

DOI: 10.13718/j.cnki.xdzk.2026.05.018

胡俊杰, 张聪, 陶章法, 等. 番茄叶片病害多目标识别模型的轻量化研究 [J]. 西南大学学报(自然科学版), 2026, 48(5): 222-234.

番茄叶片病害多目标识别模型的轻量化研究

胡俊杰¹, 张聪¹, 陶章法², 梁红蕊², 张瑞特¹, 王正³

1. 武汉轻工大学 电气与电子信息学院, 湖北 武汉 430048; 2. 武汉轻工大学 数学与计算机学院, 湖北 武汉 430048;
3. 武汉大学 计算机学院, 湖北 武汉 430072

摘要: 针对现有番茄叶片病害识别模型存在检测目标少、计算复杂度高的问题, 提出了一种结合复合卷积的目标识别改进模型。在骨干网络中对 FasterNet Block 进行改进, 提出了轻量级多尺度特征提取模块(Dil-FasterNet block), 通过结合空洞卷积(Dilated Conv, DConv)和深度可分离卷积(Depth-Separable Convolution, DSConv), 在减少计算量的同时增强局部细节, 提升捕捉图像细节信息的能力。此外, 引入分组混洗卷积(Group-Shuffle Convolution, GSConv)替换原模型卷积, 在颈部网络引入 Slim-Neck 模块, 利用 GS 瓶颈和跨阶段部分网络(Cross Stage Partial Net, CSPNet)模块控制模型时间复杂度, 并且保留每个通道的隐藏连接。最后, 设计了 FocalWise-IoU 作为损失函数, 加强检测模型对中质量锚框的预测, 减少模型对高质量锚框的过度优化, 同时保留低质量锚框产生的信息。实验结果表明: 改进模型在数据集上的浮点运算量和模型参数量分别下降了 2.1 G 和 1.1 M, 同时平均精度均值提高了 0.5 个百分点, 权值文件大小降低了 27%。

关键词: 番茄叶片病害; 分组混洗卷积; 非单调动态聚焦机制;
多目标识别; 深度可分离卷积

中图分类号: TP391.41

文献标识码: A

开放科学(资源服务)标识码(OSID):



文章编号: 1673-9868(2026)05-0222-13

Lightweight Study of a Multi-Target Recognition Model for Tomato Leaf Diseases

HU Junjie¹, ZHANG Cong¹, TAO Zhangfa²,
LIANG Hongrui², ZHANG Ruite¹, WANG Zheng³

1. School of Electrical and Electronic Engineering, Wuhan Polytechnic University, Wuhan Hubei 430048, China;
2. School of Mathematics and Computer Science, Wuhan Polytechnic University, Wuhan Hubei 430048, China;
3. School of Computer Science, Wuhan University, Wuhan Hubei 430072, China

Abstract: Addressing challenges such as limited detection targets and high computational complexity in

收稿日期: 2025-10-05

基金项目: 湖北省技术创新重大项目(2018A01038)。

作者简介: 胡俊杰, 硕士研究生, 研究从事计算机视觉、目标检测的研究。

通信作者: 张聪, 博士, 二级教授, 硕士研究生导师。

existing tomato leaf disease recognition models, this paper proposed an improved target recognition model integrating composite convolutions. For the backbone network, the Dil-FasterNet block was introduced by modifying the FasterNet Block. This lightweight, multi-scale feature extraction module combines Dilated Convolution and Depth-Separable Convolution to enhance local details while reducing computational overhead, improving the capture of image details. Additionally, Group-Shuffle Convolution replaced the original convolutional layers. For the neck component, the Slim-Neck module was adopted as the neck network. GS bottleneck and the Cross Stage Partial Net (CSP Net) module were utilized to control the temporal complexity of the model while preserving hidden connections of each channel. FocalWise-IoU was designed as the loss function to strengthen the prediction of medium-quality anchor boxes by the detection model, reduce the over-optimization of high-quality anchor boxes, and retain the information from low-quality anchors. Experimental results demonstrated that the improved model achieved a 2.1 G reduction in FLOPs and a 1.1 M decrease in parameters on the dataset, while simultaneously increasing the mAP by 0.5% and reducing the weight file size by 27%.

Key words: tomato leaf diseases; GSConv; non-monotonic dynamic focusing mechanism; multi-target recognition; depth-separable convolution

近年来,以卷积神经网络为核心的深度学习技术在计算机视觉领域取得了突破性进展,在病害特征提取、分类识别与目标定位任务中具备远超传统人工识别方法的精度与鲁棒性,为作物病害智能检测提供了全新的技术方案^[1]。随着智能农业的快速发展,利用深度学习技术对番茄叶片病害进行实时检测可显著提升番茄生产效率。

当前,国内外主流的基于深度学习的目标识别模型有单阶段目标检测算法和双阶段目标检测算法。单阶段目标检测算法包括 SSD^[2]、YOLO^[3]等,其中 YOLO 模型因检测速度快、准确度高而备受关注^[4]。然而,在农田的叶片病害检测场景中,叶片病害目标普遍存在小目标占比较高、背景复杂以及现有识别模型难以兼顾检测精度与模型轻量化的问题。

针对上述问题,本研究围绕番茄叶片病害识别任务设计了 Dil-YOLO(YOLOv8n-DilatedConv-Slim-Neck)番茄叶片病害识别模型,从 3 个方面进行改进:①在 YOLOv8 主干网络设计了 Dil-FasterNet 模块,替换原 C2f 模块,该模块将标准卷积拆分为单通道空间特征提取的深度卷积(Depth Wise Convolution, DWConv)与跨通道特征融合的逐点卷积(Point Wise Convolution, PWConv),在不破坏特征传递逻辑的前提下大幅降低单卷积层的参数量与浮点运算量;②引入高效特征颈部网络 Slim-Neck 模块^[5],该模块通过分组卷积(Grouped Convolution, GConv)与通道混洗操作的协同作用,在降低卷积运算参数量与计算复杂度的同时,保留了通道间的特征交互能力;③设计了 FocalWise-IoU 损失函数,结合 Wise-IoU 函数的动态聚焦机制^[6]和 Focal-EIoU v1 函数的动态权重分配机制^[7],加强对低质量锚框的识别能力。通过上述改进整体提升番茄叶片病害识别的准确性与鲁棒性。

1 材料与实验环境

1.1 数据集材料

本研究使用的复杂环境下番茄叶片数据集采集于陕西临潼番茄产业园,该产业园具有独特的地理位置和多样的气候条件,为获取丰富多样的番茄叶片病害数据集提供了理想环境^[8]。在不同生长期(发芽期、幼苗期、开花期和结果期)及不同场景(图 1)下对番茄叶片图像进行了采集。图像采集相关信息如表 1 所示。



图 1 图片采集场景示例

表 1 采集信息

名称	数据	名称	数据
采集地点	陕西临潼番茄产业园	拍摄距离	相机与叶片距离保持 15~50 cm
采集设备	Nikon z30、Nikon z50	拍摄角度	正视(0°)、侧视(45°)、仰视(90°)
图像分辨率	2 784×1 856、4 176×2 344	图片数量	3 451 张

由于番茄发芽期相对较短,而幼苗期和开花期持续时间长,感染病害概率大,且需在结果期前进行病害检测,故本数据集主要选取幼苗期和开花期的图像进行实验。该数据集包括 3 451 张番茄叶片病害图片,并且使用 LabelImg 工具标记 3 723 个手工数据框,由研究团队采用“初标-校验-复验”三级质控模式,对数据集进行标注后划分为训练集(2 039 张)、验证集(572 张)、测试集(840 张),从而保证训练和测试数据的一致性和准确性。数据集划分情况如表 2 所示。

表 2 数据集分类体系

生长阶段	图像数量/张		
	训练集	验证集	测试集
幼苗期	1 008	320	519
开花期	1 031	252	321
合计	2 039	572	840

1.2 YOLOv8 网络结构

本研究将 YOLOv8 系列模型参数进行了对比(表 3),其中 *FLOPs* (Floating-point Operations)为模型浮点运算量,*Params* 为模型参数量。

表 3 YOLOv8 模型参数对比

模型	Params/M	FLOPs/G	模型	Params/M	FLOPs/G
YOLOv8n	3.2	8.2	YOLOv8l	43.7	165.2
YOLOv8s	11.2	28.6	YOLOv8x	68.2	257.8

如图 2 所示, YOLOv8 结构分为 3 部分: 骨干网络、颈部网络和探测头。其中骨干网络为基于 C2f 模块构建的轻量化卷积网络, 常见的包括残差网络系列(Residual Network, ResNet)^[9] 和更高效的 EfficientNet^[10] 等。颈部网络的一个显著特点是增强了多尺度特征融合的能力。通过引入路径聚合网络(Path Aggregation Network, PANet)^[11] 等先进的特征融合技术, YOLOv8 能够更好地将低层和高层特征信息结合起来, 提升小物体和远距离物体的检测精度。本研究内容为小目标检测, 以尽可能轻量化为目的, 故采用 YOLOv8n 作为原始模型结构进行优化^[12]。

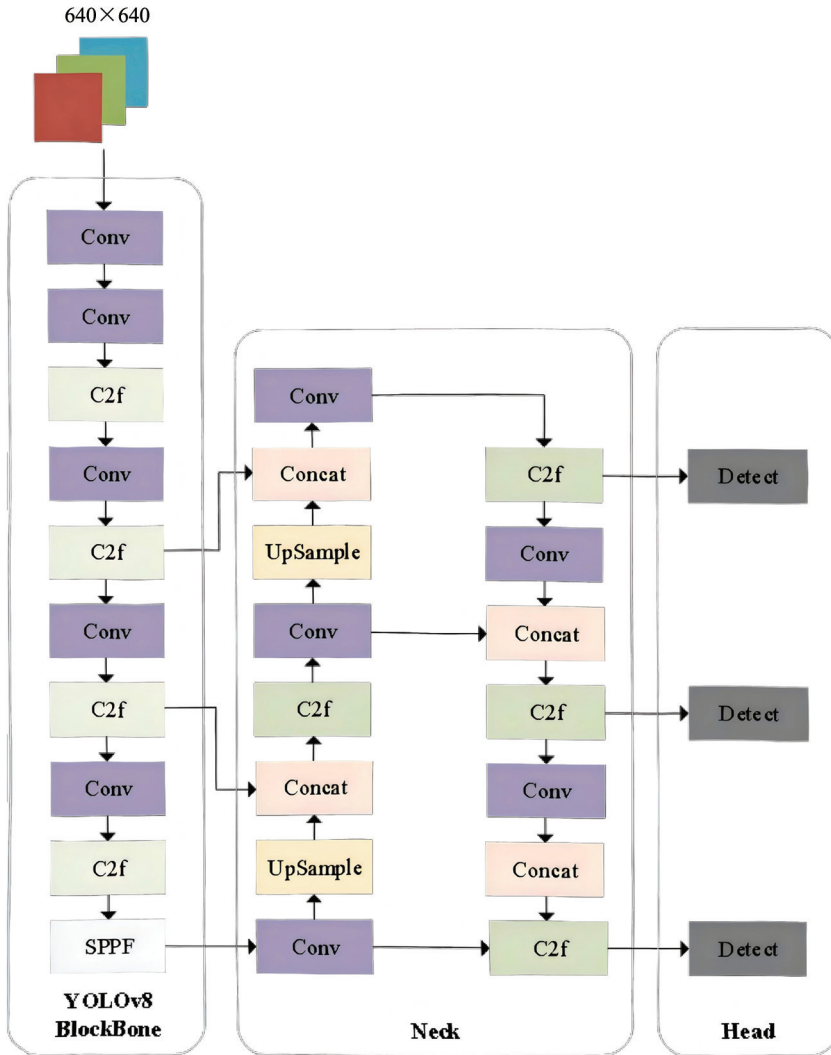


图 2 YOLOv8 结构示意图

1.3 改进的 YOLOv8n 网络

为了进一步提高模型的识别性能, 本研究基于 YOLOv8n 框架设计了 Dil-YOLO 模型(图 3)。在原模型的特征提取网络设计了新的轻量级骨干网络 Dil-FasterNet。该网络通过优化 FasterNet 模块^[13] 显著降低了模型参数量, 适用于轻量化部署环境下的识别任务。此外, 本研究在颈部网络部分引入了 Slim-Neck 空间深度卷积结构, 提高了模型的响应速度和运行能力。最后, 设计了 FocalWise-IoU 损失函数, 通过结合

Wise-IoU 的非单调动态聚焦机制和 Focal-EIoU v1 的动态权重分配机制, 减小低质量产生的异常梯度, 按梯度进行增益分配。这些设计不仅提高了模型的运行速度, 还降低了模型的存储开销, 为番茄叶片病害识别模型的轻量化提供了有效的技术支持。

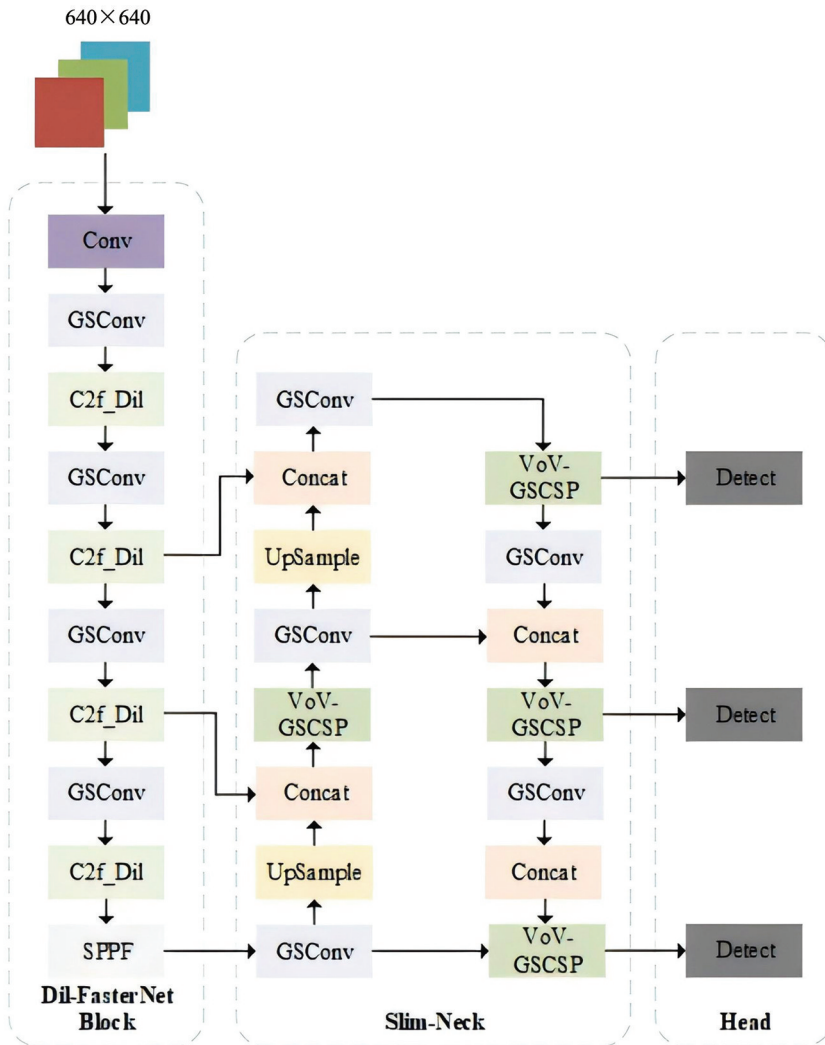


图 3 改进后模型结构图

1.3.1 Dil-FasterNet 模块

为了使农田环境下部署的检测模型在保持检测精度的情况下尽可能轻量化, 本研究设计了 Dil-FasterNet 模块。

目前优化后的网络模块, 例如空间和通道重构卷积(Spatial and Channel Reconstruction Convolution, SCConv)^[14]虽然提升了检测性能, 但由于其通道重构单元(Channel Reconstruction Unit, CRU)和空间重构单元(Spatial Reconstruction Unit, SRU)模块需要额外的全连接层或卷积层来生成权重, 增加了模型参数量, 导致训练时显存占用量显著增加, 需要引入新的网络部件进行补偿。在保持高速有效检测的情况下, Ren S 等设计出快速网络架构 FasterNet^[13], 引入部分卷积(Partial Convolution, PConv)来代替基础卷积。PConv 通过在部分输入通道中使用常规的 Conv 进行空间特征提取来减少冗余。PConv 的 $FLOPs$ 计算公式如下:

$$FLOPs = h \times w \times k^2 \times c_p^2 \quad (1)$$

式中: h 为输入图片高度; w 为输入图片宽度; k 为卷积核边长; c_p 为输入输出通道。FasterNet 在提高浮点运算速度的同时有效降低了 $FLOPs$, 从而显著提升了目标检测的速度。

在此基础上,为进一步提高模型特征提取能力,在原 FasterNet 模型的特征转换器 MLP 中设计了 Dil-FasterNet 模块。该模块利用深度可分离卷积(Depth-Separable Convolution, DSConv)将标准卷积分解为 DWConv 和 PWConv 进行维度转换,同时利用空洞卷积(Dilated Conv, DConv)^[15]增强特征提取能力。

Dil-FasterNet 在自然农田番茄叶片病害识别环境中具备两大优势。一是其使用的 DWConv 对每个输入通道单独滤波, PWConv 仅用 1×1 卷积融合通道,参数量远少于标准卷积。虽然计算量减少,但通过分离空间滤波和通道融合,仍然能有效地提取特征。二是由于扩张卷积直接在高分辨率特征图上进行操作,模型网络可在保持空间细节的同时捕捉更大范围的上下文信息,从而减少有效信息的丢失。本研究 Dil-FasterNet 模块如图 4 所示。

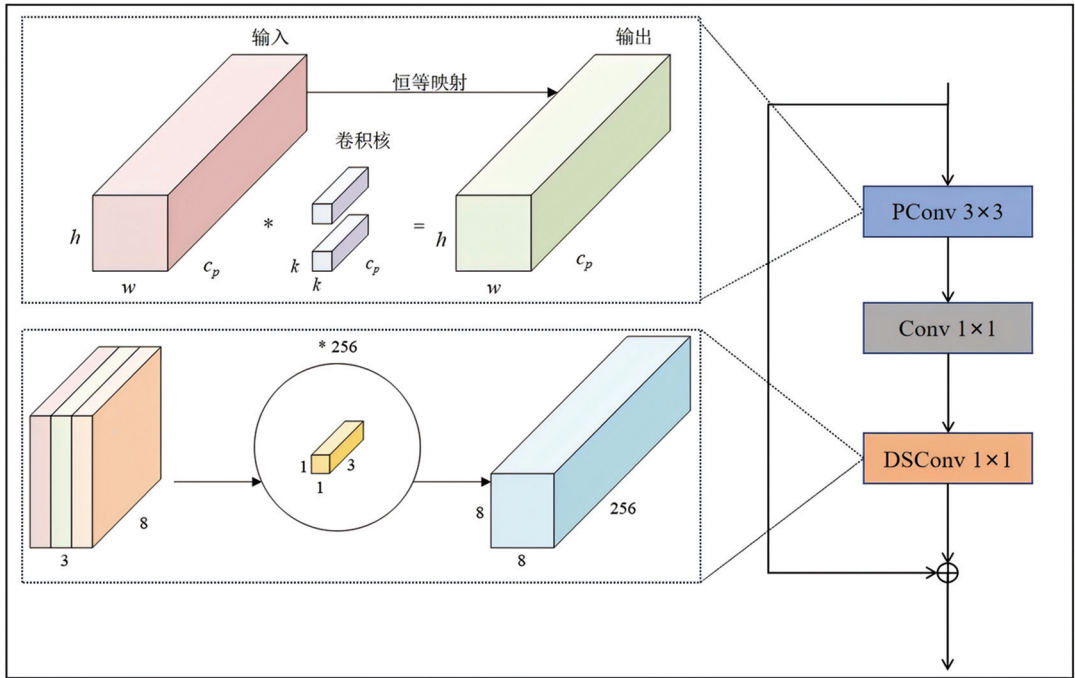


图 4 Dil-FasterNet 模块图

1.3.2 特征融合模块

为了加强物体边缘的信息捕捉能力,本研究引入了 Slim-Neck 模块。该模块通过优化特征图的空间表示和通道间的交互,提升了对图像边界框特征的捕捉能力,同时通过交互学习增强了对不同物体类别的区分。该模块基本原理如下:

分组混洗卷积(Group-Shuffle Convolution, GSConv)提高了模型图像预测能力。常规的卷积神经网络将空间信息转化为通道信息的过程中易发生语义内容丢失的情况。GSConv 能保持通道之间的隐藏连接,且不会增加模型的时间复杂度^[15],其计算原理如下:

GSConv 全局跨通道信息融合的时间复杂度为:

$$O(H \times W \times C_{in} \times C_{out} \times K^2) \quad (2)$$

GSConv 在稀疏部分通过稀疏化来降低计算量:

$$O(\alpha \times H \times W \times C_{in} \times C_{out} \times K^2) \quad (3)$$

GSConv 的总时间复杂度为:

$$O(H \times W \times C_{in} \times C_{out} \times (1 + \alpha K^2)) \quad (4)$$

其中: H 、 W 、 C_{in} 为输入特征图尺寸; α 为稀疏率; C_{out} 为输出通道; $K \times K$ 为卷积核大小。当稀疏策略仅保留局部关键点时, α 会显著降低。GSConv 模块结构如图 5 所示。

在 GSConv 之后,继续引入由 GS 瓶颈和跨阶段部分网络(Cross Stage Partial Network, CSPNet)组成

的多视角-分组混洗跨阶段部分模块(Variety of Views-Group Shuffle Cross Stage Partial, VOV-GSCSP)。GS 瓶颈通过稀疏注意力机制和动态稀疏化机制, 仅对重要区域和通道进行密集计算, 在不显著增加计算量的前提下扩展有效识别视野、减少模型冗余计算。

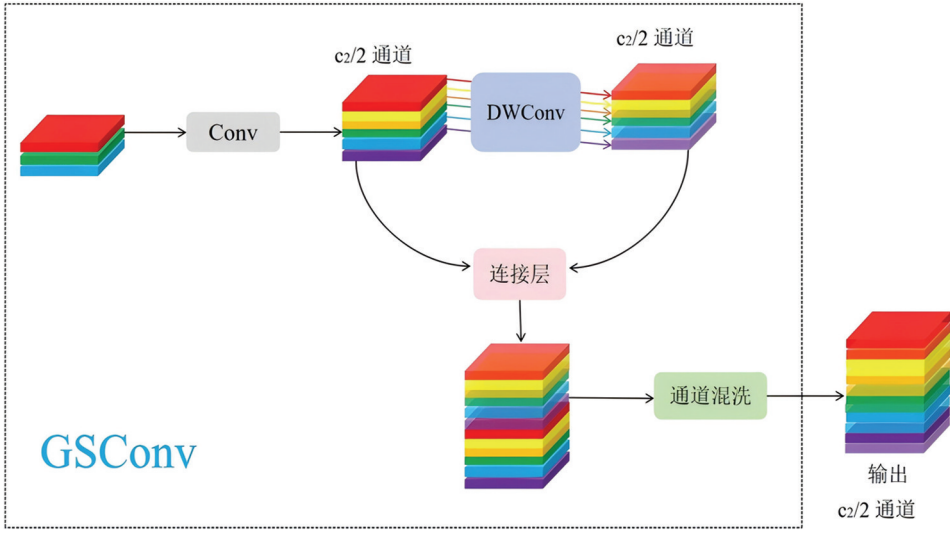


图 5 GSConv 模块原理图

Slim-Neck 模块在农田环境中如下优势: 不同尺度的特征图(P3、P4、P5)先通过 GSConv 模块处理, 然后通过上采样 VpSample 和拼接(Concat)操作与其他尺度的特征图结合, 并再次通过 GSConv 模块, 最后使用 VOV-GSCSP 模块进一步提取和融合特征, 最后检测头对接收到的特征图(P3、P4、P5)进行检测。通过以上模块化和分层的方法, Slim-Neck 架构能在保持高准确度的同时减少计算量和推理时间。

1.3.3 损失函数的改进

目标的检测与识别是深度学习的重点问题之一, 其效果与损失函数性能密切相关。目前大多数的边界框回归函数(如 IoU、SIoU^[16]、原模型 GIoU^[17]和 DIoU^[18])会忽略训练集中存在的低质量示例, 从而导致低质量示例产生的异常梯度影响算法模型的识别能力。Focal-EIoU v1 被用于解决此问题, 但由于聚焦机制不是动态的, 现有研究对于非单调聚焦机制的内在机理和潜在效能未进行充分挖掘。本研究结合非单调的动态聚焦机制 Wise-IoU 和动态权重分配机制 Focal-EIoU v1 构建了 FocalWise-IoU 损失函数。FocalWise-IoU 损失函数的主要优势在于:

非单调动态聚焦机制使用“离群度”^[19]替代 IoU 对锚框进行质量评估, 并设计了更加可靠的按梯度进行增益的分配机制。该函数降低了一部分高质量锚框的竞争力, 并减小了低质量示例产生的异常梯度, 其定义如图 6 所示。

训练数据中的低质量样本常受几何因素例如距离和长宽比的影响, 这可能会影响模型的泛化能力。一般理想的损失函数应在锚框与目标框高度重合时减少几何因素对检测的干扰, 同时减少训练干预, 从而增强模型的泛化能力。基于此构建距离注意力, 并结合动态聚焦机制获得具有两层注意力机制的损失函数 $L_{WIoU-v1}$:

$$L_{WIoU-v1} = R_{WIoU} L_{IoU} \tag{5}$$

式中: R_{WIoU} 是距离注意力权重; L_{IoU} 是标准损失函数, 为基础 IoU 损失。 R_{WIoU} 计算公式如下:

$$R_{WIoU} = \exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)^*}\right) \tag{6}$$

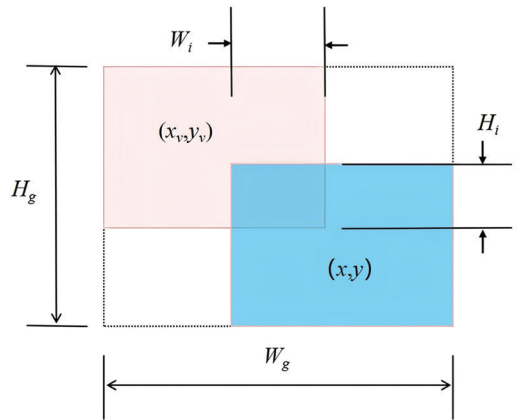


图 6 $L_{WIoU-v1}$ 定义图

式中: W_g 、 H_g 分别表示最小封闭框的宽、高; * 表示 W_g 、 H_g 仅用于归一化的数值运算, 不再对模型参数产生梯度, 从而没有引入新的度量。 $R_{wIoU} \in [1, e)$, 可显著放大普通质量锚框的 L_{IoU} 。 $L_{IoU} \in [0, 1)$ 可显著降低高质量锚框的 R_{wIoU} , 并且当锚框与目标框重合良好时, 它更注重中心点之间的距离。 $L_{wIoU-v1} \in [0, 1)$ 可明显地降低高质量锚框的 R_{wIoU} , 并且当锚框与目标框较吻合时, 它将更加集中于中心点之间的距离。

锚框的离群度定义如式(7)所示:

$$\beta = \frac{\overline{L_{IoU}}}{L_{IoU}} \in [0, +\infty] \quad (7)$$

式中: $\overline{L_{IoU}}$ 表示 IoU 损失的指数移动平均值。

之后使用离群度构造非单调聚焦系数并将其应用到 $L_{wIoU-v1}$, 得到了 $L_{wIoU-v3}$ ^[6], 其计算公式为:

$$L_{wIoU-v3} = rL_{wIoU-v1}, r = \frac{\beta}{\delta\alpha^{\beta-\delta}} \quad (8)$$

式中: r 为非单调梯度增益; α 、 δ 为超参数, 用来控制损失函数的梯度增益。

利用 *Focal* 动态加权机制^[7]解决目标检测分类中正负样本、难易样本不均衡的问题。 *Focal* 动态加权机制计算公式如式(9)所示:

$$Focal(p_i) = -(1 - p_i)^\gamma \cdot \ln(p_i) \quad (9)$$

式中: $p_i \in [0, 1)$ 表示模型对真实类别的预测概率, 越接近 1 表示预测越准确; $\gamma \in [0, 5]$ 表示聚焦参数, 用于控制对简单样本的权重衰减速率, $\gamma=0$ 时退化为标准交叉熵损失; $(1 - p_i)^\gamma$ 为动态调制因子, 用于自动降低高置信度简单样本 ($p_i \rightarrow 1$) 的损失权重, 使模型聚焦于难分类的样本 ($p_i \rightarrow 0$) 的优化。

本研究基于 $L_{wIoU-v3}$ 以及 *Focal* 动态加权机制公式得出 *FocalWise-IoU* 函数, 核心逻辑是先通过 *Focal* 调制因子聚焦低 IoU 样本, 并通过几何注意力来强化锚框的几何位置约束, 最后通过 r 优化低质量锚框的权重分配。 *FocalWise-IoU* 函数的推导公式为:

$$L_{FocalWise-IoU} = (1 - IoU)^{\gamma+1} \cdot r \cdot \exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)^*}\right) \quad (10)$$

1.4 评价指标

为了汇总分析各优化部分对模型的影响, 采用 5 种模型实验指标进行评判: 精确率 (*Precision*)、浮点运算量 (*FLOPs*)、召回率 (*R*)、模型参数量 (*Params*)、检测精度均值 (*mAP*) 作为评价指标^[20]。计算公式如下:

$$Precision = \frac{TP}{TP + FP} \quad (11)$$

$$R = \frac{TP}{TP + FN} \quad (12)$$

$$AP_c = \int_0^1 Precision(R) dR \quad (13)$$

$$mAP = \frac{1}{C} \sum_{c=1}^C AP_c \quad (14)$$

式中: TP 为预测正确的样本数; FP 为被错误预测为正例的负例样本数; FN 为被错误预测为负例的正例样本数; C 为任务中的类别总数; AP_c 为第 c 个类别的检测精度。

1.5 实验环境及参数设置

实验的硬件环境为 12th Gen Intel(R) Core(TM) i3-12100F 3.30 GHz 处理器, 显卡为 NVIDIA GeForce RTX 4060 ti, 显存为 16 GB; 操作系统为 Windows10; 机器学习软件平台为 PyTorch 2.3.1 和 CUDA 12.4。为了保证结果的有效性和真实性, 所有实验均在 PyTorch 框架下进行端到端训练。本实验参数如表 4 所示, 各个模型训练过程均未使用任何预训练权重。

表 4 软件环境参数表

参数	设置	参数	设置
迭代轮次/轮	200	初始学习率	0.01
每批训练图片数/张	16	动量因子	0.937
图像分辨率	640×640	权重衰减	0.000 5

2 结果与分析

2.1 对比实验

为了验证优化后模型的优势, 将其与目前阶段常用的 6 种目标检测模型 (SSD、Faster R-CNN、YOLOv7-tiny、YOLOv5s、YOLOv10-N、YOLOv8n) 进行对比实验。对比结果见表 5。

表 5 对比实验结果

模型	$mAP50/\%$	$mAP50:95/\%$	$R/\%$	$Precision/\%$	$FLOPs/G$	$Params/M$
SSD	79.2	37.7	85.1	53.7	62.7	26.3
Faster R-CNN	79.4	39.1	74.8	78.3	370.2	37.1
YOLOv5s	79.7	64.0	74.3	80.3	23.8	9.1
YOLOv7-tiny	77.5	56.6	76.2	80.1	13.2	6.0
YOLOv8n	79.8	63.7	72.2	79.4	8.2	3.2
YOLOv10-N	74.8	58.9	66.7	74.8	8.4	2.7
Dil-YOLO	80.3	63.9	75	79.8	6.1	2.1

注: $mAP50$ 表示在单一 IoU 阈值为 0.5 时的平均精度均值; $mAP50:95$ 在 10 个等间隔 IoU 阈值下计算 mAP , 再对这 10 个值取平均。

与 SSD、Faster R-CNN、YOLOv5s、YOLOv7-tiny、YOLOv10-N 等主流检测模型相比, 本研究提出的模型 Dil-YOLO 在 $mAP50$ 、 $mAP50:95$ 、 R 、 $Precision$ 上分别达到 80.3%、63.9%、75.0%、79.8%, 综合精度表现优异。相比基线模型 YOLOv8n, Dil-YOLO 模型的 $mAP50$ 提升 0.5 个百分点, $FLOPs$ 减少 2.1 G、参数量减少 1.1 M, 以更低开销取得更好检测效果, 兼顾了检测精度与模型轻量化。

2.2 消融实验

为了系统分析每一个改进点对模型的影响情况, 以 YOLOv8n 模型作为初始模型, 将 FasterNet 模块替换为 Dil-FasterNet 模块、特征融合模块替换为 Slim-Neck 模块、损失函数替换为 FocalWise-IoU, 改进模型的 8 次实验结果见表 6。

表 6 消融实验

序号	Dil-FasterNet	Slim-Neck	FocalWise-IoU	$mAP50/\%$	$mAP50:95/\%$	$R/\%$	$Precision/\%$	$FLOPs/G$	$Params/M$
1	—	—	—	79.8	63.7	72.2	79.4	8.2	3.2
2	√	—	—	78.3	62.4	70.8	77.6	7.2	2.6
3	—	√	—	80.1	64.0	74.6	77.2	7.4	2.8
4	—	—	√	78.9	62.9	69.8	81.3	8.2	3.0
5	√	√	—	79.5	62.6	72.5	79.2	6.9	3.2
6	—	√	√	80.0	62.3	74.9	78.8	7.0	2.6
7	√	—	√	79.2	63.1	74.7	77.2	7.3	2.7
8	√	√	√	80.3	63.9	75.0	79.8	6.1	2.1

由表 6 可知 Dil-FasterNet 的引入使模型的 $Params$ 和 $FLOPs$ 分别降低了 0.6 M、1 G, 有效减少了骨干网络的参数大小和浮点运算量。Slim-Neck 特征融合结构的引入让模型的 $mAP50$ 增加了 0.3%, $Params$ 和 $FLOPs$ 分别降低了 0.4 M、0.8 G, 在增强小目标检测精度的同时减小了模型规模。消融实验证明了

YOLOv8n 改进的有效性,在 *Params* 和 *FLOPs* 上分别下降了 1.1 M 和 2.1 G,同时提升了模型的精度。

2.3 泛化实验

为了验证改进后的模型在不同场景下的表现,本研究选取国际农业植物病害应用最广泛的公开基准数据集 PlantVillage^[21] 作为跨域测试数据源,从中筛选出与训练数据集完全独立、作物品类与病害特征存在显著的域偏移的棉花叶片病害数据集(PlantVillage Cotton Leaf Disease Dataset)与水稻叶片病害数据集(PlantVillage Rice Leaf Disease Dataset)两个标准子集作为跨域测试集。对比实验结果见表 7。

表 7 泛化实验

数据集	模型	<i>mAP</i> 50/	<i>mAP</i> 50 : 95/	<i>R</i> /	<i>Precision</i> /	<i>FLOPs</i> /	<i>Params</i> /
		%	%	%	%	G	M
棉花病害	YOLOv8n	82.3	48.9	73.8	86.0	8.2	3.0
	DiI-YOLO	83.1	51.3	74.6	87.1	6.1	2.1
水稻病害	YOLOv8n	87.3	54.7	81.5	82.1	8.2	3.0
	DiI-YOLO	87.7	57.1	87.1	87.8	6.1	2.1

从表 7 格数据可知,相比于原模型,改进后的检测模型在棉花病害数据集和水稻病害数据集保持了最佳的计算效率,验证了改进后模型在自然农田环境检测任务中具有良好的泛化性。

2.4 可视化

2.4.1 可视化热力图

为了直观呈现改进后模型在农田环境中的病害检测性能,本研究采用了 Grad-CAM^[22] 可视化技术,旨在展示神经网络检测结果的特征映射。改进模型能够检测到在原始模型中被忽略的微小目标,说明模型性能得到提升。实验热力图见图 7。

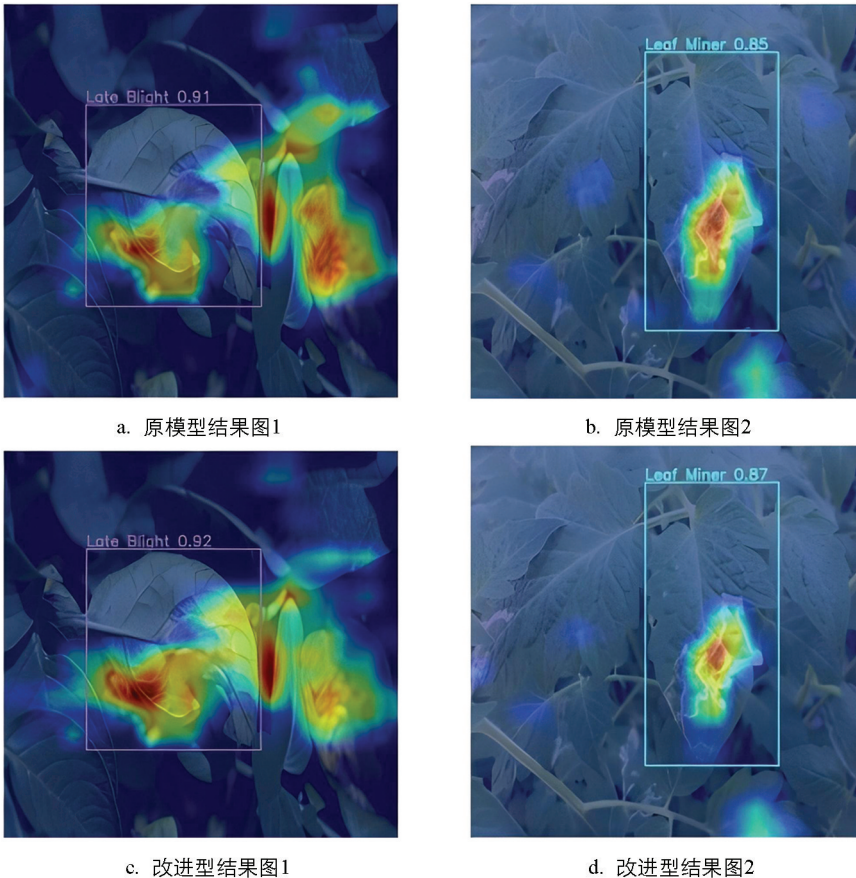


图 7 模型结果对比热力图

2.4.2 热力图泛化结果

为了进一步通过可视化热力图展示改进效果, 对 2.3 节泛化实验中所使用的两个泛化数据集采用 Grad-CAM 可视化技术, 从而更加直观地展示改进后模型的检测效果。水稻病害、棉花病害数据集结果热力图分别见图 8、图 9。

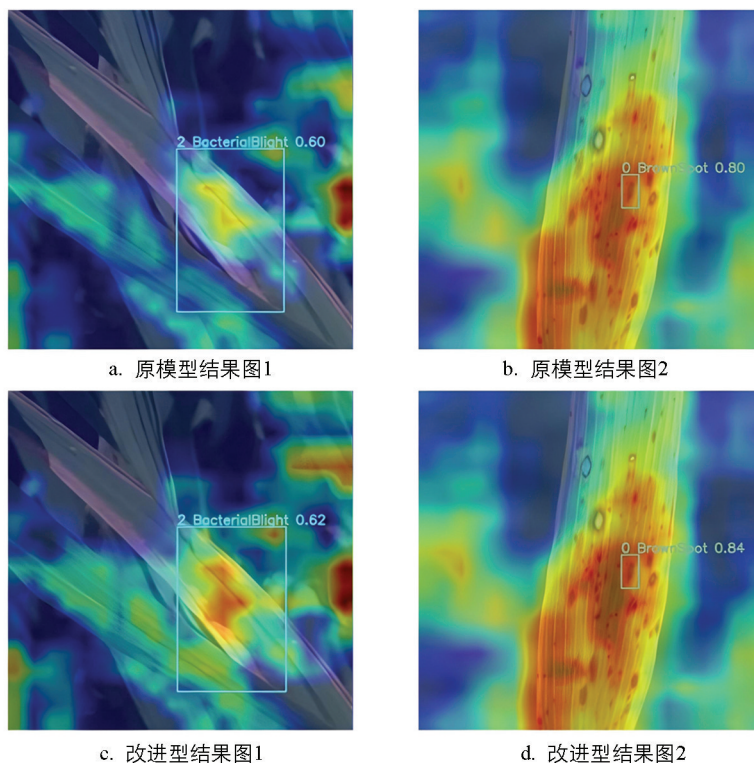


图 8 水稻泛化结果对比热力图

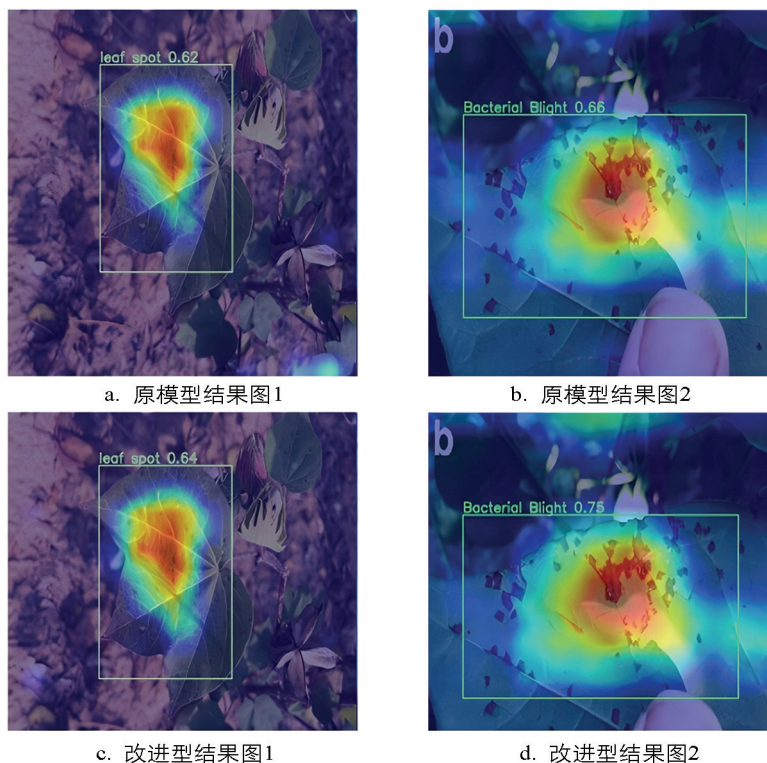


图 9 棉花泛化结果对比热力图

3 结论

随着智慧农业的高速发展,农田环境下对于高精度、低延时、低算力的目标识别模型提出了更高的需求。当前主流识别模型在目标密集、尺度差异大的农田环境中常出现识别偏差,难以兼顾模型大小等问题。基于此,本研究提出了一种基于 YOLOv8 的番茄叶片病害识别模型。在 YOLOv8 的骨干网络中设计了 Dil-FasterNet 模块,通过空洞卷积可在不增加参数量的同时捕捉多尺度特征,通过深度可分离卷积能将标准卷积拆解为深度卷积与逐点卷积以大幅削减参数量和计算量。之后在颈部网络引入 Slim-Neck 模块,通过 GSConv 来替代标准卷积,搭配 VoV-GSCSP 瓶颈结构优化特征融合,在大幅降低参数量与计算开销的同时,强化多尺度特征的高效交互与表达。最后设计了 FocalWise-IoU 损失函数,使模型聚焦低 IoU 样本,解决锚框质量导致的梯度不平衡的问题,通过对特征维度的精细化约束,提升模型对复杂目标的判别能力。与基线模型相比,Dil-YOLO 模型大幅减少了模型的计算量和参数量,Params 降低了 1.1 M,浮点运算量降低了 2.1 G,且模型 *mAP* 相比于基线模型提高了 0.5 个百分点。

在后续研究中,将进一步增加数据集中病害种类和图像数量,强化数据的多样性,并与更多先进模型进行对比分析。同时补充不同地域、不同种植模式(设施大棚、露天种植)、不同作物生育期、极端光照/阴雨/叶片严重遮挡等田间复杂环境下的病害样本,构建更完备的番茄病害专用数据集。

参考文献:

- [1] 汝刚,刘慧,沈桂龙. 用人工智能改造中国农业:理论阐释与制度创新 [J]. 经济学家,2020(4): 110-118.
- [2] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single Shot MultiBox Detector [C] //Computer Vision-ECCV 2016. Cham: Springer, 2016: 21-37.
- [3] 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 (CVPR). New York: IEEE Press, 2016: 779-788.
- [4] 马一鸣,尹爽,郭瑞,等. 烟叶烘烤阶段不同 YOLO 算法模型的实时判别性能比较 [J/OL]. 河南农业大学学报, (2025-11-04) [2025-12-21]. <https://doi.org/10.16445/j.cnki.1000-2340.20251104.001>.
- [5] LI H L, LI J, WEI H B, et al. Slim-Neck by GSConv: A Lightweight-Design for Real-Time Detector Architectures [J]. Journal of Real-Time Image Processing, 2024, 21(3): 62.
- [6] TONG Z J, CHEN Y H, XU Z W, et al. Wise-IoU: Bounding Box Regression Loss with Dynamic Focusing Mechanism [EB/OL]. (2023-01-24) [2025-10-11]. <https://arxiv.org/abs/2301.10051>.
- [7] ZHANG Y F, REN W Q, ZHANG Z, et al. Focal and Efficient IoU Loss for Accurate Bounding Box Regression [J]. Neurocomputing, 2022, 506: 146-157.
- [8] 刘洋,官志宏,黎贞发,等. 基于改进 YOLOv5 的番茄成熟度检测方法 [J]. 中国农业气象, 2024, 45(12): 1521-1532.
- [9] HE K M, ZHANG X Y, REN S Q, et al. Deep Residual Learning for Image Recognition [C] //2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE Press, 2016: 770-778.
- [10] TAN M X, LE Q V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks [EB/OL]. (2019-05-28) [2025-10-11]. <https://arxiv.org/abs/1905.11946>.
- [11] LIU S, QI L, QIN H F, et al. Path Aggregation Network for Instance Segmentation [C] //2018 IEEE/CVF Conference

- on Computer Vision and Pattern Recognition. New York: IEEE Press, 2018: 8759-8768.
- [12] 刘伯红, 郝文瑞. 面向交通目标的多尺度轻量化检测模型 [J]. 重庆邮电大学学报(自然科学版), 2025, 37(2): 185-195.
- [13] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [14] LI J F, WEN Y, HE L H. SCConv: Spatial and Channel Reconstruction Convolution for Feature Redundancy [C] // 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE Press, 2023: 6153-6162.
- [15] YU F, KOLTUN V. Multi-Scale Context Aggregation by Dilated Convolutions [EB/OL]. (2015-12-23) [2025-10-11]. <https://arxiv.org/abs/1511.07122>.
- [16] 袁泉, 杨清泉, 袁亚隆, 等. 改进 YOLOv8 的水下目标检测算法 [J]. 重庆邮电大学学报(自然科学版), 2025, 37(5): 729-740.
- [17] REZATOFIGHI H, TSOI N, GWAK J, et al. Generalized Intersection over Union: A Metric and a Loss for Bounding Box Regression [C] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE Press, 658-666.
- [18] ZHENG Z H, WANG P, LIU W, et al. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993-13000.
- [19] 杨鹏. 离群检测及其优化算法研究 [D]. 重庆: 重庆大学, 2010.
- [20] 胡佳乐, 周敏, 申飞. 面向无人机小目标的 RTDETR 改进检测算法 [J]. 计算机工程与应用, 2024, 60(20): 198-206.
- [21] SAKIB S N, HAQUE N, HOSSAIN MZ, et al. PlantVillageVQA: A Visual Question Answering Dataset for Benchmarking Vision-Language Models in Plant Science [EB/OL]. (2025-08-23) [2025-10-11]. <https://arxiv.org/abs/2508.17117>.
- [22] SELVARAJU RR, COGSWELL M, DAS A, et al. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization [J]. International Journal of Computer Vision, 2020, 128(2): 336-359.

责任编辑 张枸