

引用格式:

黄政, 张涛, 孔万仔, 赵丹枫, 魏泉苗. 基于YOLOv8n的梨树叶片病害检测模型[J]. 湖南农业大学学报(自然科学版), 2025, 51(2): 113–121.

HUANG Z, ZHANG T, KONG W Z, ZHAO D F, WEI Q M. Detection model of pear leaf disease based on YOLOv8n[J]. Journal of Hunan Agricultural University(Natural Sciences), 2025, 51(2): 113–121.

投稿网址: <http://xb.hunau.edu.cn>



## 基于YOLOv8n的梨树叶片病害检测模型

黄政<sup>1</sup>, 张涛<sup>1</sup>, 孔万仔<sup>1</sup>, 赵丹枫<sup>1\*</sup>, 魏泉苗<sup>2</sup>

(1.上海海洋大学信息学院, 上海 201306; 2.自然资源部东海局, 上海 200137)

**摘要:** 针对传统目标检测模型对自然场景下梨树叶片病害检测存在精度低、模型参数量大等问题, 提出一种基于YOLOv8n的梨树叶片病害检测改进模型。首先, 使用RepGhostNet改进主干网络, 利用结构重参数化实现特征的隐式重用, 在提升网络特征提取能力的同时使网络更加轻量化。其次, 引入双层路由注意力机制, 通过查询自适应的方式降低模型对不相关特征的关注, 提高模型对关键信息的敏感性, 增强网络的表征能力和特征融合能力。最后, 使用Inner-SIoU损失函数优化边界框回归, 加快模型收敛速度, 提高识别精度。结果表明: 改进后的模型能够有效对梨树叶片病害进行检测, 在DiaMOS Plant数据集上对梨树叶片病害的检测平均精准度mAP@50达到0.901, 相较于原模型提高了5.6%; 而模型参数量仅为 $2.4 \times 10^6$ 个, 计算量仅为7 GFLOPs, 相较于原模型分别降低了20.00%和13.58%。与SSD、Faster-RCNN、YOLOv5n、YOLOv8s等主流目标检测模型相比, 改进的模型不仅平均精准度有所提高, 而且参数量和计算量均减少。

**关键词:** 梨树叶片病害检测; YOLOv8n; 模型轻量化; RepGhostNet; 双层路由注意力机制

中图分类号: TP391.41; S432.9

文献标志码: A

文章编号: 1007-1032(2025)02-0113-09

## Detection model of pear leaf disease based on YOLOv8n

HUANG Zheng<sup>1</sup>, ZHANG Tao<sup>1</sup>, KONG Wanzai<sup>1</sup>, ZHAO Danfeng<sup>1\*</sup>, WEI Quanmiao<sup>2</sup>

(1.College of Information Technology, Shanghai Ocean University, Shanghai 201306, China; 2.East China Sea Bureau, Ministry of Natural Resources, Shanghai 200137, China)

**Abstract:** In response to the challenges of low accuracy and large model parameters number in traditional object detection models for detecting pear leaf diseases in natural scenes, an improved model for pear leaf disease detection based on YOLOv8n was proposed. Firstly, the RepGhostNet was employed to enhance the backbone network, which utilized structural reparameterization to achieve implicit feature reuse, thereby enhancing the network's feature extraction capabilities while maintaining lightweight characteristics. Secondly, the bi-level routing attention mechanism was utilized to dynamically filter out less relevant key-value pairs at a coarse region level, thereby lowering attention to irrelevant features and increasing sensitivity to essential information, and enhancing the network's representational and feature fusion capabilities. Finally, the Inner-SIoU loss function was employed to optimize bounding box regression, accelerate model convergence and improve recognition accuracy. The results showed that the improved model effectively detected pear leaf diseases, achieving an mAP@50 score of 0.901 on the DiaMOS Plant dataset. Compared to the original model, the improved model exhibited a notable 5.6% enhancement in performance, with reduced model parameter quantity( $2.4 \times 10^6$ ) and computations(7 GFLOPs), representing a 20.00% and 13.58% decrease, respectively. When compared to mainstream object detection models such as SSD, Faster-RCNN, YOLOv5n, and YOLOv8s, the enhanced model showed an increase in average precision, accompanied by reductions in both parameter and computation loads.

收稿日期: 2023-12-08

修回日期: 2024-01-04

基金项目: 国家自然科学基金青年项目(42106190)

作者简介: 黄政(1997—), 男, 江西赣州人, 硕士研究生, 主要从事深度学习相关研究, [henryhwong@foxmail.com](mailto:henryhwong@foxmail.com); \*通信作者, 赵丹枫, 博士, 副教授, 主要从事农业智能化研究, [dfzhao@shou.edu.cn](mailto:dfzhao@shou.edu.cn)

**Keywords:** pear leaf disease detection; YOLOv8n; model lightweighting; RepGhostNet; bi-level routing attention mechanism

梨因其丰富的营养价值和独特的口感在中国广受消费者喜爱,同时,梨也是我国种植面积仅次于苹果和柑橘的重要水果品种<sup>[1]</sup>。梨树的种植不仅为果农带来了经济效益,也促进了当地农业经济的发展。然而,梨树在生长过程中常常会受到各种病害的侵袭,这些病害往往最先使叶片表现出症状<sup>[2]</sup>,如果不能及时准确地诊断并采取有效的防治措施将会严重影响梨树的生长发育,进而导致果实产量和质量降低,给果农造成经济损失。以往对果树叶片进行病害检查主要依靠人工完成,然而,这种方式耗时费力,尤其是在大规模的果园中,人工检查的效率非常低,容易导致防治不及时。此外,在确定病害种类时,由于植株病害的复杂性和个体经验的限制,很容易造成误诊<sup>[3]</sup>,导致选择错误的防治方法。因此,寻求一种高效、准确可靠的梨树叶片病害检测方法是提升病害防控效率的关键。

近年来,在人工智能的快速推动下,融合图像处理技术和深度学习模型实现植物病害检测,为农业生产注入了新动能。深度学习通过多层次的神经网络结构自动提取图像中的特征信息并进行学习,提高了植物病害识别的准确性和效率。目前,利用深度学习技术识别农作物叶片病害已经取得了显著进展。例如,朱帅等<sup>[4]</sup>基于多特征融合改进深度残差网络,使其对多种果树叶片病害的识别准确率的平均值达到了99.4%;孙俊等<sup>[5]</sup>结合批归一化与全局池化改进AlexNet,对14种不同植物的26类病害叶片的识别准确率达到99.56%;方文博等<sup>[6]</sup>通过空洞卷积和重构特征金字塔结构改进YOLOv5s模型,对大豆叶片虫洞的平均识别准确率达到95.24%。虽然上述研究在植物叶片病害识别方面取得了较高的精度,但是这些工作都是基于PlantVillage数据集或自建数据集进行的,这类数据集中的叶片图像通常为单独拍摄,具有单一背景,无法真实反映自然环境下的场景。

在自然场景下,由于光线角度不同、叶片重叠、不同季节叶片颜色发生变化和茎干遮挡等原

因,果树叶片图像通常具有更加复杂的背景,这些因素的存在极大地增加了叶片病害特征的提取难度,可能会导致叶片病害的识别准确率下降。因此,如何优化模型以提高其对自然场景下叶片病害的识别准确率是当前植物叶片病害检测和识别研究面临的重要问题。当前,许多研究通过修改模型的网络结构以增强模型对特征的提取能力,并引入注意力机制以增强特征融合效果,提高模型在自然场景下叶片病害的识别精度。盛帅等<sup>[7]</sup>通过增强模型对深层语义特征的关注,提出了一种自适应的Dynamic-YOLOX模型,该模型对复杂背景下的苹果叶片病害检测准确率达84.63%。李鑫然等<sup>[8]</sup>基于特征金字塔网络深层特征融合提出了一种改进的Faster R-CNN模型,该模型对复杂背景下苹果叶片病害检测的平均精度达到82.48%。然而,上述2种模型检测精度相对较低,与农业实际应用场景的需求仍有一定差距。CHEN等<sup>[9]</sup>采用多维特征融合神经网络和注意力机制将特征融合到Inception结构的每一层,以提高模型分类能力,该模型对自然环境下油茶叶片病害的识别准确率达86.78%。LI等<sup>[10]</sup>通过融合视觉Transformer的卷积神经网络提取有效特征的方法,对自然背景下猕猴桃叶片病害的识别准确率达到98.78%。虽然这2种模型的精度均有了一定程度的提高,但是模型参数量大、计算复杂度高,难以在计算资源有限的农业终端设备上高效部署,无法满足现代农业生产的实际需求。由此可见,自然环境下植物叶片病害识别模型在高精度和低算力之间的平衡仍然是一个亟待解决的问题。

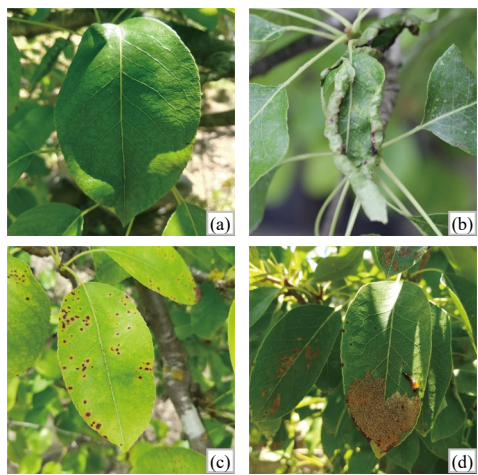
鉴于此,本文从提升真实场景下叶片病害识别精度和模型轻量化两方面入手,提出一种基于YOLOv8n的梨树叶片病害检测改进模型。该模型使用RepGhostNet<sup>[11]</sup>改进主干网络,利用结构重参数化实现特征的隐式重用,在提升网络特征提取能力的同时使网络更加轻量化;同时,引入双层路由注意力(bi-level routing attention, BRA)<sup>[12]</sup>机制,通过查询自适应的方式降低其对不相关特征的关注,提高对关键信息的敏感性,从而增强网

络的表征能力和特征融合能力,以降低复杂背景干扰,提高识别精度。此外,利用动态的查询感知稀疏性减少计算量和内存使用,在保证精度的同时尽可能地降低模型的参数量和计算量。在此基础上再采用Inner-SIoU优化损失函数,加快模型收敛速度,进一步提高识别精度,使其更符合自然场景下梨树叶片病害检测需求,以便后续更好地应用于农业生产实践。

## 1 梨树叶片病害数据集及其预处理

### 1.1 DiaMOS Plant梨树叶片病害数据集

实验采用的数据集为公开的梨树叶片病害数据集DiaMOS Plant<sup>[13]</sup>。DiaMOS Plant数据集涵盖了梨树整个生长期不同类别(包括健康、卷叶病、斑病和蛭蚰虫病害4个类别)以及这些类别中不同病害程度的3 006张梨树叶片图像。该数据集中的所有图像均是在果园的自然环境下,针对不同天气和光照角度拍摄叶片得到的,而且每张图像都引入了复杂背景的元素(例如包含多个叶片或树枝等)。图1所示为DiaMOS Plant数据集中4个类别的梨树叶片图像示例。



(a) 健康叶片; (b) 卷叶病叶片; (c) 斑病叶片; (d) 蛭蚰虫害叶片

图1 DiaMOS Plant 梨树叶片病害图像示例

Fig. 1 Example of DiaMOS Plant pear leaf disease image

### 1.2 数据预处理

为了提高数据的多样性,增强模型的泛化性和鲁棒性,采用随机旋转、随机裁剪、镜像翻转、随机颜色抖动和添加高斯噪声等数据增强方式将数据集图像扩充至6 012张;然后,使用

LabelImg软件对扩充后的图像进行人工标注,并按照1:8:1的数量比例随机将数据集分为测试集、训练集和验证集。

## 2 目标检测模型

### 2.1 YOLOv8模型

YOLOv8是一种先进的目标检测模型。目前YOLOv8模型分为YOLOv8n、YOLOv8s、YOLOv8m、YOLOv8l和YOLOv8x这5种不同参数规模的版本,其中,YOLOv8n具有最小的参数量和计算量,更易于部署在低算力的嵌入式设备中,因此,本文采用YOLOv8n作为基线模型。

YOLOv8n的网络结构包括主干网络、颈部网络和检测头3个部分。主干网络用于特征提取,颈部网络用于特征融合,检测头则用于对网络中的特征图进行处理,最终通过卷积操作生成目标的位置信息和类别信息。YOLOv8的主干网络遵循CSPDarknet53的结构设计,并采用了C2f模块来高效提取特征。颈部网络使用了PA-FPN网络结构来提升不同尺度特征的融合能力,而检测头则使用了3个不同尺度的解耦检测头,分别负责不同尺度的目标检测。

### 2.2 改进的YOLOv8n模型

由于YOLOv8n模型使用了大量的标准卷积和C2f模块,在提高模型检测精度的同时,增加了模型的参数量和计算量。此外,自然环境中的叶片图像往往具有复杂的背景,YOLOv8n模型在提取特征和进行特征融合时难以消除复杂背景的干扰,容易导致检测精度下降。针对这一问题,本文对YOLOv8n模型进行以下改进:首先,使用RepGhostNet改进主干网络,利用结构重参数化实现特征的隐式重用,在提升网络特征提取能力的同时使网络更加轻量化;其次,引入双层路由注意力机制,通过查询自适应的方式降低其对不相关特征的关注,提高其对关键信息的敏感性,从而增强网络的表征能力和特征融合能力,降低复杂背景的干扰;最后,采用Inner-SIoU损失函数替代原始CIoU损失函数,加快模型收敛速度。改进后的YOLOv8n模型结构如图2所示。

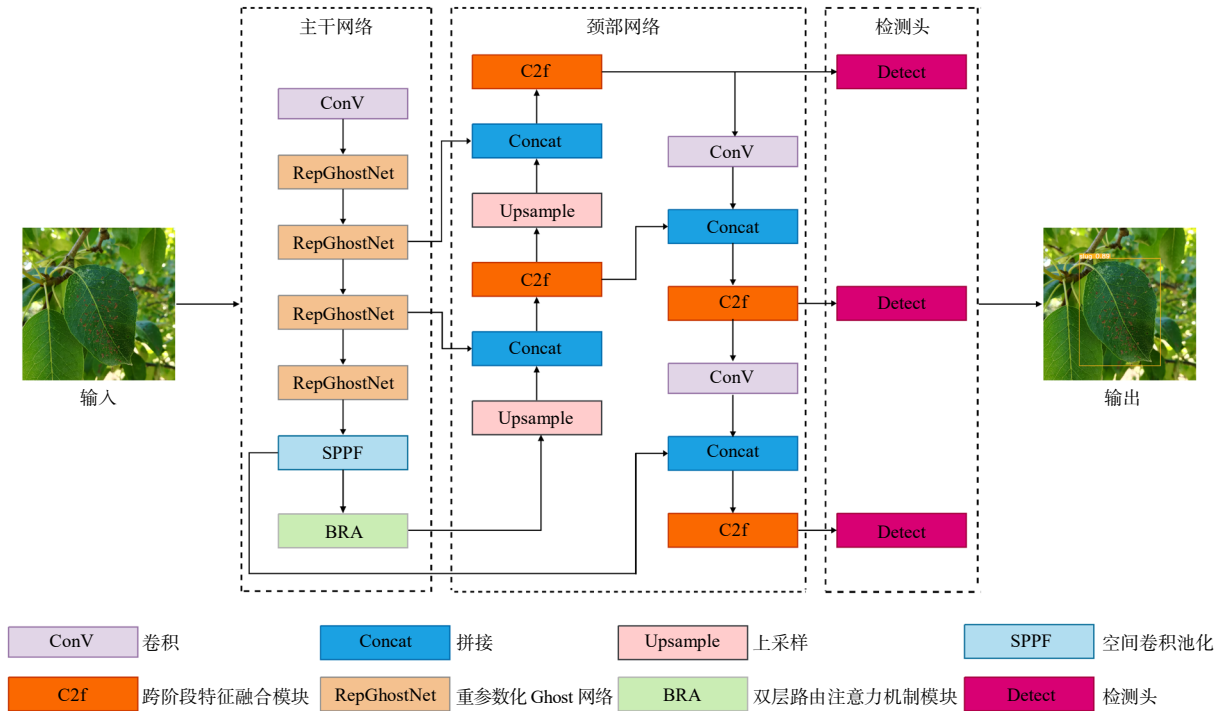


图2 改进的YOLOv8n模型

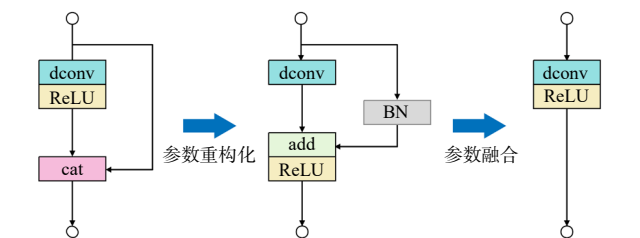
Fig.2 Improved YOLOv8n model

### 2.2.1 基于RepGhostNet的主干网络改进

YOLOv8n模型的主干网络采用了大量的标准卷积和C2f模块。在特征提取过程中，传统的卷积在生成大通道数的特征图时会耗费大量计算资源。然而，不同层的卷积输出的特征图通常包含大量未被充分利用的特征。为优化特征利用效率并减少计算量，可以通过连接来自不同层的已有特征图，从而实现特征复用。特征复用在轻量化设计中扮演关键角色。当前特征复用通常通过级联运算符以低成本重用其他层的特征映射并维持大通道数来实现。然而，基础特征复用和线性映射通过特征通道拼接来保持大通道数，虽然不引入额外的参数，但增加了硬件设备的计算成本。为了实现更高效特征复用，本文使用RepGhostNet替代YOLOv8n主干网络中的普通卷积和C2f模块。

RepGhostNet基于RepGhost<sup>[11]</sup>模块实现，而RepGhost则是在Ghost<sup>[14]</sup>模块的基础上进行结构重参数化演变而来的，其演变过程如图3所示。为提高硬件计算效率，在RepGhost模块中，特征拼接(cat)操作被替换为特征相加(add)操作，这种结构

不仅具有特征融合功能，而且不引入额外的模型参数，不增加推理时间。此外，在恒等映射分支中引入批归一化(batch normalization, BN)操作以满足重参数化规则，最终实现参数融合。这一过程使得模型在训练阶段能够充分利用复杂结构的性能优势，而在训练完成后则将其等价转换为简单的推理结构，无需额外的时间成本。基于这一特性，RepGhostNet在训练和推理阶段分别采用不同的架构，如图4所示。RepGhostNet既保留了更为丰富和抽象的特征表示能力的优势，又具备了高速推理能力。



dconv表示深度可分离卷积层(depthwise convolutional layer); ReLU表示修正线性单元(rectified linear unit)。

图3 Ghost模块演化为RepGhost模块

Fig.3 Evolution of the Ghost to the RepGhost



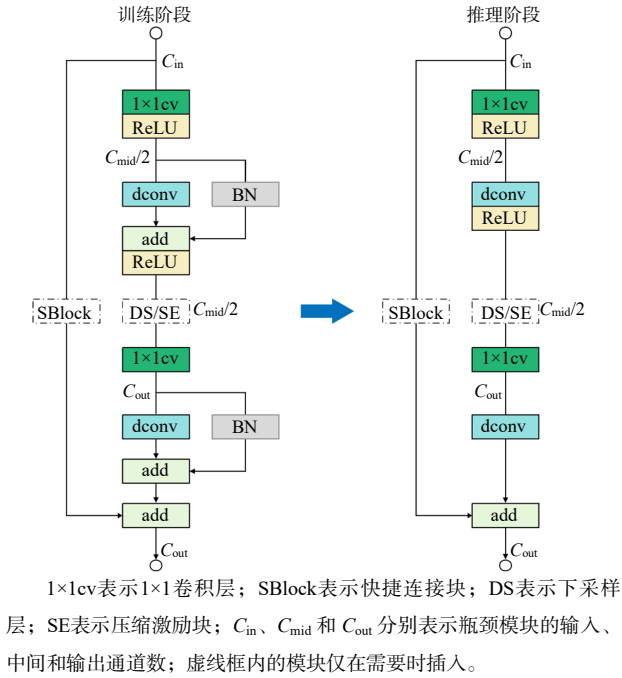


图4 RepGhostNet 网络结构

Fig.4 Network architecture of RepGhostNet

### 2.2.2 双层路由注意力机制的添加

由于自然环境下梨树叶片图像具有复杂的背景，为了使网络更加聚焦于关键特征，消除复杂背景的干扰，本文在网络中引入双层路由注意力机制。双层路由注意力机制是一种具备动态查询感知稀疏性的注意力机制，其主要思想是通过查询自适应的方式在粗区域级别上过滤掉与查询语义上最不相关的键值对向量，然后将剩余的键值对向量作为路由区，并在路由区中应用细粒度的Token-to-token注意力<sup>[12]</sup>，从而达到降低对不相关特征的关注度，提高对关键信息的敏感性的目的。

在图像中应用双层路由注意力机制获取特征图的过程见图5。假设原始输入图像的特征映射为  $X \in \mathbb{R}^{H \times W \times C}$  (其中,  $H$ 和 $W$ 分别为特征图的高度和宽度;  $C$ 为通道数), 然后通过线性投影导出3个分别代表查询、键和值的张量  $Q$ 、 $K$ 、 $V$  ( $Q, K, V \in \mathbb{R}^{H \times W \times C}$ ), 并且  $Q = X^r W^q$ ,  $K = X^r W^k$ ,  $V = X^r W^v$ , 其中  $W^q$ 、 $W^k$ 、 $W^v$  分别表示  $Q$ 、 $K$  和  $V$  对应的投影权重 ( $W^q, W^k, W^v \in \mathbb{R}^{C \times C}$ ),  $X^r$  为特征图在子区域上的嵌入矩阵。然后将  $Q$ 、 $K$ 、 $V$  分成  $S \times S$  的窗口, 为了从窗口中筛选出最相关的键值对向量, 首先计算  $Q$  和  $K$  的每个区域的平均值  $Q^r$  和  $K^r$ , 然后通过  $Q^r$  和  $K^r$  计算得到区域与区域之间的关联矩阵

$A^r$ ,  $A^r$  的计算公式如下:

$$A^r = Q^r (K^r)^T \quad (1)$$

为了从关联矩阵  $A^r$  中滤掉与查询语义上最不相关的键值对向量, 采用TopK算法筛选出每个区域中最相关的前  $k$  个连接, 从而得到路由索引矩阵  $I^r$ :

$$I^r = f_{\text{topK}}(A^r) \quad (2)$$

由于  $I^r$  中对应的路由区域可能会分散在整个特征图上, 而GPU计算依赖于凝聚内存操作, 即一次性加载数十个连续字节块, 为了让GPU更加高效地对这些区域进行处理, 收集  $K$  和  $V$  张量:

$$K^g = f_{\text{gather}}(K, I^r) \quad (3)$$

$$V^g = f_{\text{gather}}(V, I^r) \quad (4)$$

将  $K^g$  和  $V^g$  应用细粒度的Token-to-token注意力得到特征图矩阵  $O$ :

$$O = f_{\text{attention}}(Q, K^g, V^g) \quad (5)$$

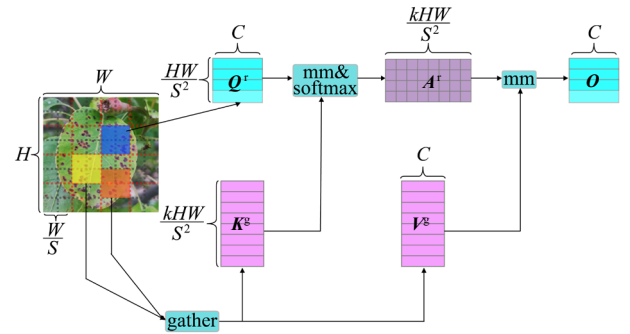


图5 双层路由注意力机制获取特征图的过程

Fig.5 Acquisition of feature map by bi-level routing attention mechanism

### 2.2.3 损失函数优化

在目标检测任务中, 通常使用矩形边界框来确定目标的位置和大小。模型预测边界框与真实边界框之间的重叠程度通常采用IoU(intersection over union)进行度量。YOLOv8n采用CIoU(complete intersection over union)<sup>[15]</sup>损失函数来优化预测边界框的位置。相较于传统IoU, CIoU不仅考虑了预测框和真实框中心点之间的欧式距离, 还考虑了预测框的纵横比与真实框相应比例的拟合程度。然而, 纵横比作为描述相对值的参数, 不能充分反映边界框长度和宽度的真实差异。在CIoU的基础上, ZHANG等<sup>[16]</sup>提出了EIoU(enhanced intersection over union), 以进一步考虑边界框长、宽差异; GEVORGYAN等<sup>[17]</sup>引入了真实框与预测框之间的向

量角度,提出了SIOU(structured intersection over union),进一步优化了损失值的计算。然而,当前基于IoU的边界框回归方法注重通过引入新的损失项以提高收敛速度,却忽视了IoU损失本身存在的局限性,如对不同检测器与检测任务的自适应能力不足以及收敛速度较慢等问题。对此,ZHANG等<sup>[18]</sup>提出了一种基于辅助框的Inner-IoU方法来计算边界框损失。该方法通过引入尺度因子,动态生成不同尺度的辅助框参与损失计算,从而扩展回归的有效范围。此外,该方法可以根据不同的检测器和检测任务进行自适应调整,具有更强的泛化能力。

由于在自然场景中,梨树叶片常出现重合交叠的情况,为了在不同重合程度的叶片图像中准确识别病害叶片,模型需要具备更强的泛化能力。Inner-IoU损失计算方法不仅能够根据不同检测器和检测任务进行自适应调整以提高模型的泛化能力,而且能够较好地与基于IoU的损失计算方法(如CIoU、EIoU、SIOU)相结合<sup>[18]</sup>。在现有的基于IoU的损失计算方法中,SIOU具有更加出色的性能,因此,本文选择结合了Inner-IoU与SIOU的优点的Inner-SIOU作为损失函数。

## 2.3 模型训练与评价指标

### 2.3.1 实验环境与参数设置

本实验使用环境主要参数为:Ubuntu 20.04.6 LTS操作系统,Intel Xeon Gold 6130 CPU,显存为11 GB的NVIDIA GeForce RTX 2080ti GPU,CUDA版本为10.2,编程语言为Python 3.8.18,深度学习框架为Pytorch 1.9.1。训练时,将输入图像分辨率设置为640像素×640像素,模型训练的最大迭代次数设置为150,初始学习速率为0.01,批次大小为64。

### 2.3.2 评价指标

实验以平均精准度(mean average precision, mAP)作为评估模型目标检测准确率的标准。mAP取值范围为[0,1],其值越大说明模型检测精度越高。mAP通常在IoU阈值为50%和50%~95%的范围内进行计算,分别记作mAP@50和mAP@50-95。模型大小采用参数量进行评估,计算量则以10亿次浮点运算(GFLOPs)为依据,参数量和计算量越小,表示模型对硬件的算力要求越低,越容易在小型设备上部署。

## 3 实验结果与分析

### 3.1 改进方法效果对比

#### 3.1.1 RepGhostNet模块对模型性能的影响

为了验证不同主干网络改进方案对模型性能的影响,设计采用不同模块改进主干网络的对比实验,结果如表1所示。MobileNetv3<sup>[19]</sup>通过倒置残差和线性瓶颈提高特征学习并降低计算成本,同时使用神经网络结构搜索策略自动发现最优网络结构。在使用MobileNetv3模块改进主干网络的方案中,虽然相较于基线模型减少了24.69%的计算量,但mAP@50却下降了2.3%,其改进效果并不理想。EfficientViT<sup>[20]</sup>是视觉Transformer的高效变体,该模块通过采用基于多头自注意力机制的视觉Transformer,并引入深度可分离卷积以增强线性注意力。然而,在使用EfficientViT改进主干网络时,模型的精度下降,同时参数量和计算量均增加,对于模型的优化没有起到积极的作用。这是因为EfficientViT为了保持全局和局部特征提取能力,使用线性注意力代替softmax注意力,而不是限制softmax注意力,从而增加了参数量和计算量。FasterNet<sup>[21]</sup>使用了一种新的部分卷积(partial convolution, PConv)替代普通卷积,减少了冗余计算,提高了访存效率,同时增强了通道通信。虽然使用FasterNet改进主干网络后,mAP@50-95相较于基线模型提高了1.08%,但是参数量和计算量却分别增加了36.67%和32.10%,这对于模型的优化来说代价过高。本文使用RepGhostNet改进主干网络,改进后的模型相较于基线模型mAP@50和mAP@50-95分别提高了0.82%和0.67%,同时参数量和计算量分别减少了23.33%和16.05%,在降低参数量和计算量的同时保证了精度。

表1 使用不同主干网络改进的模型性能对比结果

Table 1 Comparison of improved model performance using different backbone networks

| 主干网络         | mAP@50 | mAP@50-95 | 参数量/(×10 <sup>6</sup> 个) | GFLOPs |
|--------------|--------|-----------|--------------------------|--------|
| 基线模型         | 0.853  | 0.742     | 3.0                      | 8.1    |
| MobileNetv3  | 0.833  | 0.705     | 3.1                      | 6.1    |
| EfficientViT | 0.842  | 0.712     | 4.0                      | 9.4    |
| FasterNet    | 0.826  | 0.750     | 4.1                      | 10.7   |
| RepGhostNet  | 0.860  | 0.747     | 2.3                      | 6.8    |

#### 3.1.2 不同注意力机制对模型性能的影响

为验证BRA的效果,设计引入不同注意力机

制模块的对比实验,结果如表2所示。实验结果表明,虽然使用SpatialGroup-Enhance<sup>[22]</sup>、EMA<sup>[23]</sup>和SimAM<sup>[24]</sup>这3种注意力机制都没有改变模型的参数量和计算量,但模型的精度却下降了,这与引入注意力机制增强特征融合能力以提高模型精度的初衷相悖。本文在网络中引入BRA模块,尽管改进后的模型相较于基线模型参数量和计算量分别增加了6.67%和2.47%,但mAP@50和mAP@50-95分别提高了0.35%和0.54%。这是由于引入注意力机制后,需要额外的参数来学习权重和注意力分布,以调整模型对输入的关注程度,使模型更有效地捕捉输入中的相关信息,从而增强网络的表征能力和特征融合能力。然而,BRA在通过查询自适应的方式降低对不相关特征的关注时,充分利用稀疏性操作直接跳过最不相关区域的计算,这些操作涉及密集矩阵乘法,有利于服务器端的推理加速,能够有效降低计算量。因此,BRA以尽可能小的参数量和计算量实现了提升精度的目标,确保在实际应用中达到最佳性能和效率平衡。

表2 使用不同注意力机制改进的模型性能对比结果

| Table 2 Comparison of improved model performance using different attention mechanism |        |           |                            |        |
|--------------------------------------------------------------------------------------|--------|-----------|----------------------------|--------|
| 注意力机制                                                                                | mAP@50 | mAP@50-95 | 参数量/<br>( $\times 10^6$ 个) | GFLOPs |
| 基线模型                                                                                 | 0.853  | 0.742     | 3.0                        | 8.1    |
| SpatialGroup-Enhance                                                                 | 0.847  | 0.709     | 3.0                        | 8.1    |
| EMA                                                                                  | 0.842  | 0.706     | 3.0                        | 8.1    |
| SimAM                                                                                | 0.812  | 0.708     | 3.0                        | 8.1    |
| BRA                                                                                  | 0.856  | 0.746     | 3.2                        | 8.3    |

3.1.3 不同损失函数对模型性能的影响

为验证Inner-SIoU损失函数的优越性,将其与Inner-CIoU和Inner-EIoU进行比较,实验结果如表3所示。结果表明,采用3种改进的损失函数均未减少模型的参数量和计算量。使用Inner-CIoU作为损失函数反而导致模型精度下降,而使用Inner-EIoU作为损失函数时,相较于基线模型仅有mAP@50-95提高了0.4%。相比之下,使用Inner-SIoU作为损失函数时,其mAP@50和mAP@50-95相较于基线模型分别提高了0.58%和2.02%。为了更直观地体现出不同损失函数对于模型收敛性能的影响,将训练过程中的损失值变化进行可视化,结果如图6所示。由图6可以看出,Inner-SIoU相对于Inner-CIoU和Inner-EIoU表现出更快的收敛速度,并且具有更

小的边框损失值,这反映出Inner-SIoU损失函数更具有优势。

表3 使用不同损失函数的模型性能对比结果

Table 3 Comparison of improved model performance using different loss function

| 损失函数       | mAP@50 | mAP@50-95 | 参数量/( $\times 10^6$ 个) | GFLOPs |
|------------|--------|-----------|------------------------|--------|
| 基线模型       | 0.853  | 0.742     | 3.0                    | 8.1    |
| Inner-CIoU | 0.837  | 0.713     | 3.0                    | 8.1    |
| Inner-EIoU | 0.853  | 0.745     | 3.0                    | 8.1    |
| Inner-SIoU | 0.858  | 0.757     | 3.0                    | 8.1    |

注: Inner-CIoU、Inner-EIoU 和Inner-SIoU都是在尺度因子为0.7时进行计算。

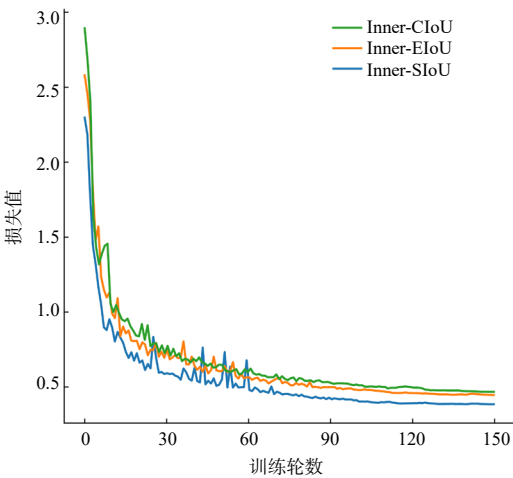


图6 不同损失函数曲线对比

Fig.6 Comparison of different loss function curves

3.2 消融实验

为了进一步探究本文提出的各部分改进方法对模型整体性能的提升效果,本实验在DiaMOS Plant数据集上进行消融实验,实验结果如表4所示。通过表4可以看出,单独采用一种改进方法均可提高模型的精度,这充分证明了这些改进方法的有效性。然而,单独引入这些模块或改进方法对模型的整体优化效果并不显著,例如,仅采用RepGhostNet改进主干网络时,虽然降低了参数量和计算量,但模型的精度提升并不明显。引入BRA虽然提高了精度,却不可避免地增加了参数量和计算量。此外,无论是单独采用Inner-SIoU损失函数改进策略还是将其与其他两种改进方法结合使用,都进一步提高了精度,但无法有效降低模型的参数量和计算量。为此,本文使用RepGhostNet改进主干网络,在此基础上,再引入BRA,使模型更有效地提取特征。同时,使用Inner-SIoU损失

函数替代原有损失函数，进一步提高模型的精度。通过这种组合策略，改进后的模型对关键特征的关注程度更加集中(图7)，与改进前相比，mAP@50和mAP@50-95分别提高了5.6%和5.1%，

同时参数量和计算量分别减少了20.00%和13.58%，在提高精度的同时兼顾了轻量化的部署需求。

表4 消融实验结果

| Table 4 Results of ablation experiments |             |     |            |        |           |                        |        |
|-----------------------------------------|-------------|-----|------------|--------|-----------|------------------------|--------|
| 模型                                      | RepGhostNet | BRA | Inner-SIoU | mAP@50 | mAP@50-95 | 参数量/( $\times 10^6$ 个) | GFLOPs |
| 基线模型                                    |             |     |            | 0.853  | 0.742     | 3.0                    | 8.1    |
| 改进模型1                                   | √           |     |            | 0.860  | 0.747     | 2.3                    | 6.8    |
| 改进模型2                                   |             | √   |            | 0.856  | 0.746     | 3.2                    | 8.3    |
| 改进模型3                                   |             |     | √          | 0.858  | 0.757     | 3.0                    | 8.1    |
| 改进模型4                                   |             | √   | √          | 0.864  | 0.751     | 3.2                    | 8.3    |
| 改进模型5                                   | √           |     | √          | 0.883  | 0.758     | 3.0                    | 8.1    |
| 本文改进模型                                  | √           | √   | √          | 0.901  | 0.780     | 2.4                    | 7.0    |

注：“√”表示模型中包含该改进方法。

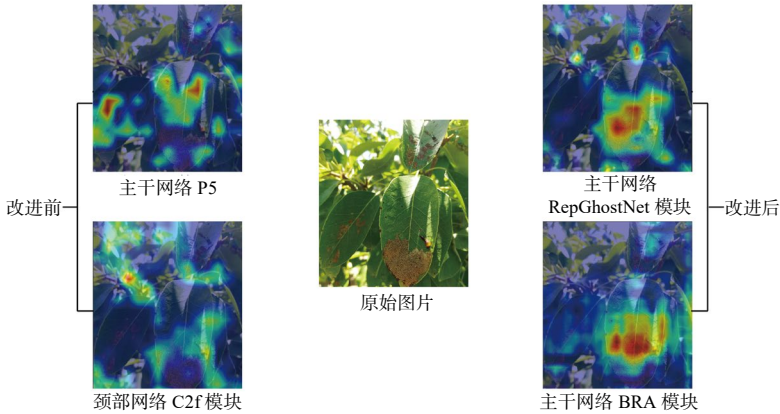


图 7 模型改进前后特征提取效果热力图

Fig.7 Heatmap of feature extraction performance before and after model improvement

3.3 不同模型之间的性能比较

为了进一步验证本文提出的基于YOLOv8n的改进模型的性能，在相同的环境配置和数据集的前提下将本文提出的改进模型与当前主流的单阶段模型SSD<sup>[25]</sup>、两阶段模型Faster R-CNN<sup>[26]</sup>以及其他版本的YOLO模型进行对比实验，实验结果如表5所示。相较于SSD和Faster R-CNN，本文提出的改进模型在mAP@50指标上分别提高了12.9%

和8.4%，且参数量分别减少了90.87%和94.00%，计算量分别减少了92.02%和96.62%，充分体现了本文改进模型的优势。此外，与同系列的YOLOv5n和YOLOv8s相比，本文提出的改进模型不仅在精度上进一步提升，而且参数量和计算量均有所减少，在高精度和模型轻量化之间取得了更好的平衡。

4 结论

本文提出了一种基于YOLOv8n的梨树叶片病害检测改进模型，旨在解决传统目标检测模型精度低、参数量大的问题。本文作者从模型轻量化的角度出发，使用RepGhostNet改进主干网络，在提升网络特征提取能力的同时减少了参数量和计算量；此外，引入双层路由注意力机制增强网络的表征能力和特征融合能力，降低复杂背景的干扰，并且使用Inner-SIoU损失函数优化边界框回归，加快模型收敛速度，提高识别精度。结果表

表5 不同模型对比结果

| Table 5 Comparison results of different model |        |           |                        |        |
|-----------------------------------------------|--------|-----------|------------------------|--------|
| 模型                                            | mAP@50 | mAP@50-95 | 参数量/( $\times 10^6$ 个) | GFLOPs |
| SSD                                           | 0.798  | —         | 26.3                   | 87.7   |
| Faster R-CNN                                  | 0.831  | —         | 40.0                   | 207.0  |
| YOLOv5n                                       | 0.822  | 0.689     | 2.5                    | 7.1    |
| YOLOv8s                                       | 0.856  | 0.760     | 11.1                   | 28.4   |
| 本文改进模型                                        | 0.901  | 0.780     | 2.4                    | 7.0    |

注：SSD 和Faster R-CNN这2个模型的常用评估方式通常只关注mAP@50，故其mAP@50-95未进行测算。



明,改进后的模型能够对梨树叶片病害进行有效检测,在DiaMOS Plant数据集上对梨树叶片病害的检测mAP@50达到0.901,相较于原模型提高了5.6%,并且改进的模型具有更少的参数量和计算量,兼顾了高精度和低算力的部署需求。

#### 参考文献:

- [1] 张放. 2020年我国水果生产统计简析[J]. 中国果业信息, 2021, 38(12): 29–39.
- [2] FENU G, MALLOCI F M. Using multioutput learning to diagnose plant disease and stress severity[J]. Complexity, 2021, 2021(1): 6663442.
- [3] HUGHES D P, SALATHÉ M. An open access repository of images on plant health to enable the development of mobile disease diagnostics through machine learning and crowdsourcing[EB/OL]. (2015–11–25). <https://arxiv.org/abs/1511.08060v1>.
- [4] 朱帅, 王金聪, 任洪娥, 等. 基于多特征融合的残差网络果树叶片病害识别[J]. 森林工程, 2022, 38(1): 108–114, 123.
- [5] 孙俊, 谭文军, 毛罕平, 等. 基于改进卷积神经网络的多种植物叶片病害识别[J]. 农业工程学报, 2017, 33(19): 209–215.
- [6] 方文博, 郭永刚, 关法春, 等. 基于改进YOLO v5s算法的大豆叶片虫洞的识别[J]. 湖南农业大学学报(自然科学版), 2023, 49(1): 127–132.
- [7] 盛帅, 段先华, 胡维康, 等. Dynamic-YOLOX: 复杂背景下的苹果叶片病害检测模型[J]. 计算机科学与探索, 2024, 18(8): 2118–2129.
- [8] 李鑫然, 李书琴, 刘斌. 基于改进Faster R-CNN的苹果叶片病害检测模型[J]. 计算机工程, 2021, 47(11): 298–304.
- [9] CHEN Y X, WANG X Y, CHEN Z B, et al. Classification of *Camellia oleifera* diseases in complex environments by attention and multi-dimensional feature fusion neural network[J]. Plants, 2023, 12(14): 2701.
- [10] LI X P, CHEN X Y, YANG J L, et al. Transformer helps identify kiwifruit diseases in complex natural environments[J]. Computers and Electronics in Agriculture, 2022, 200: 107258.
- [11] CHEN C P, GUO Z C, ZENG H E, et al. RepGhost: a hardware-efficient ghost module via re-parameterization[EB/OL]. (2022–11–11). <https://arxiv.org/abs/2211.06088v1>.
- [12] ZHU L, WANG X, KE Z, et al. BiFormer: Vision transformer with bi-level routing attention[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Vancouver, Canada: IEEE, 2023: 10323–10333.
- [13] FENU G, MALLOCI F M. DiaMOS Plant: a dataset for diagnosis and monitoring plant disease[J]. Agromony, 2021, 11(11): 2107.
- [14] 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. Seattle, USA: IEEE/CVF, 2020: 1580–1589.
- [15] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: faster and better learning for bounding box regression[EB/OL]. (2019–11–19). <https://arxiv.org/abs/1911.08287>.
- [16] 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.
- [17] GEVORGYAN Z. SIOU loss: more powerful learning for bounding box regression[EB/OL]. (2022–05–25). <http://arxiv.org/abs/2205.12740>.
- [18] ZHANG H, XU C, ZHANG S. Inner-IoU: more effective intersection over union loss with auxiliary bounding box[EB/OL]. (2023–11–14). <https://arxiv.org/abs/2311.02877v4>.
- [19] HOWARD A, SANDLER M, CHEN B, et al. Searching for MobileNetV3[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea: IEEE, 2019: 1314–1324.
- [20] LIU X, PENG H, ZHENG N, et al. EfficientViT: memory efficient vision transformer with cascaded group attention[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Vancouver, Canada: IEEE, 2023: 14420–14430.
- [21] CHEN J, KAO S, HE H, et al. Run, don't walk: chasing higher FLOPS for faster neural networks[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Vancouver, Canada: IEEE, 2023: 12021–12031.
- [22] LI X, HU X L, YANG J. Spatial group-wise enhance: improving semantic feature learning in convolutional networks[EB/OL]. (2019–05–25). <https://arxiv.org/abs/1905.09646v2>.
- [23] OUYANG D, HE S, ZHANG G, et al. Efficient multi-scale attention module with cross-spatial learning[C]//2023 IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP). Rhodes Island, Greece: IEEE, 2023: 1–5.
- [24] YANG L, ZHANG R Y, Li L, et al. SimAM: a simple, parameter-free attention module for convolutional neural networks[C]//Proceedings of the 38th International Conference on Machine Learning. Virtual: PMLR, 2021: 11863–11874.
- [25] 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.
- [26] REN S, HE K, 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.

责任编辑: 伍锦花

英文编辑: 张承平