

基于改进 YOLOv8 模型的轻量化板栗果实识别方法

李 茂，肖洋轶，宗望远^{*}，宋 宝

(1. 华中农业大学工学院，武汉 430070; 2. 农业农村部长江中下游农业装备重点实验室，武汉 430070)

摘 要：为实现自然环境下的板栗果实目标快速识别，该研究以湖北省种植板栗为研究对象，提出了一种基于改进 YOLOv8 模型的栗果识别方法 YOLOv8-PBi。首先，将部分卷积（partial convolution, PConv）引入 C2f 模块中，缩减卷积过程中的浮点数和计算量；其次，引入加权双向特征金字塔网络（bidirectional feature pyramid network, BiFPN），增强多尺度特征融合性能，最后，更改边界框损失函数为动态非单调聚焦机制 WIoU（wise intersection over union, WIoU），提高模型收敛速度，进一步提升模型检测性能。试验结果表明，改进 YOLOv8-PBi 模型准确率、召回率和平均精度分别为 89.4%、74.9%、84.2%；相比原始基础网络 YOLOv8s，模型权重减小 46.22%，准确率、召回率和平均精度分别提升 1.3、1.5、1.8 个百分点。部署模型至边缘嵌入式设备上，经过 TensorRT 加速后，检测帧率达到 43 帧/s。该方法可为板栗智能化收获过程中的栗果识别提供技术基础。

关键词：图像处理；图像识别；YOLOv8；板栗果实；目标检测；部分卷积；BiFPN

doi: 10.11975/j.issn.1002-6819.202309185

中图分类号: TP391.4; S24

文献标志码: A

文章编号: 1002-6819(2024)-01-0209-09

李茂, 肖洋轶, 宗望远, 等. 基于改进 YOLOv8 模型的轻量化板栗果实识别方法[J]. 农业工程学报, 2024, 40(1): 209-217. doi: 10.11975/j.issn.1002-6819.202309185 <http://www.tcsae.org>

LI Mao, XIAO Yangyi, ZONG Wangyuan, et al. Detecting chestnuts using improved lightweight YOLOv8[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(1): 209-217. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.202309185 <http://www.tcsae.org>

0 引 言

板栗作为中国丘陵山区种植的主要经济作物，种植面积和产量均居世界第一^[1]。当前板栗采收环节存在的人工投入大、危险性高、生产效率低等问题已经成为制约板栗产业发展的关键因素^[2]。随着信息技术的发展，传统农业朝着智慧化发展，这为板栗产业升级提供了方向，而板栗果实目标检测可为板栗收获过程提供必要的技术支持和智能化解决方案。目前针对板栗的检测主要有近红外技术和传统机器视觉。展慧等^[3]基于 BP 神经网络建立板栗识别模型，预测识别率达到 91.67%；周竹等^[4]基于遗传算法（genetic algorithms, GA）和最小二乘支持向量机（least squares support vector machine, LSSVM），对标准正态变量变化预处理后的板栗近红外光谱进行傅里叶变换，建立 GA-LSSVM 板栗识别模型，总体平均识别正确率为 97.54%；MOSCETTI 等^[5]基于近红外光谱，使用遗传算法进行特征选择，结合图像分析分级、线性判别分析（linear discriminant analysis, LDA）、二次判别分析（quadratic discriminant analysis, QDA）等

分析法，实现板栗识别误差 2.38%。近年来，板栗目标检测主要集中在板栗分级和分选环节中，自然条件下的板栗果实由于环境复杂、光线变化多样、树叶与栗果辨识度较低、遮挡严重等情况，识别难度较高。

随着物联网、大数据、AI 等的发展，基于机器视觉技术和深度学习的非接触目标检测为农业作物智能化收获提供了基础^[6]。王金星等^[7]基于 Mask R-CNN 网络改进 ResNeXt 为主干网络，引入 Boundary 损失函数构建苹果害虫检测模型，经过试验证明，改进 Mask R-CNN 平均识别精度达到 96.52%；闫建伟等^[8]基于 Faster RCNN 网络，采用交替优化训练方式（alternating optimization）将感兴趣区域池化（ROI pooling）改进为感兴趣区域校准（ROI align），并选择 VGG16 网络模型构建刺梨果实识别方法，经检测，改进 Faster RCNN 平均识别精度为 92.01%；LE 等^[9]基于 Faster RCNN 模型，采用 Inception-ResNet-V2 特征提取模块构建复杂环境下的杂草检测模型，经检测表明，改进 Faster RCNN 的平均精度达 55.5%；HTET 等^[10]基于 Mask RCNN，通过贝叶斯算法调整超参数组合，构建无人机视频中的棕榈树检测，经检测，该模型在各棕榈类别上均获得 80%~95% 的准确率。

以上均使用类似于 RCNN^[11] 的 Two-stage 算法，检测精度较高，但由于检测过程分为两个阶段，速度较慢。另外一种以 SSD^[12]、YOLO^[13] 等为代表性的 One-stage 网络，由于一阶段检测的特性，检测速度较快。ARAKAWA 等^[14]基于 YOLOv4 目标检测网络，针对无

收稿日期: 2023-09-25 修订日期: 2023-12-26

基金项目: 湖北省重点研发计划项目 (2020BED027); 湖北省自然科学基金 (2023AFB871)

作者简介: 李茂, 博士生, 研究方向为现代农业电气化与自动化。

Email: leemone@webmail.hzau.edu.cn

*通信作者: 宗望远, 博士, 教授, 博士生导师, 研究方向为现代农业装备设计与测控。Email: zwy@mail.hzau.edu.cn

人机采集的图像训练, 训练模型 R^2 0.98, 均方根 6.3, 平均检测精度为 84.27%; 王梦妮等^[15] 基于 YOLOv5s 网络, 替换空间池化结构 (spatial pyramid pooling-fast, SPPF) 为空洞卷积池化结构 (atrous spatial pyramid pooling, ASPP), 采用双向特征金字塔网络 (bidirectional feature pyramid network, BiFPN), 添加卷积注意力模块 (convolutional block attention module, CBAM), 构建 Tea-YOLOv5s 茶叶嫩芽识别方法, 经检测, 识别准确率达 85%, 推理速度 84 帧/s; 黄杰等^[16] 基于 YOLOv4 网络, 替换特征提取网络为 GhostNetV2, 替换特征融合网络中的卷积模块为深度可分离卷积 (depthwise separable convolution, DW), 采用有角度代价的边界框损失函数 (SCYLLA-IoU, SIoU), 构建轻量化马铃薯种薯芽眼检测算法, 经检测, 模型平均检测精度为 89.13%, 检测时间为 0.148 s; BAZAME 等^[17] 为了获取阿拉比卡咖啡成熟度, 对比 YOLOv3、YOLOv3-tiny、YOLOv4、YOLOv4-tiny 三种检测模型, 检测精度分别为 78%、77%、81%、79%。

以 YOLO 为代表的目标检测模型在检测精度和速度上效果较好, 其在农业领域目标检测的成功应用为板栗果实目标检测提供了借鉴, 同时可以修改网络结构进一步减小模型参数量, 提高目标检测速度, 部署到算力有限的边缘设备上。本文基于最新的 YOLOv8 模型, 以湖北省各地适收期板栗为研究对象, 对自然条件下的板栗果实目标检测模型开展轻量化工作, 以实现减小参数量, 节约设备资源, 满足机械化收获过程中的板栗目标检测要求。

1 材料与方法

1.1 数据集建立

试验用板栗果实图像于 2022 年 9 月 2 日至 8 日 (8:00 至 18:00) 在湖北省板栗主产区黄冈市罗田县和红安县、孝感市安陆市、荆门市京山市等地采集。2022 年 9 月 23 日, 又在华中农业大学板栗种植园采集了部分数据。采集设备为 Canon 200D 和 Fuji X-T20 相机。由于自然环境背景的复杂性和板栗果实目标特征的不明显性, 采集过程中选择了不同的光线条件、拍摄角度、拍摄距离等, 进一步增加目标数据多样性和模型泛化能力, 图 1 展示了部分采集的板栗图像。

采用 Make Sense 软件标注自然环境下板栗果实目标, 得到包含板栗目标中心点坐标和标注框宽度、高度信息的 TXT 标签文件, 并将采集板栗目标数据集分为训练集和验证集, 数据集划分如表 1。

YOLOv8 的数据增强方法包括马赛克 (Mosaic)、自适应锚框计算和自适应目标缩放。马赛克数据增强是将四张目标图片按照随机的比例和位置拼接成一张新的图片, 如图 2 所示。这种方法可以有效地扩充数据集, 提高网络训练效率, 减少模型内存消耗。

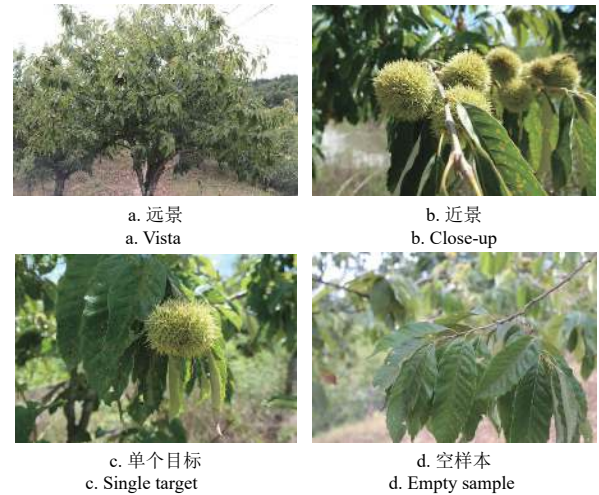


图 1 板栗果实采集图像示例

Fig.1 Chestnut fruit collection image example

表 1 板栗数据集分布表

Table 1 Chestnut dataset distribution table

类别 Categories	图像数量 Number of images	标签数量 Number of labels
训练 Training	2 624	28 089
验证 Validation	1 164	19 005

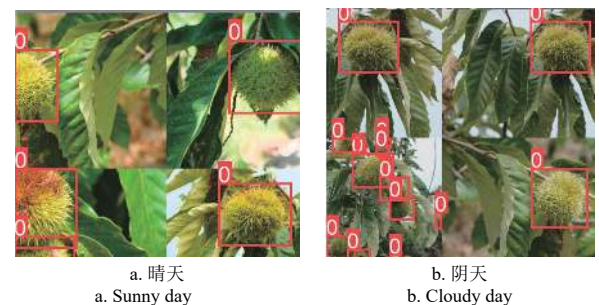


图 2 Mosaic 数据增强效果

Fig.2 Mosaic data enhancement effect

1.2 板栗果实目标检测方法

1.2.1 YOLOv8 卷积神经网络模型

YOLOv8 是 Ultralytics 公司开发的一种 SOTA (state-of-the-art) 模型, 于 2023 年 1 月正式发布, 继承了 YOLO 系列的优点, 同时加入了新的特性和改进。与上一代 YOLOv5 相比, 主要的变化有: 在主干网络中使用 C2f 模块代替 C3 模块; 去掉了上采样过程中的卷积操作; 采用解耦合头结构 (DeCouple-Head), 将分类和检测任务分开, 进一步降低了模型复杂度。

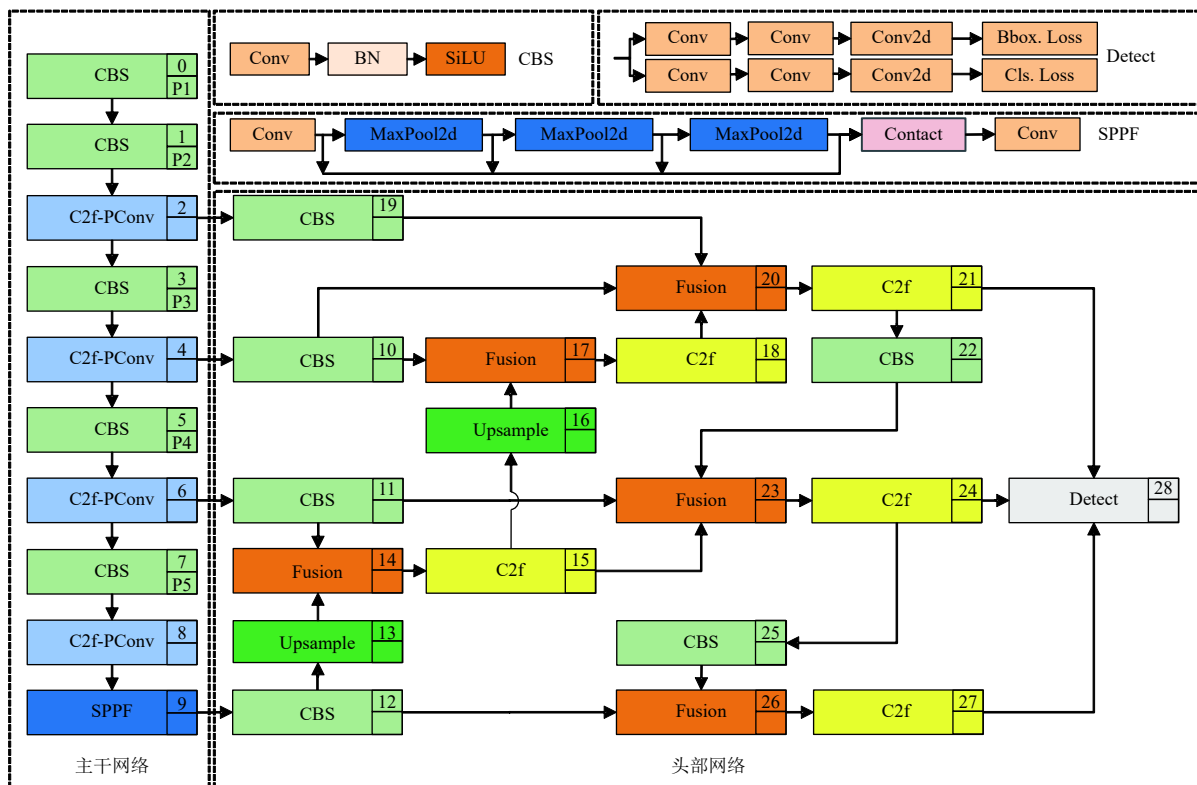
YOLOv8 网络结构由 3 个部分组成: 输入端 (Input)、主干网络 (Backbone) 和头部网络 (Head)。输入端将经过 Mosaic 数据增强的板栗果实图像送入网络, 主干网络使用 DarkNet53 结构自下而上提取特征, 并使用 C2f 模块替换 C3 模块。C2f 模块包含 n 个 Bottleneck 和 3 个卷积层, 以及一个 Split 和多个跳层连接。SPPF 模块通过不同核大小的池化抽取多尺度特征, 并进行叠加融合。

头部网络主要通过路径聚合网络（path aggregation network, PANet）和预测头（Detect）实现板栗的特征融合和目标检测，PANet 由 FPN（feature pyramid networks）和 PAN（bottom-up path aggregation）组成，FPN 自上而下提取特征层信息，将上层特征和下层特征融合，提升网络对于不同尺度目标的能力。PAN 自下而上提取特征层信息，获取位置信息。预测头以 3 个不同尺寸的分支

输出分类结果和目标坐标信息。

1.2.2 网络模型的改进

本文改进后的 YOLOv8-PBi 结构如图 3 所示，通过引入 PConv 到 C2f 模块，替换 BiFPN 为特征融合网络及替换损失函数为 WIoU，有效提升模型检测能力和速度，减少漏检和错检。



Note: CBS is convolution + normalization + activation function; SPPF is the spatial pyramid pooling module; C2f-PConv is the C2f module with partial convolution; Fusion is the fast normalization fusion module; Upsample is the upsampling module; Detect is the detection head; Maxpool2 d is the maximum pooling; Conv is the convolution; Contact is the feature connection module; Bbox. Loss and Cls. Loss are bounding box losses and classification losses, respectively.

图 3 YOLOv8-PBi 网络结构图

Fig.3 YOLOv8-PBi network structure diagram

本文主要从以下 3 个方面进行改进:

1) C2f-PConv 模块

YOLOv8 采用的 C2f 模块含有更多的 Bottleneck 结构, 在提取更多特征的同时导致了通道信息的过度冗余。现阶段 MobileNet^[18]、ShuffleNet^[19]、GhostNet^[20] 等主流轻量化网络利用深度卷积或组卷积来提取空间特征, 在减少 FLOPs 的同时, 算子造成内存访问增加, 碎片计算效率低下。FasterNet 中的 PConv 利用特征图中的冗余信息, 只在部分通道上执行规则卷积, 而不影响其他通道, 这样可以更好地利用设备的计算能力, 减少计算冗余和内存访问^[21]。PConv 只需要在输入的部分通道上进行卷积操作, 从而提取空间特征, 而保持其他通道不变。假设 h 、 w 表示输入特征图的长和宽, c 表示输入通道数, c_p 表示参与卷积的通道数, k 为卷积核尺寸, r 表示参与卷积率, 其计算量 F_{PConv} 表达式如式 (1), 内存访问量 M_{AC} 计算式如式 (3)。

$$F_{\text{PConv}} = h \cdot w \cdot k^2 \cdot c_p^2 \quad (1)$$

$$r = \frac{c_p}{c} \quad (2)$$

$$M_{AC} = h \cdot w \cdot 2c_p + k^2 \cdot c_p^2 \approx h \cdot w \cdot 2c_p \quad (3)$$

一般典型参与卷积率为 1/4, PConv 的 FLOPs 仅为常规卷积的 1/16, 由于卷积过程占用内存较少, 内存访问量约为常规卷积 1/4。

由于只有 c_p 通道参与空间特征提取，对于后续特征通道信息并未有流失，同时可以大大减少计算量和内存访问量。

串联的 Bottleneck 模块可以实现不同尺度特征的提取及融合, 提高特征图的表示能力, 但由于 YOLOv8 中含有大量 Bottleneck 结构, 带来更大的计算开销。Bottleneck 结构如图 4。

本文使用 FasterNet 中的 PConv 设计 Faster Block,

在特征提取 C2f 模块处, 用类继承的方式将 Faster Block 替换 Bottleneck 作为主要的梯度流分支, 缩减模型浮点数和计算量。通过构建 C2f-PConv 引入 PConv, 进一步缩减卷积过程中的浮点数和计算量。C2f-PConv 模块结构如图 5 所示。

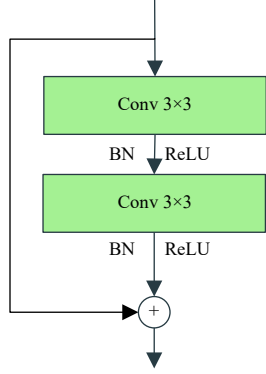
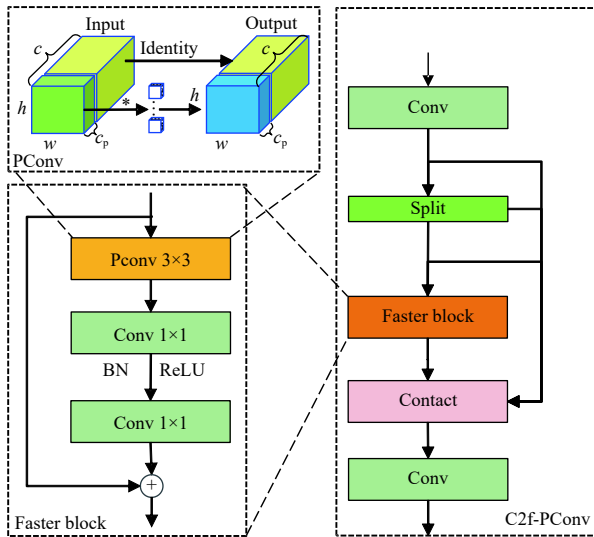


图 4 Bottleneck 结构图
Fig.4 Bottleneck structure diagram

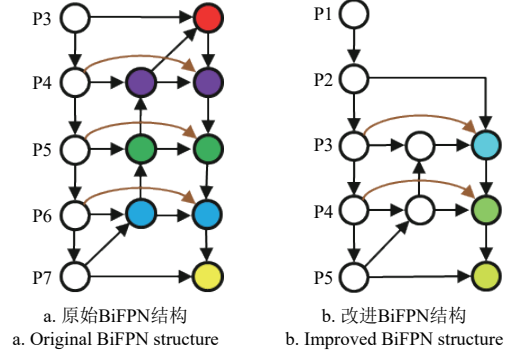


注: c 为输入特征图的通道数; c_p 为参与卷积的通道数; $*$ 为卷积操作; h, w 分别为特征图的高和宽尺寸; Conv 为常规卷积模块; PConv 为部分卷积; Faster Block 为含有部分卷积的结构; Split 为通道分割模块。
Note: c is the number of channels in the input feature map; c_p is the number of channels participating in the convolution; $*$ is the convolution operation; h, w are the height and width dimensions of the feature map; Conv is the general convolution module; PConv is the partial convolution; Faster Block is the structure with partial convolution; Split is the channel splitting module.

图 5 C2f-PConv 模块结构图
Fig.5 C2f-PConv module structure diagram

2) 特征融合网络 BiFPN

多尺度特征融合旨在聚合不同分辨率的特征, PANet 有自下而上和自上而下两条路径, 效果较 FPN 更好, 但计算量更大, BiFPN 基于 PANet, 移除仅有一个输入的节点, 此种节点由于未进行特征融合, 对特征融合网络的贡献较小, 同时为融合更多的特征, 额外添加一条路径连接同级网络输入节点与输出节点^[22], 原始 BiFPN 结构如图 6a 所示, 本文改进 BiFPN 结构增加 P2 大尺寸特征图层, 增加板栗果实远景目标特征融合能力, 改进 BiFPN 结构如图 6b 所示。



注: P1~P7 为不同层级的输入特征图。
Note: P1~P7 are the input feature maps of different layers.

图 6 BiFPN 结构
Fig.6 BiFPN structure

特征融合过程中, 不同输入特征分辨率不同, 对于输出特征的贡献不等, 自然条件下采集的多尺度栗果目标较多, BiFPN 采用快速归一化融合模块平衡不同特征权重, 挖掘栗果目标的深层信息, 减少因环境复杂造成的误检和漏检, 其输入与输出的关系式为:

$$O = \sum_i \frac{\omega_i}{\epsilon + \sum_j \omega_j} \cdot I_i \quad (4)$$

式中 ω_i 为输入特征 I_i 对应的学习权重, 通过后级应用 ReLU 来确保 $\omega_i \geq 0$ 。为避免数值不稳定, 设置初始学习率 $\epsilon=0.0001$, 归一化权重的值在 0~1 之间下降。

以特征层 P_4 为例, P_4 的中间特征 P_4^{td} 和输出特征 P_4^{out} 式分别为 (5)、(6):

$$P_4^{td} = \text{Conv} \left(\frac{\omega_1 \cdot P_4^{in} + \omega_2 \cdot \text{Resize}(P_5^{in})}{\omega_1 + \omega_2 + \epsilon} \right) \quad (5)$$

$$P_4^{out} = \text{Conv} \left(\frac{\omega_1' \cdot P_4^{in} + \omega_2' \cdot P_4^{td} + \omega_3' \cdot \text{Resize}(P_3^{out})}{\omega_1' + \omega_2' + \omega_3' + \epsilon} \right) \quad (6)$$

式中 P_i^{in} 为第 i 级输入特征, Resize 为上采样或下采样操作。

同时为了进一步提高效率, 特征融合阶段使用深度可分离卷积, 卷积后添加批量归一化和激活操作。本文特征融合网络替换为 BiFPN, 能够为不同尺度特征分配不同权重, 并进行跨尺度权重抑制或特征表达加强特征融合, 进一步提高目标检测性能。

3) WIoU 损失函数

YOLOv8 网络预测边框坐标损失 L_{CIoU} 采用的是 CIoU (complete intersection over union) Loss^[23], 其计算如式 (7) 所示:

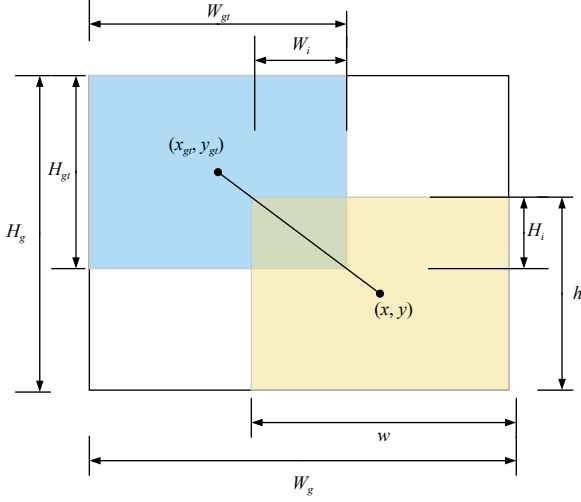
$$L_{CIoU} = L_{IoU} + \frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)} + \alpha v \quad (7)$$

$$\alpha = \frac{v}{L_{IoU} + v} \quad (8)$$

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w}{h} - \arctan \frac{w_{gt}}{h_{gt}} \right)^2 \quad (9)$$

$$L_{IoU} = \frac{W_i H_i}{wh + w_{gt} h_{gt} - W_i H_i} \quad (10)$$

式中 L_{IoU} 表达如图 7 所示, α 为权重函数, 用于平衡参数, v 为纵横比度量函数, 用于衡量高宽比一致性, CIoU 损失函数虽然考虑了重叠面积、中心距离、纵横比, 但在预测框回归过程中, 存在预测框与真实框高宽比呈线性关系时, CIoU 的惩罚项退化为 0 的情况, 同时无论是高质量或者低质量锚框都会对回归损失有害^[24]。



注: $w, h, (x, y)$ 分别表示预测框的宽、高尺寸和中心坐标; $w_{gt}, h_{gt}, (x_{gt}, y_{gt})$ 分别表示真实框的宽、高尺寸和中心坐标; w_i, h_i 分别表示交集宽、高尺寸; W_g, H_g 分别表示最小边框宽、高尺寸。
Note: $w, h, (x, y)$ represent the width and height dimensions and center coordinates of the prediction box, respectively; $w_{gt}, h_{gt}, (x_{gt}, y_{gt})$ represent the width and height dimensions and center coordinates of the real box, respectively; w_i, h_i represent the intersection width and height dimensions, respectively; W_g, H_g represent the minimum border width and height dimensions respectively.

图 7 真实框和预测框交并集面积

Fig.7 The area of intersection and union of the real box and the prediction box

本文选择动态非单调聚焦机制 WIoU 替换原损失函数, 其损失函数如式 (11) 所示:

$$L_{WIoU} = r R_{WIoU} L_{IoU}, R_{WIoU} \in [1, e], L_{IoU} \in [0, 1] \quad (11)$$

式中距离聚焦机制 R_{WIoU} 用于放大普通适量锚框的 L_{IoU} , 非单调聚焦系数 r 用于聚焦普通质量的锚框。 R_{WIoU} 定义如式 (12), r 定义如式 (13)。

$$R_{WIoU} = \exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)^*}\right) \quad (12)$$

$$r = \frac{\beta}{\delta \alpha^{\beta - \delta}} \quad (13)$$

式中离群度 $\beta = \frac{L_{IoU}^*}{L_{IoU}} \in [0, +\infty)$ 其中 L_{IoU}^* 表示动态滑动平均值, β 表示锚框的质量, 在值大或值小时都分配较小的梯度增益, 降低对边界框回归的影响。 α, δ 表示超参数。 r 通过降低高质量样本对损失值贡献, 动态给予边界框梯度增益, 并在训练后期减少低质量锚框产生的有害梯度, 聚焦于普通质量的锚框, 提高模型定位能力。

WIoU 去除了 CIoU 中的纵横比惩罚项, 同时平衡高质量和低质量锚框对于模型回归的影响, 增强模型的泛化能力, 提高模型的整体性能。因此本文采用 WIoU 替换原始模型的边界损失函数以对模型损失函数进行优化。

1.3 训练平台与嵌入式平台

训练模型平台硬件配置为 I9 9900K CPU, Nvidia A4000 GPU, 8GB 显卡, 软件环境为 win10 系统, 64 位, Python 3.9 版本, Pytorch 2.0.0 框架, Cuda 11.7 版本。训练参数如表 2 所示。

表 2 训练参数

训练参数 Training parameters	数值 Values
初始学习率 Initial learning rate	0.01
优化器 Optimizer	SGD
优化器动量 Optimizer momentum	0.937
优化器权重衰减系数 Optimizer weight decay rate	0.000 5
训练批次数 Number of image per batch	16
迭代周期数 Number of epochs	100

模型迁移部署设备采用 Nvidia Jetson nano, 配置为 A57 四核 ARM CPU, 128 核心 Maxwell GPU, 运行内存 4GB, 64 位 LPDDR4。软件环境位 Ubuntu 18.04 系统, 运行环境的配置位 Jetpack 4.5、Python 3.6、Pytorch 1.8、TensorRT 8.0.1.6。

2 结果与分析

2.1 消融试验

消融试验结果如表 3 所示, 从表中可知, 利用部分卷积改进 C2f 模块后, 减小了参数数量和模型大小, 提高模型推理速度, 采用 BiFPN 特征融合网络增加重复双向跨尺度连接和带权重的特征融合, 进一步减小参数数量和模型大小, 提高模型召回率。利用 WIoU 损失函数提供合理的梯度增益分配策略, 迁移学习策略调整初始化模型网络结构, 提升模型检测精度、召回率、平均精度等。

由表 3 可知, 试验 2 在 YOLOv8s 网络中将原来 C2f 模块替换为 C2f-PConv 后, 模型体积缩小为基线网络的 74.67%, 模型召回率和平均精度有所下降, 其原因在于仅有部分通道进行卷积操作, 有效减小模型体积和计算冗余的同时, 丢失剩余通道可能含有的少量特征, 导致召回率和平均精度略有减小; 试验 3 替换 WIoU 损失函数后, 模型精确率、召回率、平均精度均有提升, 表明 WIoU 可以提高模型的拟合程度, 提高模型识别精度; 试验 4 替换特征融合网络为 BiFPN 后, 模型平均精度下降 0.2 个百分点, 但模型体积缩小为基线网络的 66.7%, 表明 BiFPN 能够融合更多特征, 并通过带权重的特征融合机制为正确特征分配到了更多权重; 试验 5 替换 C2f-PConv 和 WIoU, 相比于试验 2, 模型各项指标均有提升, 试验 6 在试验 5 的基础上增加 BiFPN, 模型召回率提升 0.9 个百分点, 同时模型体积减小为基线网络的 53.78%; 试验 7、8、9 在添加迁移学习的基础上, 分别加入 PConv、WIoU、BiFPN, 其中召回率相较于于

验 2、3、4 分别增加 0.6、1.3、0.7 个百分点, 平均精度分别增加 0.5、0.8、0.2 个百分点, 结合试验 10 可知, 迁移学习在不牺牲模型性能的基础上, 召回率、平均精度等指标都有提升, 表明迁移学习可以将训练达到饱和,

获得最佳的模型检测精度和泛化能力。与基线网络相比, 经过改进后, 模型体积缩小为基线网络的 53.78%, 模型精确率提升 1.3 个百分点, 召回率提升 1.5 百分点, 平均精度提升 1.8 百分点。

表 3 消融试验

Table 3 Ablation test

试验编号 Test NO.	基线网络 Base model	PConv	WIoU	BiFPN	迁移学习 Transfer learning	精确率 Precision P/%	召回率 Recall R/%	平均精确率均值 Mean average precision mAP/%	权重 Weights /MB	帧率 Frames per second FPS/(帧·s ⁻¹)
1	YOLOv8s	×	×	×	×	88.1	73.4	82.4	22.5	81.9
2	YOLOv8s	√	×	×	×	88.4	72.4	81.7	16.8	103.5
3	YOLOv8s	×	√	×	×	88.6	74.1	83.0	22.5	88.5
4	YOLOv8s	×	×	√	×	88.3	73.4	82.2	15.0	96.9
5	YOLOv8s	√	√	×	×	88.6	72.8	82.5	16.8	104.9
6	YOLOv8s	√	√	√	×	88.5	73.7	82.6	12.1	105.5
7	YOLOv8s	√	×	×	√	88.8	73.0	82.2	16.8	103.1
8	YOLOv8s	×	√	×	√	88.9	75.4	83.8	22.5	88.9
9	YOLOv8s	×	×	√	√	88.2	74.1	82.4	15.0	97.2
10	YOLOv8s	√	√	√	√	89.4	74.9	84.2	12.1	108.1

注: √, 使用该算法; ×, 不使用该算法。

Note: √, use this algorithm; ×, not using this algorithm.

2.2 其他轻量化网络对比

本试验以改进 YOLOv8 目标检测网络为基础网络, 替换主流轻量化特征提取主干网络, 如 MobileNetV3、MobileNetV2、GhostNet、ShuffleNetV2。在保持除主干网络外其他参数一致的基础上, 对比不同主干网络对于目标训练效果。

由表 4 可知, 使用 C2f-PConv 改进方法相较于 MobileNetV3、MobileNetV2、GhostNet、ShuffleNetV2, 训练精确率、召回率、平均精度都有较大优势。平均精度较其他网络分别提升 3.5、9、7.5、9.9 个百分点, 部分轻量化模型改进网络体积小于本方法, 但模型提取特征能力较差, 不能满足部署要求。

表 4 不同轻量化特征提取主干网络对比

Table 4 Comparison of different lightweight feature extraction backbone networks

主干网络 Backbone network	P /%	R /%	mAP /%	计算量 Calculation amount /GB	权重 Weights /MB
MobileNetV3	84.6	70.9	78.2	21.3	19.0
MobileNetV2	81.8	64.2	72.7	17.3	13.9
GhostNet	83.9	65.5	74.2	16.4	13.0
ShuffleNetV2	81.9	63.5	71.8	16.6	13.1
C2f-PConv	88.4	72.4	81.7	21.4	16.8

2.3 损失函数对比

改进前 YOLOv8 目标检测模型使用 CIoU 作为边界框回归损失函数, 在训练过程中拟合能力较强, 但由于预测框高宽比描述的是一个相对值, 计算时存在不确定性, 同时高质量和低质量锚框对回归损失不利, 本试验对比了 YOLOv8-PBi 模型采用 WIoU、CIoU、EIoU^[25]、SIoU^[26] 损失函数时的收敛情况, 试验结果如图 8 所示。

由图 8 可知, 采用 EIoU 时, 收敛速度最慢, 且收敛后的损失值最高; 采用 SIoU 和 CIoU 收敛速度略高于 EIoU, 收敛后损失值略低于 EIoU; 采用 WIoU 后, 训练时模型的梯度下降速度最快, 在前 15 轮迭代周期快速收敛, 收敛后的损失值远小于另外 3 种损失函数, 结合表 3 可知, 采用 WIoU 后, 准确率、召回率、平均精度等评

价指标均有提升。

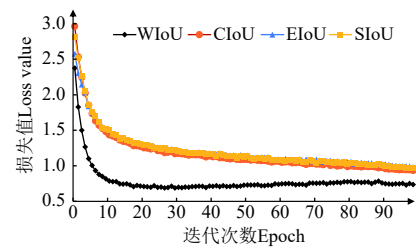


图 8 改进损失函数对比

Fig.8 Comparison of improved loss function

2.4 模型特征可视化

为更加直观的观察模型特征融合网络 BiFPN 对于识别能力提升, 采用 Grad-CAM^[27] (gradient-weighted class activation mapping) 绘制热力图, 可以较为直观的看到网络对于不同目标的学习情况。Grad-CAM 利用训练权重反向传播, 将得到的梯度矩阵进行空间维度全局平均池化, 并对特征层的各通道进行加权激活后, 得到热力图, 通过热力图中某一区域的亮度深浅可以表现出图像对于模型输出影响更大的部分。添加 BiFPN 前后的热力图如图 9 所示。

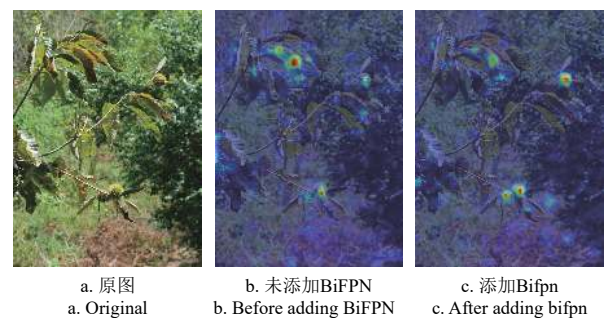


图 9 模型热力图可视化

Fig.9 Model heat map visualization

与图 9b 未替换 BiFPN 相比, 图 9c 的板栗目标处颜色更亮, 响应程度更高, 错误提取的叶片特征处亮度变

低。替换 BiFPN 后，采用快速归一化融合模块，使得模型对于正确目标的感知加强，同时抑制了错误样本对于整体预测的影响，使得模型更为准确的关注栗果目标特征。

2.5 检测模型对比试验

将基于 YOLOv8 的改进 YOLOv8-PBi 模型与 SSD、RCNN、YOLOv5s、YOLOv5 m、YOLOv7-tiny 等主流目标检测网络模型进行对比试验，结果如表 5 所示。

表 5 不同模型栗果检测结果对比
Table 5 Comparison of chestnut detection results of different models

模型 Models	P/%	R /%	mAP /%	计算量 Calculation amount /GB	权重 Weights /MB
SSD	87.77	43.81	62.87	61.2	90.6
Fast-RCNN	53.06	37.37	36.38	139.8	108.0
YOLOv5s	84.8	71.6	79.8	16.5	14.4
YOLOv5 m	87.4	73.7	82.3	49.0	42.2
YOLOv7-tiny	84.0	72.0	78.0	13.2	12.3
YOLOv8s	88.1	73.4	82.4	28.8	22.5
YOLOv8-PBi	89.4	74.9	84.2	20.8	12.1

由表 5 可知，YOLOv8-PBi 的模型体积均小于其他模型，精确率、召回率、平均精度均大于其他模型，其中平均精度分别比 SSD、RCNN、YOLOv5s、YOLOv5 m、YOLOv7-tiny、YOLOv8s 分别高出 21.33、47.82、4.4、1.9、6.2、1.8 个百分点，同时模型体相较于 SSD、RCNN、YOLOv5s、YOLOv5 m、YOLOv7-tiny、YOLOv8s，分别减小了 86.64%、88.80%、15.97%、71.33%、1.63%、46.22%。Two-stage 网络模型 Fast-RCNN 的体积和参数量较大，且识别精度较低。One-Stage 网络模型 SSD、YOLOv5s、YOLOv5 m、YOLOv7-tiny、YOLOv8s 的体积和参数量小于 Fast-RCNN，同时检测精度更高。改进后的 YOLOv8-PBi 相较于其他主流网络，对于板栗果实目标的检测效果更好。

2.6 边缘设备部署

为了验证改进 YOLOv8-PBi 模型边缘设备部署情况，同时为了提高模型的检测速度，选择 TensorRT 推理库进行加速。TensorRT 是 Nvidia 公司发布的高性能推理优化框架，能够在 Nvidia GPU 上为模型提供低延迟和高吞吐量的部署推理加速。

将 YOLOv8-PBi 模型训练权重文件转化为 wts 中间文件导入 Jetson Nano 执行编译操作，序列化模型对象生成 engine 推理引擎。对 engine 文件执行反序列化操作即可实现推理与后处理操作。设备部署检测情况如表 6 所示。

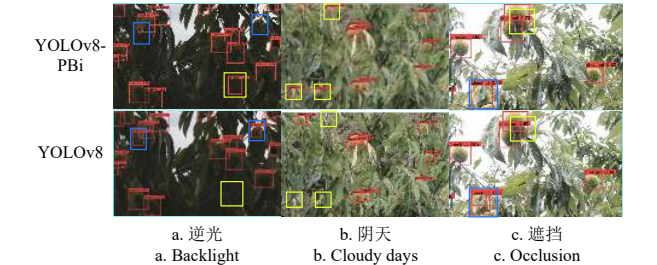
表 6 设备部署检测帧率对比
Table 6 Comparison of device deployment detection frame rates (帧·s⁻¹)

模型 Models	台式计算机 Desktop computers	嵌入式设备 Embedded devices	TensorRT
YOLOv8s	81.9	5.4	未加速
YOLOv8-PBi	108.1	6.9	43

由表 6 可知，TensorRT 加速前，改进 YOLOv8-PBi 模型由于嵌入式设备算力有限，检测速度比较低，加速后，模型检测速度提升 6.23 倍，检测帧率为 43 帧/s，检

测速度为 23.26 ms。

为进一步验证改进 YOLOv8-PBi 对复杂环境下板栗果实的检测性能，选择较难检测的远景下逆光、阴天、遮挡的板栗图像进行测试，图 10 为部分 YOLOv8-PBi 与 YOLOv8s 模型在 Jetson Nano 上的检测结果。



注：红色框是预测框，蓝色框为误检目标，黄色框为漏检目标。
Note: The red box is the prediction box, the blue box is the false detection target, and the yellow box is the missed detection target

图 10 YOLOv8-PBi 与 YOLOv8s 检测效果对比
Fig.10 Comparison of detection effects of YOLOv8-PBi and YOLOv8s

由图 10 可知，逆光时，YOLOv8s 检出错误目标 2 个、未正确检出目标 1 个，YOLOv8-PBi 预测结果均正确。在阴天光线不充足时，YOLOv8s 存在 3 处漏检，YOLOv8-PBi 未出现漏检。遮挡时，YOLOv8s 存在 1 处误检、1 处漏检，YOLOv8-PBi 未出现误检，存在 1 处漏检。3 种条件下改进模型的检测能力均优于原模型。

3 结 论

1) 本文提出了基于改进 YOLOv8 卷积神经网络的轻量化板栗目标检测算法 YOLOv8-PBi，在测试集上的准确率为 89.4%，模型权重为 12.1 MB，与原模型相比，改进后模型权重缩小为基线网络的 53.78%，精确率提升 1.3 个百分点，平均精度提升 1.8 个百分点，与 SSD、Faster RCNN、YOLOv5s、YOLOv5 m、YOLOv8s 主流目标检测网络模型相比，也具有一定的优势。

2) 引入 PConv 构建 C2f-PConv，模型权重减小 25.33%，检测性能优于替换为 MobileNetV3、MobileNetV2、GhostNet、ShuffleNet 轻量化主干网络；替换 BiFPN 网络，加强模型特征融合能力，模型权重减小为原始 YOLOv8s 的 66.7%。

3) 更改 WIoU 损失函数，精确率提升 0.5 个百分点，梯度下降速度和收敛后损失值均优于 CIoU、EIoU、SIoU 损失函数；采用迁移学习，精确率提升 0.9 个百分点，提高模型拟合速度；使用 TensorRT 推理库部署模型后，检测帧率为 43 帧/s，满足边缘设备部署要求。

[参 考 文 献]

[1] Food and Agriculture Organization. Food and Agriculture Organization of the United Nations Statistical Databases [DB/OL](2023-03-24) [2023-08-26].<https://www.fao.org/faostat/en/>.
[2] 宗望远, 黄木昌, 肖洋轶, 等. 板栗收获拍打式落果装置

- 设计与试验[J]. 农业工程学报, 2021, 37(18): 1-10.
- ZONG Wangyuan, HUANG Muchang, XIAO Yangyi, et al. Design and experiment of the fruit-beating dropping device for chestnut harvesters[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(18): 1-10. (in Chinese with English abstract)
- [3] 展慧, 李小昱, 王为, 等. 基于机器视觉的板栗分级检测方法[J]. 农业工程学报, 2010, 26(4): 327-331.
- ZHAN Hui, LI Xiaoyu, WANG Wei, et al. Determination of chestnuts grading based on machine vision[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2010, 26(4): 327-331. (in Chinese with English abstract)
- [4] 周竹, 李小昱, 李培武, 等. 基于 GA-LSSVM 和近红外傅里叶变换的霉变板栗识别[J]. 农业工程学报, 2011, 27(3): 331-335.
- ZHOU Zhu, LI Xiaoyu, LI Peiwu, et al. Near-infrared spectral detection of moldy chestnut based on GA-LSSVM and FFT[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2011, 27(3): 331-335. (in Chinese with English abstract)
- [5] MOSCETTI R, HAFF R P, SARANWONG S, et al. Nondestructive detection of insect infested chestnuts based on NIR spectroscopy[J]. *Postharvest Biology and Technology*, 2014, 87: 88-94.
- [6] 周济. 智能制造-“中国制造 2025”的主攻方向[J]. 中国机械工程, 2015, 26(17): 2273-2284.
- [7] 王金星, 马博, 王震, 等. 基于改进 Mask R-CNN 的苹果园害虫识别方法[J]. 农业机械学报, 2023, 54(6): 253-263, 360.
- WANG Jinxing, MA Bo, WANG Zhen, et al. Pest Identification Method in apple orchard based on improved mask R-CNN[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(6): 253-263, 360. (in Chinese with English abstract)
- [8] 闫建伟, 赵源, 张乐伟, 等. 改进 Faster-RCNN 自然环境下识别刺梨果实[J]. 农业工程学报, 2019, 35(18): 143-150.
- YAN Jianwei, ZHAO Yuan, ZHANG Lewei, et al. Recognition of Rosa roxbunghii in natural environment based on improved Faster RCNN[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2019, 35(18): 143-150. (in Chinese with English abstract)
- [9] LE V N T, TRUONG G, ALAMEH K. Detecting weeds from crops under complex field environments based on Faster RCNN[C]//2020 IEEE eighth international conference on communications and electronics (ICCE). Phu Quoc Island, Vietnam: IEEE, 2021: 350-355.
- [10] HTET K S, SEIN M M. Toddy palm trees classification and counting using drone video: Retuning hyperparameter mask-RCNN[C]//2021 7th International Conference on Control, Automation and Robotics (ICCAR). Singapore: IEEE, 2021: 196-200.
- [11] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Columbus, OH, USA: IEEE, 2014: 580-587.
- [12] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector[C]//Proceedings of the 14th European Conference on Computer Vision. Netherlands: ECCV, 2016: 21-37.
- [13] 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. Las Vegas, NV, USA: IEEE, 2016: 779-788.
- [14] ARAKAWA T, TANAKA T S T, KAMIO S. Detection of on-tree chestnut fruits using deep learning and RGB unmanned aerial vehicle imagery for estimation of yield and fruitload[J]. *Agronomy Journal*, 2023: 1-9.
- [15] 王梦妮, 顾寄南, 王化佳, 等. 基于改进 YOLOv5s 模型的茶叶嫩芽识别方法[J]. 农业工程学报, 2023, 39(12): 150-157.
- WANG Mengni, GU Jinan, WANG Huajia, et al. Method for identifying tea buds based on improved YOLOv5s model[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2023, 39(12): 150-157. (in Chinese with English abstract) (in Chinese with English abstract)
- [16] 黄杰, 王相友, 吴海涛, 等. 基于轻量型卷积神经网络的马铃薯薯芽眼检测算法[J]. 农业工程学报, 2023, 39(9): 172-182.
- HUANG Jie, WANG Xiangyou, WU Haitao, et al. Detecting potato seed bud eye using lightweight convolutional neural network(CNN)[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2023, 39(9): 172-182. (in Chinese with English abstract) (in Chinese with English abstract)
- [17] BAZAME H C, MOLIN J P, ALTHOFF D, et al. Detection of coffee fruits on tree branches using computer vision[J]. *Scientia Agricola*, 2022, 80: e20220064.
- [18] HOWARD A G, ZHU M, CHEN B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications [EB/OL]. (2017-04-17)[2023-01-12]. <https://arxiv.org/abs/1704.04861>.
- [19] ZHANG X, ZHOU X, LIN M, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 6848-6856.
- [20] 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, WA, USA: IEEE, 2020: 1580-1589.
- [21] CHEN J, KAO S, HE H, et al. Run, Don't Walk: Chasing higher FLOPS for faster neural networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern

- Recognition. Vancouver, BC, Canada: IEEE, 2023: 12021-12031.
- [22] TAN M, PANG R, LE Q V. Efficientdet: Scalable and efficient object detection[C]// The IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE, 2020: 10781-10790.
- [23] ZHENG Z H, WANG P, LIU W, et al. Distance-IOU loss: Faster and better learning for bounding box regression[C]// Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto, California USA: AAAI Press, 2020: 12993-13000.
- [24] TONG Z, CHEN Y, XU Z, et al. Wise-IOU: Bounding box regression loss with dynamic focusing mechanism [EB/OL]. [2023-10-20]<https://arxiv.org/abs/2301.10051>.
- [25] ZHANG Y F, REN W, ZHANG Z, et al. Focal and efficient IOU loss for accurate bounding box regression[J]. *Neurocomputing*, 2022, 506: 146-157.
- [26] ZHORA G. SIOU Loss: More powerful learning for boundingbox regression [EB/OL]. (2022-05-25)[2023-06-10]<https://arxiv.org/ftp/arxiv/papers/2205/2205.12740.pdf>.
- [27] SELVARAIU R R, COGSWELL M, DAS A, et al. Grad-cam: Visual explanations from deep networks via gradient-based localization[C]//Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017: 618-626.

Detecting chestnuts using improved lightweight YOLOv8

LI Mao, XIAO Yangyi, ZONG Wangyuan^{*}, SONG Bao

(1. College of technology, Huazhong Agricultural University, Wuhan 430070, China; 2. Key Laboratory of agricultural equipment in the middle and lower reaches of the Yangtze River, Ministry of agriculture and rural areas, Wuhan 430070, china)

Abstract: Chestnut industry has been confined to low productivity, high labor expenses, and operational hazards. Information technology can be expected to deal with these obstacles in the chestnut production. Fruit object detection can also offer essential technical assistance and solutions to the intelligent division of chestnut harvesting. In this study, a chestnut fruit object detection (YOLOv8-PBi) was proposed using the lightweight improved YOLOv8s. The research object was taken in the natural environment of Hubei Province, China. This approach brought various advancements, where the YOLOv8 was used as the basis network. Initially, the C2f-PConv module was constructed to minimize the floating-point number and computation amount using partial convolution (PConv), in order to improve the utilization of computational capacity. A weighted bidirectional feature pyramid network (BiFPN) was then employed to enhance the cross-scale connections and feature fusion performance. Thirdly, the convergence speed and detection performance were enhanced to change the bounding box loss function, in order to employ the dynamic non-monotone focusing technique Wise Intersection over Union (WIoU). Lastly, transfer learning was used to raise the accuracy and generalizability of the model. The enhanced YOLOv8-PBi model was scored by 89.4%, 74.9%, and 84.2% in the accuracy, recall, and average precision, respectively. The model weight was 46.22% less, compared with the original base network YOLOv8s. The gains were observed in the accuracy, recall, and average precision of 1.3, 1.5, and 1.8 percentage points, respectively. Additionally, the PC inference speed increased from 81.9 to 108 frames per second. The average detection accuracy increased by 0.8 percentage points after substituting the original bounding box loss function, CIOU, for WIoU. Both the gradient descent and fitting speeds increased significantly. BiFPN was replaced with the original feature fusion network, in order to increase the model recall rate by 0.9 percentage points. The model was allowed to concentrate more precisely on the attributes specific to chestnut fruit objects. The enhanced performance was achieved in the C2f-PConv approach, with an average accuracy of 3.5, 9.0, 7.5, and 9.9 percentage points greater than the standard lightweight feature extraction networks of MobileNetV3, MobileNetV2, GhostNet, and ShuffleNetV2. The enhanced WIoU loss function demonstrated that the WIoU had the lowest loss value and the fastest iterative convergence speed, compared with the CIOU, EIOU, and SIOU. The model size was the smallest, compared with the mainstream object identification models, such as SSD, Fast-RCNN, YOLOv5s, YOLOv5m, YOLOv7-tiny, and YOLOv8s. The average accuracy was found to be 21.33, 47.82, 4.4, 1.9, 6.2, and 1.8 percentage points greater than those of the other models, respectively. The model was finally deployed on the edge-embedded devices using TensorRT acceleration. The better performance was achieved in the detection frame rate of 43 frames per second and a detection speed of 23.26 ms, fully meeting the device deployment requirements. YOLOv8-PBi enhanced the misdetection and missed detection scenarios in the environments of backlight, overcast, and occlusion, compared with YOLOv8s. This finding can provide a technical foundation to identify the chestnut fruits in the process of intelligent chestnut harvesting.

Keywords: image processing; image recognition; YOLOv8; chestnut; Object detection; PConv; BiFPN