doi:10.6041/j.issn.1000-1298.2025.11.052

基于改进 YOLO v8n 的轻量化番茄叶片病害检测方法

张　净[1]　华　玟[1]　刘晓梅[2]　孙月平[1]

(1.江苏大学电气信息工程学院,镇江 212013;2.江苏科茂信息技术有限公司,镇江 212000)

摘要:针对温室环境下番茄叶片重叠有遮挡,部分番茄病害病斑较小导致检测精度不高、现有检测模型参数量多和难以部署至计算资源有限的嵌入式设备端等问题,提出一种基于改进 YOLO v8n 的轻量化番茄叶片病害检测方法。首先,使用 HGNetV2 网络替换 YOLO v8n 模型的主干网络,并引入轻量化卷积 GhostConv 结构,提高对多尺度病害特征的提取能力,同时减少模型的参数和计算量;其次,在颈部网络部分使用改进的 C2f_Star 模块,实现进一步轻量化的同时增强模型的特征表达能力;然后,引入轻量级 SimAM 注意力机制模块,加强模型在识别过程中对小病斑特征的有效提取与表征能力;最后,设计混合损失函数 Wise_Focaler_MPDIoU 代替原损失函数,提高网络边界框回归性能,加快收敛速度。试验结果表明,改进模型(GSSW-YOLO v8n)的参数量及模型权重文件大小相比原模型分别降低 28.9% 和 26.9%,同时在本研究所使用的番茄叶片病害数据集上的平均检测精度达 97.4%,相较原模型提高 3.0 个百分点,对比其他目标检测算法 Faster R-CNN、YOLO v5n、YOLO v6n 和 YOLO v7-tiny,改进模型的平均检测精度分别高出 13.3、5.4、9.9、9.2 个百分点。改进模型(GSSW-YOLO v8n)具有更高的检测精度和更少的参数量,可为番茄叶片病害检测模型在嵌入式设备端上的部署和应用提供参考。

关键词:番茄病害;轻量化;YOLO v8;目标检测;SimAM

中图分类号:TP391.4　　**文献标识码**:A　　**文章编号**:1000-1298(2025)11-0538-12　　**OSID**:

Lightweight Tomato Leaf Disease Detection Method Based on Improved YOLO v8n

ZHANG Jing[1]　HUA Wen[1]　LIU Xiaomei[2]　SUN Yueping[1]

(1. *School of Electrical and Information Engineering, Jiangsu University, Zhenjiang* 212013, *China*
2. *Jiangsu Kemao Information Technology Co., Ltd., Zhenjiang* 212000, *China*)

Abstract: Aiming at the problems of overlapping tomato leaves with shading in greenhouse environment, small spots of some diseases leading to low detection accuracy, and large number of parameters of the existing detection model, which is difficult to be deployed to the embedded end with limited computational resources, a lightweight tomato leaf disease detection method based on the improvement of YOLO v8n was proposed. Firstly, the HGNetV2 network was used to replace the backbone network of the YOLO v8n model, and a lightweight convolutional GhostConv structure was introduced to improve the extraction of multi-scale disease features while reducing the number of parameters and computation of the model. Secondly, an improved C2f_Star module was used in the neck network part to achieve further lightweighting while enhancing the feature representation of the model. Then the lightweight attention mechanism SimAM module was added to strengthen the feature extraction ability of the model for small lesions. Finally, the hybrid loss function Wise_Focaler_MPDIoU was designed to replace the original loss function to improve the performance of the network bounding box regression and accelerate the convergence speed. The experimental results showed that the number of parameters and model size of the improved model (GSSW-YOLO v8n) were reduced by 28.9% and 26.9%, respectively, compared with the original model, while the average detection accuracy on the tomato leaf disease dataset used reached 97.4%, which was an improvement of 3.0 percentage points, comparing with other target detection algorithms of Faster R-CNN, YOLO v5n, YOLO v6n, and YOLO v7-tiny by 13.3, 5.4, 9.9, and 9.2 percentage points, respectively. The improved model (GSSW-YOLO v8n) had high

收稿日期:2024-07-24　修回日期:2024-09-06

基金项目:国家自然科学基金项目(62173162)

作者简介:张净(1975—),女,副教授,博士,主要从事计算机应用研究,E-mail:jszj08062000@163.com

detection accuracy and less number of parameters, which can provide a reference for the deployment and application of the tomato leaf disease detection model on the embedded device side.
Key words: tomato diseases; lightweighting; YOLO v8; object detection; SimAM

0　引言

为了满足市场对优质番茄的需求,南北方很多地区均采取大棚种植模式,这样可以使外界环境因素对番茄植株生长的影响得到最大程度的降低,但大棚内潮湿闷热的环境也会容易滋生病菌,导致番茄发生如早疫病、细菌性斑点病、黄化曲叶病等病害,会对番茄植株的正常生长造成严重的影响,从而影响番茄的生产质量[1]。因此,番茄病害的及时诊断与防治至关重要。

传统的番茄病害诊断方法主要依赖于人工经验判断,但由于叶片数量众多且一些病害类别难以区分,人工识别手段不仅耗费大量的人力资源与物质成本,还因其高度的主观性而难以保证识别结果的稳定性与准确性[2]。近些年来,随着机器学习、深度学习的迅速发展,许多目标检测算法[3]开始应用到番茄的病害检测当中,并取得了一定的进展。例如,金婷婷等[4]提出一种番茄叶部早期病斑诊断算法,采用 Mosaic 9 增加数据多样性,引入 GSConv 和 Slim - Neck 网络、SimAM 注意力机制和自适应空间特征融合模块优化 YOLO v5s 模型结构,改进模型识别精度达 95.1%;QI 等[5]提出了一种改进的SE - YOLO v5 网络模型,加入挤压激励(SE)模块实现关键病害特征提取,改进模型对番茄病毒病平均识别精度为 94.10%,参数量近 4.2×10^7;韩鑫等[6]为实现番茄病害早期诊断,提出一种结合红外热成像和改进 YOLO v5 的识别模型,通过采用 SPD - Conv 模块替代主干及颈部网络中的特定卷积层,融入 SE 机制和 SPPF 模块深化特征提取效率,优化后模型平均检测精度达 95.7%,但其番茄病害种类较为单一;张领先等[7]通过在原 YOLO v5 模型中引入 BiFPN 改进 Neck 结构、嵌入 CBAM 机制并调整检测层的输入通道,显著提升了模型在识别番茄各类病害时的整体性能,优化后病害识别率达 96.4%,但对于特征较为微小的红蜘蛛损伤,其识别精度却有所下降。以上研究都提升了番茄病害检测精度,但针对小病斑的检测性能仍然较低,且网络结构也更加复杂,存在参数量大,实时性差的问题。

在实际农业应用中,移动设备和嵌入式设备的计算资源有限,而深度学习算法内存占用量大,存在着部署困难的问题[8],因此,模型轻量化研究也逐渐受到广泛关注。王会征等[9]为实现番茄叶片病虫害的精准与快速检测,提出一种 SLP - YOLO v7 - tiny 模型,通过将部分传统的 3×3 卷积层 Conv2D 替换为分布式偏移卷积 DSConv2D,降低了计算复杂度,引入无参数的 SimAM 注意力机制和 Focal - EIoU 损失函数,加快模型收敛。改进后的模型整体识别精度达 95.9%,计算量减少 62.6%,但参数并未减少,且其识别背景较为简单;方晨晨等[10]针对卷积神经网络内存占用量普遍较高的问题,设计了一种优化的深度残差网络。通过采用多尺度卷积和深度可分离卷积替换传统单一尺度卷积和部分标准卷积,显著提升了模型对番茄病害的识别效果,平均识别准确率高达 98.5%。然而,尽管模型权重文件大小缩减了 18.8% 至 19.0 MB,其资源占用仍相对较大;王娜等[11]为解决当前番茄叶片病害识别面临的精度不足和模型参数大等挑战,提出了一种基于 YOLO v5n 的轻量化改进方法。该方法采用 C3Ghost 轻量化模块替换原有的 C3 模块以减轻卷积运算负担,同时在颈部引入 GSConv 与 VOV - GSCSP 2 种轻量级卷积模块,并引入 PAGCP 算法实施全局通道剪枝,有效压缩了参数量并降低了训练成本,参数量减少 66.6%,但平均精度均值下降 0.5 个百分点,且数据集相对单一,未考虑到实际番茄病害识别环境的复杂性;杨佳昊等[12]通过替换主干网络、剪枝和知识蒸馏建立了一种轻量化的番茄黄化曲叶病检测方法,经轻量化处理的模型内存占用量仅为 YOLO v5s 模型的 10.0%,但精确率下降 3.7 个百分点。上述研究虽然在一定程度上实现了模型轻量化,但模型检测准确率也受到了制约,且识别场景较为单一。

针对现有大多数番茄病害研究方法仍存在识别场景较为单一,小病斑漏检率高、难以平衡计算复杂度与精度,不便于在算力资源受限的设备上部署的问题,本研究提出一种基于改进 YOLO v8n 的轻量化番茄叶片病害检测方法,从而实现计算量更小、检测精度更高的病害检测,以期为番茄叶片病害检测及硬件部署提供参考。

1　材料与方法

1.1　数据集制作

1.1.1　数据来源

采用的番茄叶片病害数据集由两部分构成:①获取自公开网络(https://roboflow.com),包括 7 类病害叶片和健康叶片,共 767 幅图像,病害类别

包括早疫病、晚疫病、细菌性斑点病、叶霉病、斑枯病、花叶病毒和黄化曲叶病。②采集于镇江市江苏科茂信息技术有限公司乡下大棚,数据采集时间为 2024 年 2—5 月,采集设备为 iphone 手机,采集的图像分辨率为 3 024 像素 ×4 032 像素,保存格式为“JPG”。拍摄时采用顺光、逆光、近距离、远距离和不同角度等多种拍摄方式,以丰富数据的多样性。经过对采集到的原始图像进行筛选整理后,共得到 176 幅图像,包括早疫病、黄化曲叶病、健康叶片 3 类。

本研究所收集的两部分数据集共计包含 923 幅番茄叶片图像样本,其中部分数据集图像如图 1 所示。



(a) 早疫病 (b) 晚疫病 (c) 细菌性斑点病 (d) 斑枯病

(e) 叶霉病 (f) 花叶病毒 (g) 黄化曲叶病 (h) 健康

图 1 部分番茄数据集图像

Fig. 1 Partial tomato dataset images

1.1.2 数据预处理

对获取到的数据集样本图像采用 LabelImg 软件进行标注,采用最大水平矩形框框住图像中的病害目标所在区域,并标注上对应的病害类别标签,然后保存为 XML 文件,最后转化为 TXT 文件,作为 YOLO v8n 模型输入。

为避免因数据集数量过少、样本不平衡导致模型出现过拟合及泛化能力差等问题[13],本研究使用添加噪声、变换亮度、旋转角度、镜像翻转和平移 5 种数据增强技术扩充数据量。经数据增强后获得图像共 4 615 幅,按照比例 8∶1∶1 随机划分为训练集(3 692 幅)、测试集(462 幅)和验证集(461 幅)。

1.2 YOLO v8 网络

YOLO v8 不仅继承了 YOLO v5 的强大优势,还融入了诸多创新功能与优化措施,实现了模型性能显著提升[14]。该网络主要包括:输入端(Input):负责图像输入;主干网络(Backbone):进行特征提取;颈部网络(Neck):用于特征融合与增强;以及头部网络(Head):用于最终的检测与分类任务。相较于 YOLO v5 而言,YOLO v8 主干网络中的第 1 次卷积的卷积核尺寸缩小为 3 ×3,且将所有 C3 模块进行了升级替换,采用 C2f 模块显著增加了跳层连接数量,并融入额外的 Split 操作,以增强信息流动与特征多样性。头部网络使用目前主流的解耦头结构,将分类和检测头分离成两条分支,同时由原先的 Anchor - Based 的策略转变为 Anchor - Free。此外,在损失函数部分,YOLO v8 使用 Task - Aligned Assigner 正负样本匹配方式,并引入了 Distribution Focal Loss。

为满足多样化应用场景的需求,YOLO v8 同样灵活设计了 n/s/m/l/x 5 种不同缩放级别的模型版本,这些模型在深度与宽度上呈现出递增趋势,检测精确度得以逐步提升,然而,模型参数与计算复杂度也随之增加。考虑到硬件设备的计算资源有限性,选用参数最少且检测速度最优的 YOLO v8n 模型作为基准模型。

2 轻量化番茄叶片病害检测模型

本研究提出一种基于改进 YOLO v8n 的轻量化番茄叶片病害检测方法,主要在以下 4 个方面进行优化和改进,以实现更高的检测性能和更低的计算成本。

(1)引入轻量级网络 HGNetV2 增强模型对多尺度病害特征的提取能力,并减少模型参数量和计算量,同时利用 GhostConv 模块对 HGNetV2 进行优化,

使用优化后的 Ghost_HGNetV2 网络作为 YOLO v8n 的主干网络，实现模型的轻量化。

(2)为进一步减少模型参数，同时提升模型的特征表达能力，引入 StarNet 网络改进 C2f 模块，并使用改进后的 C2f_Star 模块替换颈部网络中原有的 C2f 模块。

(3)为了增强模型对细微病害特征的识别能力，在颈部网络结构中嵌入 SimAM 注意力模块，使模型更加聚焦于小目标病害的细节特征，进而提升对微小病斑的检测精度与准确性。

(4)为提升模型边界框回归性能，设计混合损失函数 Wise_Focaler_MPDIoU。通过综合考虑分类信息和边界框回归信息，Wise_Focaler_MPDIoU 损失函数可以使模型更加关注病害特征的位置和大小，从而提高检测精度。本研究将改进后的模型命名为 GSSW－YOLO v8n 网络，其模型结构如图 2 所示。

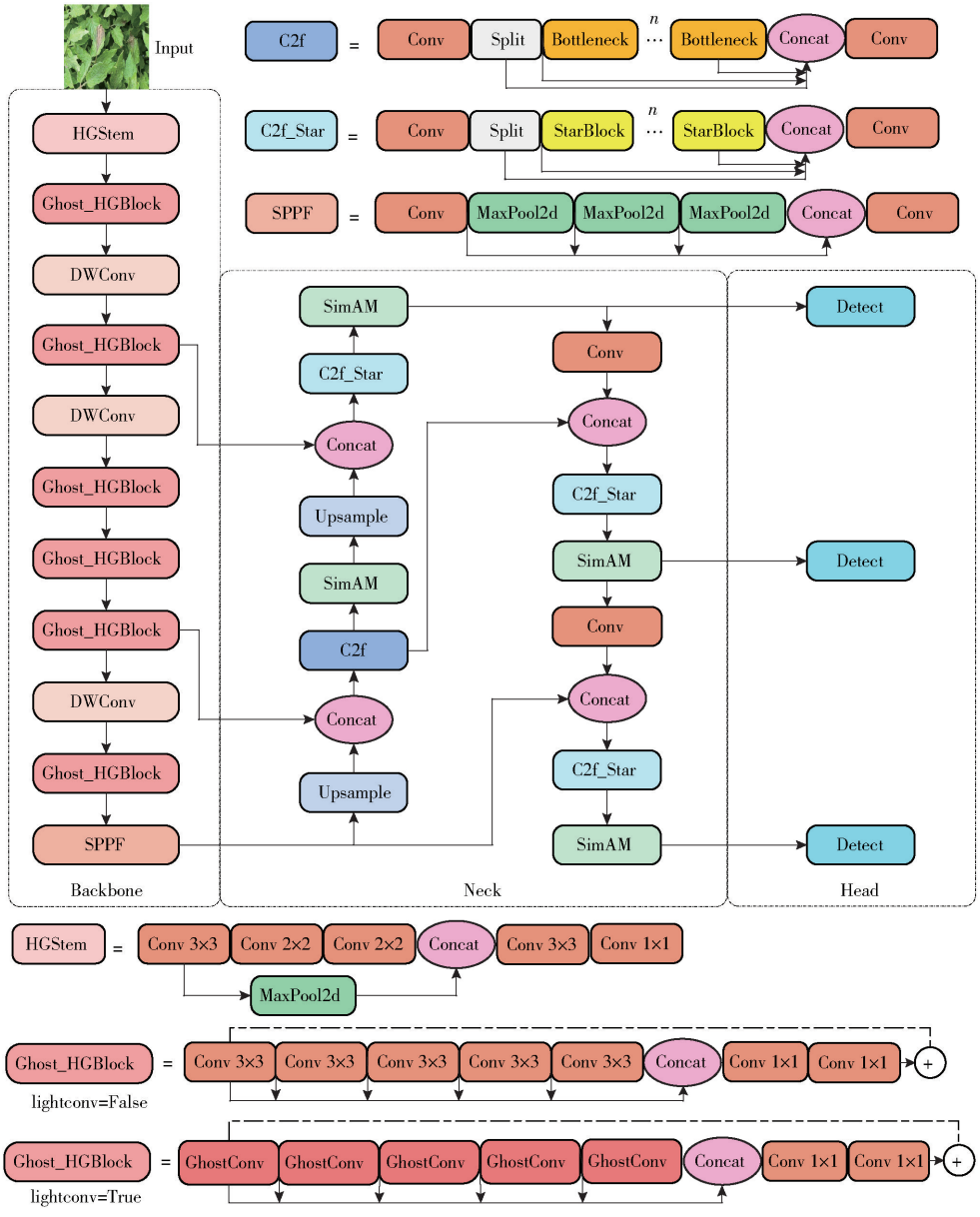


图 2　GSSW－YOLO v8n 模型结构

Fig. 2　GSSW－YOLO v8n model structure

2.1　轻量级主干网络 Ghost_HGNetV2

PP－HGNetV2(High Performance GPU Network V2)是百度飞桨视觉团队自研的轻量级高性能骨干网络，在其发布的 RT－DETR 算法的性能指标中，表现最好的 2 个模型的 Backbone 都使用 HGNetV2 网络[15]。引入 HGNetV2 作为 YOLO v8n 的主干网络，通过轻量化的卷积操作和多尺度的特征提取策略提高网络性能和泛化能力，同时减少模型参数量和计算量。

HGNetV2 整体结构由 1 个 HGStem 和 4 个 HGStage 模块构成。HGStem 用于在网络初期阶段进行快速的特征提取，共有 5 个卷积层和 1 个池化

层,其使用的卷积核包括不同的尺寸,能够使网络捕获到更多尺度的病害特征,学习到更为丰富的信息。网络的第 2 到第 5 层由 HGStage 构成,每个 HGStage 主要由深度可分离卷积和 HGBlock 构成。HGBlock 构建了多个标准卷积层,并通过挤压卷积对多个卷积层的输出进行合并,融合多尺度特征。然后使用激励卷积进一步提取特征,最后通常包含残差连接,减轻深层网络的梯度消失现象,优化模型训练的效率。

GhostConv 模块采用一种高效的线性变换策略,以较小的计算成本额外生成丰富的特征映射,可显著减少参数总量及整体计算复杂度,在保持性能的同时实现轻量化设计,更适用于内存和计算资源有限的嵌入式设备。因此,本文引入 GhostConv 对 HGNetV2 网络中的 HGBlock 结构进行优化,替换原有的 LightConv 得到 Ghost_HGBlock 模块,降低模型的计算和存储开销。

GhostConv 模块包括 3 个关键步骤:常规卷积预处理、Ghost 特征生成以及特征图拼接[16],其卷积结构如图 3 所示。首先,通过常规的 1×1 卷积操作对输入特征图进行通道维度的压缩,提取出本征特征图。随后,对每个本征特征图应用一系列低成本的线性变换,包括深度可分离卷积(Depthwise Convolution)、批量归一化(Batch Normalization)以及非线性激活函数(ReLU),它们协同工作以产生额外的 Ghost 特征图。最后,将 Ghost 特征图与本征特征图进行拼接,有效融合原始特征与通过高效变换生成的新特征,使得最终的输出特征图既包含丰富的信息又保持较低的计算复杂度。

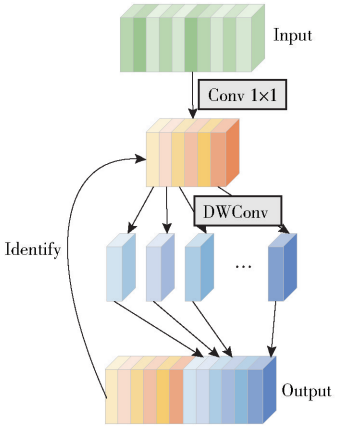


图 3 GhostConv 结构

Fig. 3 GhostConv structure

2.2 特征融合模块 C2f_Star

StarNet[17]是微软 2024 年提出的一种基于元素乘法的高效网络,通过元素乘法融合不同子空间特征的操作称为星操作(Star operation)。Star operation 具有强表达能力,可以将输入映射到极高维的非线性特征空间,与增加网络宽度的传统神经网络不同,Star operation 类似于在不同通道上进行成对特征乘法的核函数,特别是多项式核函数。在紧凑的特征空间内计算,同时又受益于隐含的高维度。

引入 StarNet 网络与 C2f 模块进行融合,使用 StarBlock 块改进 C2f 中的 Bottleneck 块,构成新的 C2f_Star 模块,加强模型的特征表达能力。StarBlock 块结构见图 4,输入首先经过深度卷积处理和归一化操作,然后分成 2 条分支,再进行星操作,实现高维和非线性特征空间的映射,而无需增加计算复杂度。当 StarBlock 通过多层堆叠时,每一层都会带来隐式维度复杂性的指数增长。只需几层,Star operation 就可在紧凑的特征空间内实现几乎无限的维度。相比于原 C2f 模块,改进的 C2f_Star 模块在保持计算效率的同时,能够获得更丰富和表达力更强的特征表示。

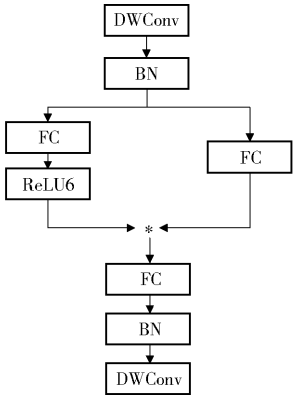


图 4 StarBlock 模块结构

Fig. 4 StarBlock module structure

2.3 SimAM 注意力机制

当前注意力机制模块可大致归类为两大类别:①通道注意力,沿着通道维度操作,平等地对所有位置,生成一维注意力权重。②空间注意力,沿着空间维度操作,对所有通道同等对待,生成二维注意力权重。而三维权重既考虑通道维度,又考虑空间维度,拥有更强的辨别线索的能力,因此优于传统的一维和二维权重注意力[18]。

SimAM 作为一种轻量级且无额外参数添加的卷积神经网络注意力机制[19],与既有的通道级与空间级注意力模块相区别,其不同之处在于能够在不扩充基础网络参数量的前提下,为网络各层特征图动态生成三维注意力权重。番茄叶片的部分病害如细菌性斑点病,初期常以微小斑点形态出现,这对模型捕捉细微病变区域的能力有更高的要求。在网络的颈部网络中融入 SimAM 注意力机制,为目标特征

的空间维度与通道维度自动分配差异化的关注权重，使网络更加聚焦于番茄叶片病害图像中的关键病变区域，同时有效削弱背景信息的干扰，提升模型对微小病斑的检测精确度。SimAM 注意力机制中 3D 权重示意图如图 5 所示。

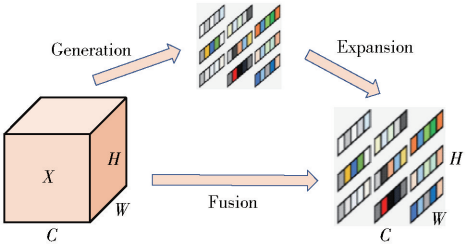


图 5　SimAM 中 3D 注意力权重示意图

Fig. 5　Schematic of 3D attention weights in SimAM

为了根据神经网络每一层中的特征图推导出三维注意力权重，需要计算单个神经元的重要性。在神经科学领域内，信息承载量高的神经元往往展现出与邻近神经元截然不同的放电模式。此外，当某个神经元被激活时，在空间上倾向于对周边神经元施加抑制性影响，这一特性即空域抑制效应。因此，定义能量函数公式寻找重要神经元，即

$$e_t(w_t,b_t,y,x_i)=(y_t-\hat{t})^2+\frac{1}{M-1}\sum_{i=1}^{M-1}(y_0-\hat{x}_i)^2 \tag{1}$$

其中

$$\begin{cases}\hat{t}=w_t t+b_t\\ \hat{x}_i=w_t x_i+b_t\end{cases} \tag{2}$$

式中　w_t——某个神经元变换时的“权重”

b_t——神经元变换时的“偏差”

x_i——输入特征的其他神经元

t——输入特征 X 的目标神经元

i——空间维度上的索引

M——在某个通道上的所有的神经元的个数

训练同一通道内神经元 t 与其他神经元之间的线性可分性，引入二进制标签代替 y_t 和 y_0，其中：$\begin{pmatrix}y_t=1\\ y_0=-1\end{pmatrix}$，并添加正则项，得到最终的能量函数公式为

$$e_t(w_t,b_t,y,x_i)=\frac{1}{M-1}\sum_{i=1}^{M-1}[-1-(w_t x_i+b_t)]^2+[1-(w_t x_i+b_t)]^2+\lambda w_t^2 \tag{3}$$

理论上，每个通道有 $M=HW$ 个能量函数，公式(3)的解析解为

$$\begin{cases}w_t=-\dfrac{2(t-\mu_t)}{(t-\mu_t)^2+2\sigma_t^2+2\lambda}\\ b_t=-\dfrac{1}{2}(t+\mu_t)w_t\end{cases} \tag{4}$$

其中

$$\begin{cases}\mu_t=\dfrac{1}{M-1}\sum_{i=1}^{M-1}x_i\\ \sigma_t^2=\dfrac{1}{M-1}\sum_{i=1}^{M-1}(x_i-\mu_t)^2\end{cases} \tag{5}$$

最小能量计算式为

$$e_t^*=\frac{4(\partial^2+\lambda)}{(t-\hat{\mu})^2+2\partial^2+2\lambda} \tag{6}$$

因此，可以根据 $1/e_t^*$ 判断神经元重要性。神经元 t 重要性越高，与邻近神经元区别就越大，最小能量就越低。

最后，根据注意力定义，使用单调函数 Sigmoid 对特征进行增强处理，计算公式为

$$\widetilde{X}=\text{Sigmoid}\left(\frac{1}{E}\right)X \tag{7}$$

式中　E——所有通道和空间维度上的能量函数的总和

2.4　Wise_Focaler_MDPIoU 损失函数

使用合适的损失函数对于提升模型检测性能至关重要。传统损失函数在评估时往往局限于预测框与真实框之间的交并比，却忽视了分类信息的潜在贡献，这在一定程度上限制了模型精度的提升。YOLO v8n 使用 CIoU Loss 计算边界框的回归损失[20]。当预测边界框和真实边界框的宽度和高度不同却具有相同的宽高比时，基于宽高比的惩罚项就无法发挥作用，CIoU 将失去有效性。

MPDIoU——一种基于最小点距离的边界框相似度比较度量[21]，简化了两个边界框之间的相似性比较，帮助算法选择最合适的边界框准确定位目标。考虑了重叠或不重叠区域、中心点距离、宽度和高度偏差等现有损失函数中的所有相关因素，同时还简化了计算过程。MPDIoU 损失函数计算式为

$$MPDIoU=IoU-\frac{d_1^2}{w^2+h^2}-\frac{d_2^2}{w^2+h^2} \tag{8}$$

其中

$$d_1^2=(x_1^{prd}-x_1^{gt})^2+(y_1^{prd}-y_1^{gt})^2 \tag{9}$$

$$d_2^2=(x_2^{prd}-x_2^{gt})^2+(y_2^{prd}-y_2^{gt})^2 \tag{10}$$

$$L_{\text{MPDIoU}}=1-MPDIoU \tag{11}$$

式中　d_1^2、d_2^2——预测框、真实框的左上角点之间和右下角点之间的欧氏距离

w、h——输入图像宽度、高度

(x_1^{prd},y_1^{prd})、(x_2^{prd},y_2^{prd})——预测框左上点和右下点的坐标

(x_1^{gt},y_1^{gt})、(x_2^{gt},y_2^{gt})——真实框左上点和右下点的坐标

CIoU 通过利用边界框之间的几何关系提高回归性能，忽略了难易样本分布对边界框回归的影响，当目标检测训练集中含有低质量示例时，会使模型

检测性能降低。Focaler - IoU 方法通过关注不同的回归样本,可以提高检测器性能[22]。对于简单样本占主导的检测任务,在边框回归过程聚焦简单样本,对于困难样本占比较高的检测任务中聚焦困难样本的边框回归,有助于提升检测效果。Focaler - IoU 损失函数公式为

$$IoU^{Focaler}=\begin{cases}0 & (IoU<d)\\ \dfrac{IoU-d}{u-d} & (d\leqslant IoU\leqslant u)\\ 1 & (IoU>u)\end{cases} \tag{12}$$

式中$[d,u]\in[0,1]$,通过调节 d 与 u 能够使得 Focaler - IoU 聚焦不同的回归样本。

CIoU 依赖于一种静态的单调聚焦机制,而 WIoU 损失函数引入了动态非单调聚焦机制,利用“离群程度”评估锚框的质量。为了克服几何特性(如距离与纵横比)可能带来的过度惩罚问题,设计了更为优越的梯度增益分配策略[23],该策略不仅降低了高质量锚框在训练过程中的竞争优势,还有效减轻了低质量样本所产生的不利梯度影响,使得 WIoU 能够更专注于中等质量的锚框,从而提升检测器整体性能。WIoU 损失函数计算公式为

$$L_{WIoU}=rR_{WIoU}L_{IoU} \tag{13}$$

其中

$$R_{WIoU}=\exp\frac{(x^{prd}-x^{gt})^2+(y^{prd}-y^{gt})^2}{c_w^2+c_h^2} \tag{14}$$

$$r=\frac{\beta}{\delta\alpha^{\beta-\delta}} \tag{15}$$

$$\beta=\frac{L_{IoU}^{*}}{L_{IoU}}\in[0,+\infty) \tag{16}$$

式中 c_w、c_h——预测框与真实框最小外接矩形的宽和高

在实际番茄叶片的病害检测过程中,可能存在叶片重叠有遮挡等情况,容易导致漏检,这类样本属于困难样本,传统的损失函数难以发挥作用。将 Focaler - IoU 损失应用于 MPDIoU 边界框回归损失函数中,能够更加聚焦不同难易样本。定义 Focaler_MPDIoU 损失函数为

$$Focaler_MPDIoU=IoU^{Focaler_IoU}-\frac{d_1^2}{w^2+h^2}-\frac{d_2^2}{w^2+h^2} \tag{17}$$

$$L_{\text{Focaler_MPDIoU}}=1+\frac{d_1^2}{w^2+h^2}+\frac{d_2^2}{w^2+h^2}-IoU^{Focaler_IoU} \tag{18}$$

然后融合 WIoU 与 Focaler_MPDIoU 得到更加聚焦的混合损失函数 Wise_ Focaler_MPDIoU,通过结合 3 种损失函数的优势,进一步提高番茄叶片病斑目标定位精度,并加速模型收敛速度,使模型具有更好的泛化能力,有效降低漏检问题。新的损失函数为

$$L_{Wise_Focaler_MPDIoU}=\frac{\beta}{\delta\alpha^{\beta-\delta}}\exp\left(\frac{(x^{prd}-x^{gt})^2+(y^{prd}-y^{gt})^2}{c_w^2+c_h^2}\right)\times\left(1-IoU^{Focaler_IoU}+\frac{d_1^2}{w^2+h^2}+\frac{d_2^2}{w^2+h^2}\right) \tag{19}$$

3 试验

3.1 试验环境

在本试验中,所有模型的训练均在相同环境条件下进行,GPU 为 NVIDIA GeForce RTX 4090,CPU 为 Intel(R) Xeon(R) 8352V,显存为 24 GB,操作系统为 Linux,以 Python 3.8.10 作为编程语言,深度学习框架 Pytorch 2.0.0,cuda11.8。

训练参数配置为:输入图像尺寸 640 像素 × 640 像素,迭代次数 300,每批次处理的数据量(batch size)16,优化器选用 SGD,初始学习率 0.01。

3.2 评价指标

采用多种性能指标评估模型性能,包括精确度(Precision,P)、召回率(Recall,R)以及平均精度均值(mean Average Precision,mAP)。此外,为了衡量模型计算效率与资源占用情况,指标还包括浮点运算量(FLOPs)、模型参数量(Parameters)以及模型权重文件大小(Model_size),这些指标可为评估模型实际部署可行性提供重要参考。

3.3 试验结果与分析

3.3.1 替换主干网络对模型的影响

为验证优化后的 Ghost_HGNetV2 对模型的影响,分别将 YOLO v8n 的主干网络 Darknet - 53 替换为 HGNetV2 及 Ghost_HGNetV2,设计了对比试验,试验结果如表 1 所示。

表 1 替换主干网络对模型的影响

Tab.1 Impact of replacing backbone network on model

主干网络	P/%	R/%	mAP@0.5/%	mAP@0.5 ~ 0.95/%	浮点运算量	参数量	模型权重文件大小/MB
Darknet - 53	93.6	88.1	94.4	75.9	8.2×10^9	3.01×10^9	6.3
HGNetV2	94.2	91.5	95.8	77.6	7.0×10^9	2.35×10^9	5.0
Ghost_HGNetV2	94.4	90.6	95.9	78.7	6.9×10^9	2.31×10^9	4.9

由表 1 可知，使用 HGNetV2 作为 YOLO v8n 的主干网络，较原模型精确度、召回率、mAP@0.5、mAP@0.5～0.95 分别提高 0.6、3.4、1.4、1.7 个百分点，浮点运算量、参数量、模型权重文件大小分别下降 14.6%、21.9%、20.6%，实现了模型轻量化。并且，使用 GhostConv 改进 HGNetV2 后，模型浮点运算量和参数量进一步减少，模型准确率也略有提高。

3.3.2　改进特征融合模块对模型的影响

为探究改进后的 C2f_Star 模块对模型的影响，设计了 4 组在颈部网络中使用不同特征融合模块的试验，其中，4 个模块均为原 C2f 模块为试验 1，4 个模块均为改进的 C2f_Star 为试验 2，后 3 个 C2f 模块为改进的 C2f_Star 为试验 3，后 2 个 C2f 模块为改进的 C2f_Star 为试验 4，试验结果如表 2 所示。

表 2　改进特征融合模块对模型的影响

Tab. 2　Impact of improving feature fusion module on model

颈部网络	*P*/%	*R*/%	mAP@0.5/%	mAP@0.5～0.95/%	浮点运算量	参数量	模型权重文件大小/MB
C2f	94.4	90.6	95.9	78.7	6.9×10^9	2.31×10^6	4.9
C2f_Star_4	93.8	90.3	95.9	78.9	6.6×10^9	2.11×10^6	4.5
C2f_Star_3	94.6	91.6	96.3	78.9	6.7×10^9	2.14×10^6	4.6
C2f_Star_2	94.0	90.1	95.5	78.7	6.7×10^9	2.15×10^6	4.6

由表 2 可知，与原有模块相比，全部使用改进的 C2f_Star 模块，模型具有最少的浮点运算量、参数量和模型权重文件大小，但其精确度和召回率分别下降 0.6、0.3 个百分点；使用部分改进的 C2f_Star 模块，模型浮点运算量、参数量、模型权重文件大小分别下降 2×10^8、1.7×10^5 和 0.3 MB，但后 2 个为改进的 C2f_Star 模块的模型精度、召回率和平均精度均有下降，而后 3 个为改进的 C2f_Star 模块的模型精度、召回率、mAP@0.5、mAP@0.5～0.95 分别提高 0.2、1.0、0.4、0.2 个百分点，因此，本研究使用改进的 C2f_Star 替换颈部网络中的后 3 个特征融合模块，有效地实现了更强的特征表达，同时使模型更加轻量。

3.3.3　不同注意力机制对模型的影响

为了探究添加注意力机制对病害特征提取能力的提升效果，设计了添加不同注意力机制的对比试验，将 CBAM[24]、EMA[25]、SimAM 注意力机制分别添加到模型的颈部网络中，试验结果如表 3 所示。

表 3　不同注意力机制对模型的影响

Tab. 3　Impact of embedding different attention mechanisms on model

注意力机制	*P*/%	*R*/%	mAP@0.5/%	mAP@0.5～0.95/%	浮点运算量	参数量	模型权重大小/MB
未添加	94.6	91.6	96.3	78.9	6.7×10^9	2.14×10^6	4.6
CBAM	93.0	92.3	96.3	79.9	6.7×10^9	2.24×10^6	4.8
EMA	93.1	91.6	96.5	80.3	6.9×10^9	2.16×10^6	4.6
SimAM	95.2	91.8	96.5	80.9	6.7×10^9	2.14×10^6	4.6

由表 3 可知，与未添加注意力机制相比，添加 CBAM 和 EMA 后模型的精度均有下降，而添加 SimAM 注意力机制后模型精确度提高 0.6 个百分点，mAP@0.5、mAP@0.5～0.95 分别提升 0.2、2.0 个百分点，其他性能均优于 CBAM 和 EMA 注意力机制，且具有更小的浮点运算量和参数量，这表明 SimAM 注意力机制在提升模型性能方面表现更佳。

为了直观展现不同注意力机制嵌入后的效果，采用 GradCAMPlus 生成热力图进行可视化对比分析，结果如图 6 所示。此热力图通过色彩编码直观反映各区域的数据密度或数值强度，其中暖色调（如红色）通常表示较高的数值或密集度，而冷色调（如蓝色）则对应较低的数值或稀疏度。观察图 6 可见，在引入 SimAM 注意力机制后，热力图中的红色区域显著加深，这表明模型捕捉到了更为丰富且显著的病斑特征信息，同时有效削弱了背景噪声的

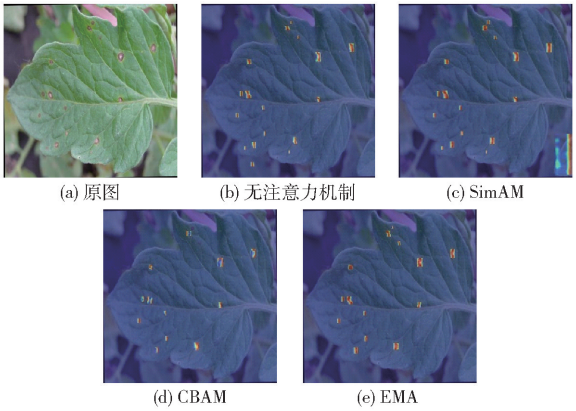


图 6　嵌入不同注意力机制可视化结果

Fig. 6　Embedding different attention mechanisms to visualize results

干扰,而嵌入 CBAM 与 EMA 注意力机制后模型都存在一定程度的漏检现象。

3.3.4 不同损失函数对模型的影响

为分析不同损失函数对模型检测结果的影响,选用 CIoU、MPDIoU、Focaler - IoU 和 Wise - IoU、Wise_Focaler_MPDIoU 4 种损失函数进行对比试验,试验结果如表 4 所示。

表 4 不同损失函数对模型的影响

Tab. 4 Impact of different loss functions on model

%

损失函数	*P*	*R*	mAP@0.5	mAP@0.5~0.95
CIoU	95.2	91.8	96.5	80.9
MPDIoU	93.5	91.4	95.7	78.3
Focaler - IoU	95.0	92.5	95.9	79.9
Wise - IoU	95.6	91.0	96.4	79.5
Wise_Focaler_MPDIoU	96.0	93.0	97.4	80.9

由表 4 可知,对损失函数的改进可以很好地提升模型性能,使用混合损失函数 Wise_Focaler_MPDIoU 后,模型精确率、召回率、mAP@0.5 和 mAP@0.5~0.95 均达到最优水平。

图 7 为不同损失函数对模型损失影响的对比曲线。对图 7 中曲线分析发现,MPDIoU 和 YOLO v8n 原本使用的 CIoU 损失值相近,但略优于 CIoU,而 WIoU 和 Focaler - IoU 的损失值明显低于 CIoU。而本研究所设计的混合损失函数 Wise_Focaler_MPDIoU 结合了 3 个损失函数的优势,展现出更低的检测框损失值,检测性能最佳。

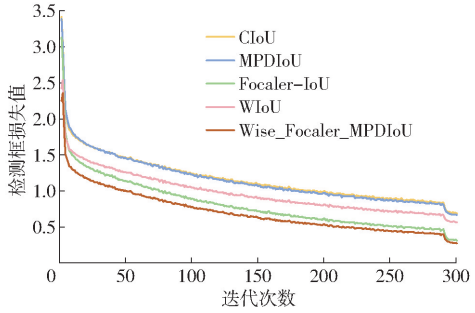


图 7 不同损失函数的损失变化曲线

Fig. 7 Loss variation curves for different loss functions

3.3.5 消融试验

为验证本研究提出的各改进模块的有效性,以原 YOLO v8n 为基准模型,设计了 5 组消融试验,试验结果如表 5 所示。

表 5 消融试验结果

Tab. 5 Ablation test results

编号	Ghost_HGNetV2	C2f_Star	SimAM	Wise_Focaler_MPDIoU	*P*/%	*R*/%	mAP@0.5/%	mAP@0.5~0.95/%	浮点运算量	参数量	模型权重文件大小/MB
1	×	×	×	×	93.6	88.1	94.4	75.9	8.2×10^9	3.01×10^6	6.3
2	√	×	×	×	94.4	90.6	95.9	78.7	6.9×10^9	2.31×10^6	4.9
3	√	√	×	×	94.6	91.6	96.3	78.9	6.7×10^9	2.14×10^6	4.6
4	√	√	√	×	95.2	91.8	96.5	80.9	6.7×10^9	2.14×10^6	4.6
5	√	√	√	√	96.0	93.0	97.4	80.9	6.7×10^9	2.14×10^6	4.6

注:√表示采用此模块,×表示不采用此模块。

由表 5 可知,使用 Ghost_HGNetV2 模型改进主干网络后,模型精确度、召回率、mAP@0.5、mAP@0.5~0.95 分别提高 0.8、2.5、1.5、2.8 个百分点,浮点运算量、参数量以及模型权重文件大小分别降低 15.8%、23.3% 和 22.2%。在改进主干网络的基础上,改进网络特征融合为 C2f_Star 模块后,模型的参数量、浮点运算量和模型权重文件大小进一步降低,而模型精确度、召回率、mAP@0.5、mAP@0.5~0.95 分别提高 0.2、1.0、0.4、0.2 个百分点。继续添加 SimAM 注意力机制后,模型参数量未发生变化,但精确度、召回率、mAP@0.5、mAP@0.5~0.95 分别提高 0.6、0.2、0.2、2.0 个百分点。最后使用混合损失函数 Wise_Focaler_MPDIoU 后,模型精确度、召回率、mAP@0.5 和 mAP@0.5~0.95 分别提升 0.8、1.2、0.9、0 个百分点。

综上,改进后的 GSSW - YOLO v8n 模型与原始 YOLO v8n 基准模型相比,精确度、召回率、mAP@0.5、mAP@0.5~0.95 分别提高 2.4、4.9、3.0、5.0 个百分点,并且模型浮点运算量、参数量和模型权重文件大小分别下降 18.3%、28.9% 和 26.9%,有效地提高了对番茄叶片病害检测的准确率,同时实现了轻量化。

3.3.6 不同算法对比试验

为验证本研究模型的有效性,选取当前较为热门的目标检测模型进行对比。将本研究模型 GSSW - YOLO v8n 与 Faster R - CNN、YOLO v5n、YOLO v6n、YOLO v7 - tiny、YOLO v8n 模型在相同的条件下进行对比试验,试验结果如表 6 所示。

由表 6 可知,Faster R - CNN 在精确度、召回率方面表现一般,浮点运算量和参数量最多,不太符合本研究的轻量化要求;相较于 YOLO v6n、YOLO v7 -

表 6　不同检测模型检测结果

Tab. 6　Results of different detection models

模型	*P*/%	*R*/%	mAP@0.5/%	mAP@0.5～95/%	浮点运算量	参数量	模型权重文件大小/MB
Faster R－CNN	85.8	73.6	84.1	72.6	2.393×10^{11}	4.16×10^{7}	321.2
YOLO v5n	90.3	86.8	92.0	71.3	7.2×10^{9}	2.51×10^{6}	5.3
YOLO v6n	88.5	80.1	87.5	66.8	1.19×10^{10}	4.23×10^{6}	8.7
YOLO v7－tiny	85.9	82.0	88.2	55.9	1.32×10^{10}	6.03×10^{6}	12.3
YOLO v8n	93.6	88.1	94.4	75.9	8.2×10^{9}	3.01×10^{6}	6.3
GSSW－YOLO v8n	96.0	93.0	97.4	80.9	6.7×10^{9}	2.14×10^{6}	4.6

tiny 模型，YOLO v8n 和 YOLO v5n 具有更小的浮点运算量和更少的参数量，精确度、召回率、mAP 反而表现更佳；而 YOLO v5n 的浮点运算量、参数量及模型权重文件最小，但是精确度、召回率、mAP 表现均不如 YOLO v8n；改进模型 GSSW－YOLO v8n 相比模型 YOLO v5n 的浮点运算量、参数量及模型权重文件更小，并且精确度、召回率、mAP 相比于 YOLO v8n 模型更高。

提出的 GSSW－YOLO v8 模型在使用的番茄叶片病害数据集上的平均检测精度达 97.4%，相较于 Faster R－CNN、YOLO v5n、YOLO v6n、YOLO v7－tiny 模型分别高 13.3、5.4、9.9、9.2 个百分点，且模型浮点运算量、参数量和模型权重文件最小。

3.3.7　算法验证

为了更直观地对比改进后模型检测效果，选取消融试验中综合检测性能表现较优的 YOLO v5n、YOLO v8n 以及改进的 GSSW－YOLO v8n 算法进行检测效果可视化对比。在番茄叶片病害数据集的测试集中选取 8 种类别的番茄叶片病害图像进行测试，其检测结果如图 8 所示。

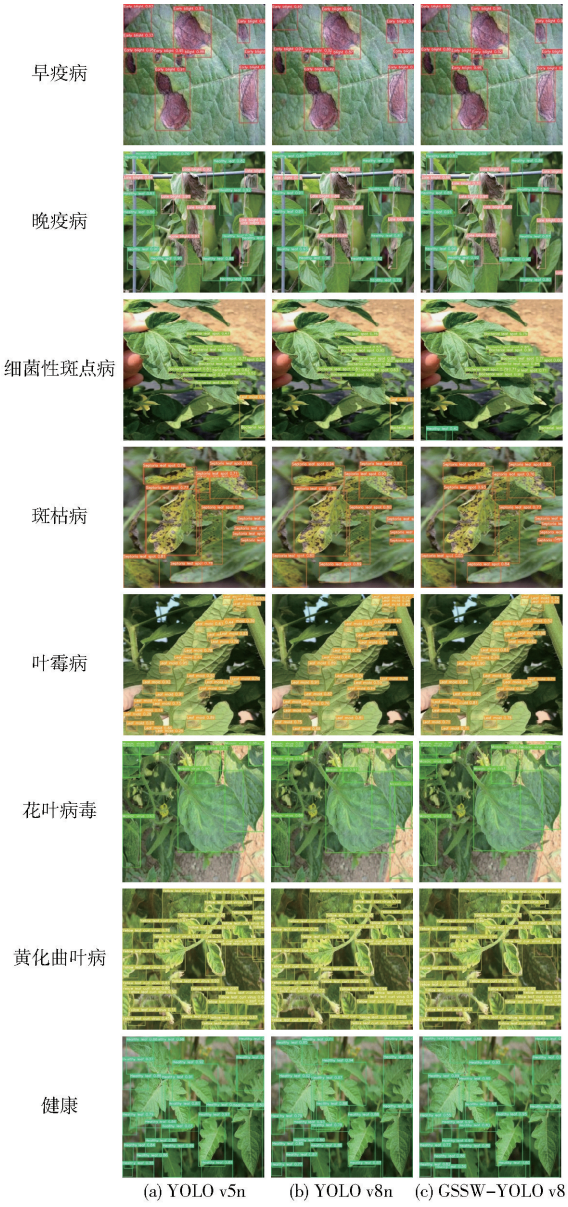


图 8　不同模型检测番茄叶片病害效果

Fig. 8　Effectiveness of different models for detecting leaf diseases in tomato

由图 8 可以看出，对于早疫病识别，GSSW－YOLO v8n 的整体识别精度更高；对于背景较为复杂的晚疫病识别场景，YOLO v5n 和 YOLO v8n 均存在漏检现象；对于病斑较小的细菌性斑点病识别，YOLO v5n 存在 2 处误检，YOLO v8n 存在 1 处误检，而 GSSW－YOLO v8n 对于小目标检测无漏检；对于斑枯病识别，YOLO v5n 检测精度较低，YOLO v8n 存在一处漏检；对于叶霉病、花叶病毒和健康叶片识别，YOLO v5n 和 YOLO v8n 识别精度均低于GSSW－YOLO v8n；对于较多目标场景的黄化曲叶病识别，YOLO v8n 存在 1 处漏检，GSSW－YOLO v8n 具备更高的识别精度。综上分析表明本研究提出的改进模型在小目标、多目标、背景复杂的场景下均具有更好的检测性能。

4　结论

（1）在相同的试验条件下，改进的 GSSW－YOLO v8n 模型相比于原 YOLO v8n 基线模型，精确度、召回率、mAP@0.5、mAP@0.5～0.95 分别提高 2.4、4.9、3.0、5.0 个百分点，并且模型浮点运算量、参数量和模型权重文件大小分别下降 18.3%、28.9% 和 26.9%。

（2）通过与其他主流目标检测模型 Faster R－CNN、YOLO v5n、YOLO v6n、YOLO v7－tiny 对比，改

进后的模型具有最少的浮点运算量和参数量,并且平均检测精度分别高出 13.3、5.4、9.9、9.2 个百分点。

(3)本研究提出的 GSSW - YOLO v8 模型可为番茄叶片病害检测在嵌入式设备上的部署应用提供方法支持。

参 考 文 献

[1] 朱立强. 大棚番茄种植管理与病虫害防治技术研究[J]. 河北农机,2023(24):162 - 164.

[2] 张林鎧,巴音塔娜,曾庆松. 基于 StyleGAN2 - ADA 和改进 YOLO v7 的葡萄叶片早期病害检测方法[J]. 农业机械学报,2024,55(1):241 - 252.
ZHANG Linxuan, BA Yintana, ZENG Qingsong. Grape disease detection method based on StyleGAN2 - ADA and improved YOLO v7 [J]. Transactions of the Chinese Society for Agricultural Machinery, 2024,55(1):241 - 252. (in Chinese)

[3] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779 - 788.

[4] 金婷婷,房建东,赵于东. 注意力特征融合的番茄叶部早期病斑诊断算法[J]. 电子测量技术,2024,47(4):156 - 164.
JIN Tingting, FANG Jiandong, ZHAO Yudong. Early lesion diagnosis algorithm of tomato leaf based on attention feature fusion [J]. Electronic Measurement Technology,2024,47(4):156 - 164. (in Chinese)

[5] QI J, LIU X, LIU K, et al. An improved YOLOv5 model based on visual attention mechanism: application to recognition of tomato virus disease[J]. Computers and Electronics in Agriculture, 2022, 194: 106780.

[6] 韩鑫,徐衍向,封润泽,等. 基于红外热成像和改进 YOLO v5 的作物病害早期识别[J]. 农业机械学报,2023,54(12):300 - 307,375.
HAN Xin, XU Yanxiang, FENG Runze, et al. Early identification of crop diseases based on infrared thermography and improved YOLO v5[J]. Transactions of the Chinese Society for Agricultural Machinery,2023,54(12):300 - 307,375. (in Chinese)

[7] 张领先,景嘉平,李淑菲,等. 基于图像自动标注与改进 YOLO v5 的番茄病害识别系统[J]. 农业机械学报,2023,54(11):198 - 207.
ZHANG Lingxian, JING Jiaping, LI Shufei, et al. Tomato disease recognition system based on image automatic labeling and improved YOLO v5[J]. Transactions of the Chinese Society for Agricultural Machinery,2023,54(11):198 - 207. (in Chinese)

[8] 王法安,何忠平,张兆国,等. 基于 PN - YOLO v8s - Pruned 的轻量化三七收获目标检测方法[J]. 农业机械学报,2024,55(11):171 - 183.
WANG Faan, HE Zhongping, ZHANG Zhaoguo, et al. Lightweight object detection method for *Panax notoginseng* based on PN -YOLO v8s - Pruned[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024,55(11):171 - 183. (in Chinese)

[9] 王会征,孙良晨,李新龙,等. 基于改进 YOLOv7 - tiny 的番茄叶片病虫害检测方法[J]. 农业工程学报,2024,40(10):194 - 202.
WANG Huizheng, SUN Liangchen, LI Xinlong, et al. Detecting tomato leaf pests and diseases using improved YOLOv7 - tiny [J]. Transactions of the CSAE, 2024, 40(10): 194 - 202. (in Chinese)

[10] 方晨晨,石繁槐. 基于改进深度残差网络的番茄病害图像识别[J]. 计算机应用,2020,40(增刊 1):203 - 208.
FANG Chenchen, SHI Fanhuai. Image recognition of tomato diseases based on improved deep residual network[J]. Journal of Computer Applications,2020,40(Supp. 1):203 - 208. (in Chinese)

[11] 王娜,陈勇,崔艳荣,等. 基于改进轻量化 YOLO v5n 的番茄叶片病害识别方法[J]. 江苏农业科学,2024,52(8):192 - 199.
WANG Na, CHEN Yong, CUI Yanrong, et al. Tomato leaf disease identification method based on improved lightweight YOLO v5n[J]. Jiangsu Agricultural Sciences,2024,52(8):192 - 199. (in Chinese)

[12] 杨佳昊,左昊轩,黄祺成,等. 基于 YOLO v5s 的作物叶片病害检测模型轻量化方法[J]. 农业机械学报,2023,54(增刊 1):222 - 229.
YANG Jiahao, ZUO Haoxuan, HUANG Qicheng, et al. Lightweight method for crop leaf disease detection model based on YOLO v5s[J]. Transactions of the Chinese Society for Agricultural Machinery,2023,54(Supp. 1):222 - 229. (in Chinese)

[13] 王金鹏,何萌,甄乾广,等. 基于改进 COF - YOLO v8n 的油茶果静态与动态检测计数方法[J]. 农业机械学报,2024,55(4):193 - 203.
WANG Jinpeng, HE Meng, ZHEN Qianguang, et al. *Camellia oleifera* fruit static and dynamic detection counting based on improved COF - YOLO v8n[J]. Transactions of the Chinese Society for Agricultural Machinery,2024,55(4):193 - 203. (in Chinese)

[14] 王旺,王福顺,张伟进,等. 基于改进 YOLO v8s 的羊只行为识别方法[J]. 农业机械学报,2024,55(7):325 - 335,344.
WANG Wang, WANG Fushun, ZHANG Weijin, et al. Sheep behavior recognition method based on improved YOLO v8s[J]. Transactions of the Chinese Society for Agricultural Machinery,2024,55(7):325 - 335,344. (in Chinese)

[15] ZHAO Y, LV W, XU S, et al. Detrs beat Yolos on real-time object detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 16965 - 16974.

[16] HAN K, WANG Y, TIAN Q, et al. Ghostnet: more features from cheap operations[C]// Proceedings of the IEEE/CVF

Conference on Computer Vision and Pattern Recognition, 2020: 1580 - 1589.
[17] MA X, DAI X, BAI Y, et al. Rewrite the stars[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 5694 - 5703.
[18] YANG L, ZHANG R Y, LI L, et al. Simam: a simple, parameter-free attention module for convolutional neural networks[C]// International Conference on Machine Learning. PMLR, 2021: 11863 - 11874.
[19] 张蔚然,杜岳峰,栗晓宇,等. 基于 FSLYOLO v8n 的玉米籽粒收获质量在线检测方法研究[J]. 农业机械学报,2024, 55(8):253 - 265.
ZHANG Weiran, DU Yuefeng, LI Xiaoyu, et al. Online detection method of corn kernel quality based on FSLYOLO v8n[J]. Transactions of the Chinese Society for Agricultural Machinery,2024,55(8):253 - 265. (in Chinese)
[20] ZHENG Z, WANG P, REN D, et al. Enhancing geometric factors in model learning and inference for object detection and instance segmentation[J]. IEEE Transactions on Cybernetics, 2021, 52(8): 8574 - 8586.
[21] SILIANG M, YONG X. MPDIoU: a loss for efficient and accurate bounding box regression[J]. arXiv preprint arXiv:2307. 07662, 2023.
[22] ZHANG H, ZHANG S. Focaler - IoU: more focused intersection over union loss[J]. arXiv preprint arXiv:2401. 10525, 2024.
[23] TONG Z, CHEN Y, XU Z, et al. Wise - IoU: bounding box regression loss with dynamic focusing mechanism[J]. arXiv preprint arXiv:2301. 10051, 2023.
[24] WOO S, PARK J, LEE J Y, et al. Cbam: convolutional block attention module[C]// Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3 - 19.
[25] OUYANG D, HE S, ZHANG G, et al. Efficient multi-scale attention module with cross-spatial learning[C]// ICASSP 2023—2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2023: 1 - 5.