  $x_c$  为预测框的中心点  $x$  坐标;  $w$  为预测框的原始高度;  $y_c$  为预测框的中心点  $y$  坐标;  $h$  为预测框的原始高度。  $inter$  为辅助框与辅助预测框的交集面积;  $union$  为辅助框与辅助预测框的并集面积;  $ratio$  为缩放因子,用来调整辅助框的尺寸。



与 IoU 损失相比,当缩放因子  $ratio$  小于 1 时,如图 6 所示,生成的辅助框小于原检测框,回归的有效范围小于 IoU 损失,能够加速高 IoU 更好收敛;相反当比例因子  $ratio$  大于 1 时,相对于大场景下的小

目标能够辅助进行更好的回归任务。因此在引入 SIoU 的同时,增加辅助边框应用于 SIoU,计算方式如公式(6)所示:

$$L_{\text{Inner-SIoU}} = L_{\text{SIoU}} + \text{IoU} - \text{IoU}_{\text{inner}} \quad (6)$$

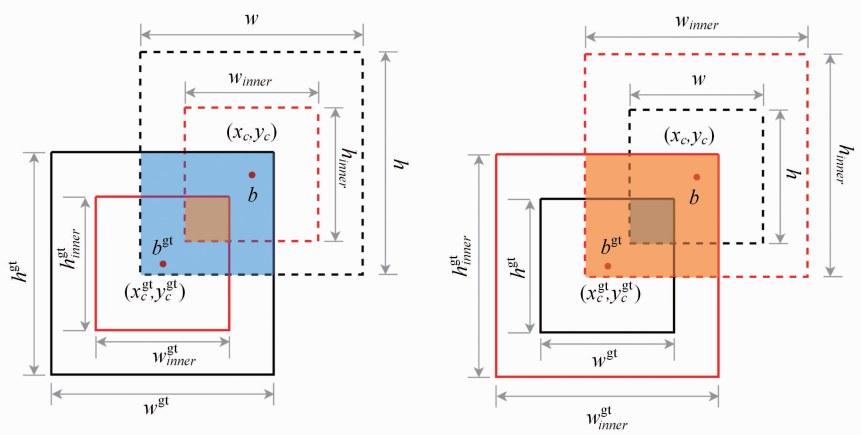


图6 Inner-IoU 原理

## 2.4 注意力机制

在番茄成熟度检测过程中,使用 YOLO v8 算法进行识别,由于灌浆期番茄同时具有其他 2 类成熟度果实的表征,所以在识别过程中容易出现识别错误的现象,这是由于骨干网络中的特征提取能力不足和信息筛选能力不足造成的结果。视觉注意力网络(VAN)具有大卷积核注意力(LKA)模块在一系列视觉检测的任务上表现出了卓越的性能,但是 LKA 模块中的深度卷积层随着卷积核大小的增加而导致计算和内存占用进一步的增加。

为了解决这些问题,并同时能继承 LKA 的优秀表现,Kin 等提出了一系列的 Large Separable Kernel Attention 模块<sup>[20]</sup>,改进如图 7 所示,称为 LSKA。公式(7)、(8)分别为 LKA 和 LSKA 的参数量计算方程。

$$\text{Param} = (2d-1)^2 \times C + \left(\frac{k}{d}\right)^2 \times C + C \times C; \quad (7)$$

$$\text{Param} = (2d-1) \times C + \left(\frac{k}{d}\right) \times C + C \times C. \quad (8)$$

式中: $C$  表示输入特征图的通道数; $k$  为核大小; $d$  为膨胀率。

LSKA 模块将 LKA 的前 2 层分解为 4 层结构,每层改成由 2 个 1D 卷积层组成,在保留了 LKA 稳定表现的同时,将计算量进行了降低,得到更高的计算效率和更小的内存占用空间。本研究引入 LSKA 注意力机制到骨干网络 SPPF 结构之后,在骨干网络提取到重要的图像信息后,进一步由注意力机制模块捕捉信息。此处改进在获得提升图像提

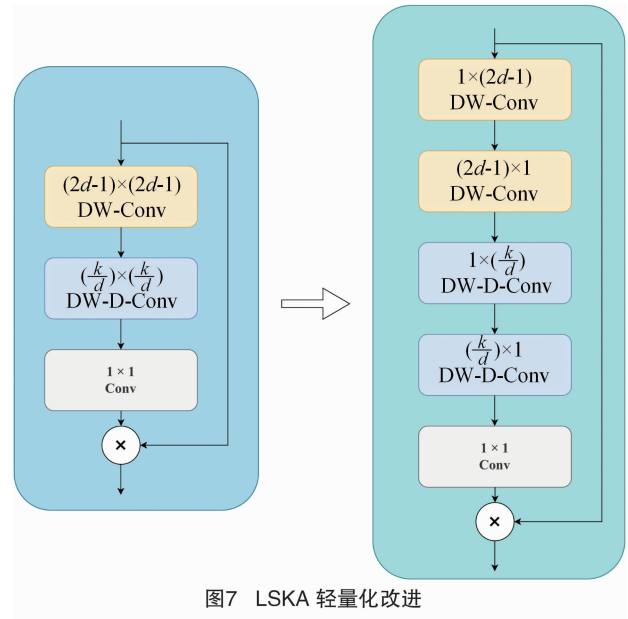


图7 LSKA 轻量化改进

取能力的同时,保证了结构的轻量化。

## 2.5 改进后的模型结构

模型主要改动如下:(1)将 C2f 中 Bottleneck 模块中的卷积更换为可变形卷积 DCNv2,并将其命名为 C2f-DCNv2,将原模型骨干网络中的第 3 个和第 4 个 C2f 模块替换成 C2f-DCNv2,增强模型的特征提取能力。(2)在骨干网络特征金字塔模块 SPPF 后添加注意力机制 LSKA,在 Neck 层前捕捉重要图像特征。(3)将原模型的 CIoU 损失函数更换为 Inner-SIoU,在增加角度惩罚项的基础上增加辅助检测框(图 8)。

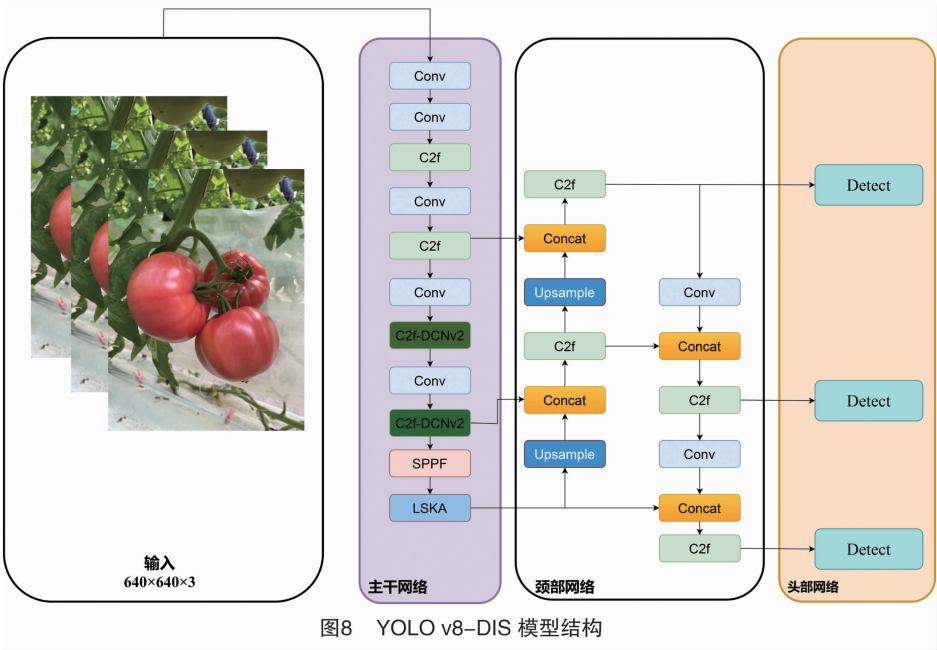


图8 YOLO v8-DIS 模型结构

3 试验结果与分析

为了验证本研究方法对模型提升的有效性,在相同试验环境下设计了 3 组试验:目标检测领域主流模型对比试验、消融试验、回归辅助边框的尺度试验。

3.1 试验环境及参数设置

试验时间为 2024 年 1—3 月,试验地点为南京市浦口区宁六路 216 号南京信息工程大学自动化学学院,服务器环境配置见表 1。在模型的训练过程中,输入的图像尺寸为 640 × 640 像素,初始学习率为 0.01,动量因子为 0.937,权重衰减因子为 0.000 5,批大小为 16,优化器为 SGD,迭代次数为 200。

表 1 环境配置

训练机配置名称	试验环境
操作系统	Windows 11
图形处理器	NVIDIA TITAN V(12 GB)
处理器	Intel® Core™ i9-9900K
内存	32 GB
Python	3.8
深度学习框架	Pytorch 1.13 + cuda 11.6

3.2 评估指标

本次试验采用精确率 (precision,  $P$ )、召回率 (recall,  $R$ )、平均精度均值 (mean average precision,  $mAP$ ),以及模型参数量作为模型的评估指标。精确率代表模型预测的数据集中预测正确的比例,该数值越高表示模型的检测效果越好。召回率指的是样本中的正例有多少被正确的预测出来,模型正确

判断出正例 ( $TP$ ) 占数据集所有正例的比例。

$P$ 、 $R$ 、 $mAP$  的计算公式见公式(9)至公式(11)。

$$P = \frac{TP}{TP + FP} \times 100\%; \quad (9)$$

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

$$mAP = \frac{\sum_{i=1}^K AP_i}{K} \times 100\%。 \quad (11)$$

式中: $TP$  代表正确识别番茄成熟度的数量; $FP$  代表值的是该样本是负样本但被预测为正样本个数; $FN$  代表本来是正样本但被预测为负样本的数量; $AP$  是精确率和召回率曲线下方的面积; $K$  为检测类别数量。

3.3 主流模型对比试验

为了验证本模型 YOLO v8-DIS 改进的有效性,设计了和当前主流目标检测模型的对比试验,采用统一数据集、试验环境分时和 YOLO v7-tiny、YOLO v5s、YOLO v7 模型以及原模型 YOLO v8n 进行了对比试验,试验结果见表 2。

表 2 主流模型对比试验

模型	精确率 (%)	召回率 (%)	平均精度均值 (%)	模型体积 (MB)
YOLO v7-tiny	75.6	74.3	80.1	11.70
YOLO v5s	81.8	78.9	85.4	13.70
YOLO v8	80.3	80.4	85.4	5.95
YOLO v8-DIS	82.9	80.7	86.6	6.59

对比结果显示,本研究算法 YOLO v8-DIS 在精确率方面相较于原始网络 YOLO v8 提升 2.4 百

分点,召回率基本持平,平均精度均值提升 1.1 百分点,模型体积仅增加 0.64 MB;和同等定位小体积的算法 YOLO v7 - tiny、YOLO v5s 相比,在模型体积更小的情况下,精确率分别高出 7.1、0.9 百分点,召回率提升 6.2、1.6 百分点;在模型体积方面分别降低 5.11、7.11 MB。可见本研究所提出的算法 YOLO v8 - DIS 目标检测网络相较于其他几种主流模型,在检测精度和模型体积方面有更加平衡的能力,同时在检测番茄之间相互遮挡问题时有更好的效果;模型体积虽然略高于原始模型 YOLO v8,但精确率和平均精度均值有明显提升。

### 3.4 辅助边框尺寸试验

本研究通过引入辅助回归边框优化模型,由尺度因子 *ratio* 控制辅助边框的大小,针对低 IoU 的番茄数据,将 *ratio* 比例因子设置为大于 1,生成大于原回归框的辅助边框,通过生成更大的辅助边框加速模型的收敛,提升模型的性能。不同的尺度因子会影响模型的性能,因此需要探究不同大小的尺度因子对于模型的影响。

表 3 为 Inner - SIoU 辅助边框不同大小的模型效果,通过对比明显看出,相较于原始算法的 CIoU 函数,在引入 Inner - SIoU 函数后,模型的检测精度有明显的提升,通过试验对比例因子调整,控制辅助回归边框大小,当比例因子为 1.30,即辅助边框为原始回归边框的 1.30 倍大小时,其模型的性能最优,精确率提升 2.7 百分点,平均精度均值提升 1.1 百分点。

表 3 不同尺度因子损失函数影响

损失函数	精确率 (%)	平均精度均值 (%)
原始 CIoU	80.2	85.4
Inner - SIoU(比例因子 = 1.25)	82.0	86.1
Inner - SIoU(比例因子 = 1.30)	82.9	86.5
Inner - SIoU(比例因子 = 1.35)	82.5	85.9

由图 9 可以看出,通过在验证集上测试,CIoU 经过 100 轮的迭代训练后,损失值最终稳定在 0.70 左右;更换损失函数为 Inner - SIoU,相较于原损失函数值明显下降,当比例因子为 1.30 时效果最优,最终稳定在 0.63;优化损失函数后选择合适的比例因子,Inner - SIoU 使模型收敛速度更快并且损失值更低。

### 3.5 模块消融试验

设计消融试验,对每个改进策略分别试验证明对模型性能的影响。DCNV、LSKA、Inner - SIoU 分别代表了利用可变形卷积、LSKA 注意力机制、带辅

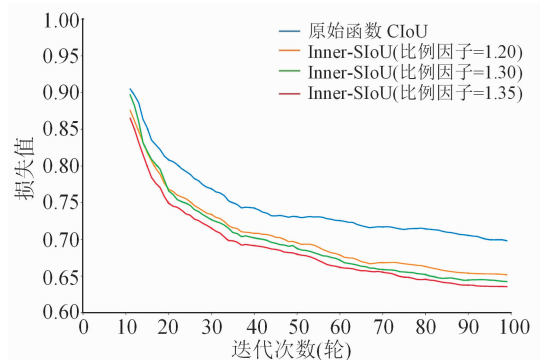


图9 不同比例因子损失函数曲线

助边框的 SIoU 损失函数改进基线模型。

由表 4 可以看出,与基线模型 YOLO v8 相比,仅使用可变形卷积、LSKA 注意力机制和带有辅助边框的 SIoU 损失函数,精确率提升 1.2 ~ 1.8 百分点,平均精度均值提升 0.3 ~ 0.5 百分点,每一种改进方法都有其特定的性能提升。在同时使用 2 种改进方法的试验中,同时使用可变形卷积和损失函数使精确率提升 2.2 百分点,平均精度均值提升 0.9 百分点。本研究提出的 YOLO v8 - DIS 模型与 YOLO v8 相比,精确率提升 2.4 百分点,平均精度均值提升 1.2 百分点,而导致模型体积的略微增长可以忽略不计。

表 4 消融试验

基线模型	DCNV	LSKA	Inner - SIoU	精确率 (%)	平均精度均值 (%)	模型体积 (MB)
✓				80.3	85.4	5.95
✓	✓			82.0	85.9	6.06
✓		✓		81.5	85.7	6.46
✓			✓	82.1	85.9	5.95
✓	✓		✓	82.5	86.3	6.06
✓		✓	✓	82.3	86.2	6.46
✓	✓	✓		82.6	86.0	6.58
✓	✓	✓	✓	82.7	86.6	6.59

### 3.6 不同模型可视化检测

使用训练好的 YOLO v8 - DIS、YOLO v7 - tiny、YOLO v5s、YOLO v8 模型对设施环境下的不同状态番茄果实进行检测,得到的识别结果见图 10。在多个成熟果实检测以及单个成熟果实检测过程中,本研究 YOLO v8 - DIS 相较于基线 YOLO v8 模型检测效果更精确,检测精度优于其他模型,图 10 中第 3 列在更复杂、多果实、有遮挡的环境下,YOLO v8 出现了成熟果实的漏检情况,而 YOLO v8 - DIS 很好地解决了果实遮挡状况下识别精度低的问题。





图10 不同模型可视化检测对比

#### 4 结论

本研究针对设施环境下目标检测模型对多果、单果、遮挡果实成熟度识别精度低的问题,构建了不同状态下的番茄成熟度数据集,提出了 YOLO v8 - DIS 算法模型,试验表明,本研究方法精确率( $P$ )均值达到 82.7%,平均检测精度均值( $mAP$ )达到 86.6%,对不同状态下的番茄果实识别更准确,相较于原 YOLO v8 算法以及 YOLO v5、YOLO v7 等不同体积的模型,识别精度更高。首先针对 C2f 模块的原传统卷积方式进行改进,利用可变形卷积的原理改进 C2f 模块,用于更好地提取传统卷积注意不到的图像信息。试验表明,仅针对 C2f 改进后算法精确率( $P$ )提升,验证了可变形卷积能更好地注意图像信息。在骨干网络后增加 LSKA 注意力机制模块,提升模型的泛化能力。对原模型的损失函数进行改进,使用新的 SIoU 函数进行改进模型,引入角度惩罚因子,同时增加回归辅助边框加速模型的训练,达到更好的模型效果。试验结果表明,YOLO v8 - DIS 模型在精确率方面相较于原始网络 YOLO v8 提升 2.4 百分点,平均精度均值提升 1.2 百分点,模型体积仅增加 0.64 MB,提升检测效果的同时保证了模型轻量化,对提升采摘机器人的工作效率有重要的影响。

#### 参考文献:

[1] 霍建勇. 中国番茄产业现状及安全防范[J]. 蔬菜, 2016(6): 1 -

4.

- [2] 郭文忠. 聚焦番茄产业症结,调整生产结构布局[J]. 山西农业大学学报(自然科学版), 2023, 43(5): 1 - 2.
- [3] 刘东琴, 杨震, 邹超, 等. 陕西省现代化果园生产技术应用概述[J]. 农业技术与装备, 2022(9): 70 - 72.
- [4] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014: 580 - 587.
- [5] Girshick R. Fast R - CNN [C]//International Conference on Computer Vision. Santiago, 2015: 1440 - 1448.
- [6] He K, Gkioxari G, Dollar P, et al. Mask r - cnn [C]//Proceedings of the IEEE International Conference on Computer Vision. Venice, 2017: 2961 - 2969.
- [7] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real - time object detection [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 779 - 788.
- [8] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector [C]//Computer Vision - ECCV 2016. Cham: Springer International Publishing, 2016: 21 - 37.
- [9] Redmon J, Farhadi A. YOLO v3: an incremental improvement [EB/OL]. (2018 - 04 - 08) [2024 - 01 - 22]. <https://arxiv.org/abs/1804.02767v1>.
- [10] Wang C Y, Bochkovskiy A, Liao H M. YOLO v7: trainable bag - of - freebies sets new state - of - the - art for real - time object detectors [C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver: IEEE, 2023: 7464 - 7475.
- [11] 冀汶莉, 刘洲, 邢海花. 基于 YOLO v5 的农田杂草识别轻量化方法研究[J]. 农业机械学报, 2024, 55(1): 212 - 222, 293.



黄蕊,李贤辉,杨彦鑫,等. 基于 FEACConv-YOLO v5 的柑橘果实目标检测与识别研究[J]. 江苏农业科学,2025,53(5):54-61.  
doi:10.15889/j.issn.1002-1302.2025.05.007

# 基于 FEACConv-YOLO v5 的柑橘果实目标检测与识别研究

黄蕊<sup>1</sup>,李贤辉<sup>1</sup>,杨彦鑫<sup>1</sup>,李文峰<sup>1</sup>,黄兆波<sup>1</sup>,魏永葵<sup>2</sup>,刘宏斌<sup>3</sup>

(1. 云南农业大学机电工程学院,云南昆明 650201; 2. 华宁县农村经济经营管理站,云南华宁 652899;  
3. 华宁县柑桔科学研究所,云南华宁 652899)

**摘要:**为解决现阶段柑橘生育期信息的获取主要依靠人工观测的效率低、主观性强等问题。首先提出一种基于 FEACConv-YOLO v5 的柑橘果实目标检测算法,提出 FEConv 与 FEACConv 等 2 种特征增强卷积核,替代 YOLO v5 网络中卷积核,然后进行消融试验验证各改进点的有效性,再与 Faster-RCNN、YOLO v4、SSD 等 3 个网络模型进行检测能力的对比,最后用测试集验证改进后的模型性能。其次,设计一种融合大气环境特征与果实表型特征的多模态特征果实生育期识别模型,利用 ResNet18 作为 CNN 的分类网络,提取柑橘表型特征;利用 ANN 作为大气环境数据特征提取网络,对 FEACConv-YOLO v5 所检测到的柑橘目标进行分类识别,并用混淆矩阵评价模型的性能。结果表明,改进后的目标检测模型收敛效果好,损失值比 YOLO 算法更低,消融试验中将 FEConv 替代主干部分卷积核后,准确率、召回率和平均精度均值分别较原模型提高 0.9、0.6、0.7 百分点;将 FEACConv 替代 Neck 部分卷积核后,准确率、召回率和平均精度均值分别提高 1.4、1.0、1.5 百分点;同时融入 FEConv 与 FEACConv 后,准确率、召回率和平均精度均值分别提高 2.1、2.2、2.7 百分点。在对比试验中,改进模型的平均精度均值达到了 90.3%,检测速率达到了 2.72 帧/ms,在 4 个网络模型里面显示最优。其次,生育期识别模型中幼果期分类精度为 96.7%,果实膨大期的分类精度为 93.6%,果实成熟期的分类精度为 100%,平均分类精度为 96.8%。FEACConv-YOLO v5 网络可以在满足时间要求的前提下,有效提升检测精度,减少小柑橘目标与遮挡目标的漏检问题。同时,多模态融合的果实目标发育期识别网络具有较高的分类精度。

**关键词:**柑橘果实;目标检测;YOLO;生育期识别

**中图分类号:**S126;TP391.41 **文献标志码:**A **文章编号:**1002-1302(2025)05-0054-08

近年来,云南省玉溪市华宁县的早熟柑橘市场

发展较快,为当地取得了良好的经济效益和社会效益,在有效缓解当地淡季供需矛盾的同时,可满足消费者对新鲜柑橘果品的周年需求。然而,早熟柑橘由于生育期短、成熟早,造成果实甜度普遍较低等问题,所以对柑橘的生育期进行监测尤为重要,另外,柑橘的产量和果实品质均取决于各生育

收稿日期:2024-03-23

基金项目:云南省科技重大专项(编号:202202AE090021)。

作者简介:黄蕊(1996—),女,云南曲靖人,硕士研究生,从事柑橘水肥灌溉研究。E-mail:2353403547@qq.com。

通信作者:黄兆波,硕士,副教授,主要从事果树水肥一体化灌溉研究。E-mail:896816678@qq.com。

[12] Li Q X, Ma W J, Li H, et al. Cotton-YOLO: improved YOLO v7 for rapid detection of foreign fibers in seed cotton[J]. Computers and Electronics in Agriculture, 2024, 219: 108752.

[13] Jia W K, Tian Y Y, Luo R, et al. Detection and segmentation of overlapped fruits based on optimized mask R-CNN application in apple harvesting robot[J]. Computers and Electronics in Agriculture, 2020, 172: 105380.

[14] 王小荣,许燕,周建平,等. 基于改进 YOLO v7 的复杂环境下红花采摘识别[J]. 农业工程学报, 2023, 39(6): 169-176.

[15] 司永胜,孔德浩,王克俭,等. 基于 CRV-YOLO 的苹果中心花和边花识别方法[J]. 农业机械学报, 2024, 55(2): 278-286.

[16] 杨国亮,王吉祥,聂子玲. 基于改进型 YOLO v5s 的番茄实时识

别方法[J]. 江苏农业科学, 2023, 51(15): 187-193.

[17] 王昱,姚兴智,李斌,等. 基于改进 YOLO v7-tiny 的甜椒畸形果识别算法[J]. 农业机械学报, 2023, 54(11): 236-246.

[18] 宋梦圆,宁国法,田永强,等. 不同成熟度番茄果实在贮藏期间的品质变化[J]. 中国蔬菜, 2023(2): 45-50.

[19] Zhu X Z, Hu H, Lin S, et al. Deformable ConvNets v2: more deformable, better results[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019: 9300-9308.

[20] Lau K W, Po L M, Rehman Y A U. Large separable kernel attention: rethinking the large kernel attention design in CNN[J]. Expert Systems with Applications, 2024, 236: 121352.