

基于改进 YOLOv8s 的轻量级果园李子检测方法

张冬妍, 陈 诺, 张 淇, 吴晨旭, 张榄翔

(东北林业大学计算机与控制工程学院, 哈尔滨 150040)

摘 要: 为了解决果园李子受枝叶和果实遮蔽、环境变化等因素影响, 难以准确检测的问题, 该研究提出了一种基于改进 YOLOv8s 的轻量级果园李子检测模型。首先, 采用自设计主干网络 Faster-EMA 缩减模型复杂度、提高检测精度。其次, 引入焦点调制网络 (focal modulation) 替换原模型中的 SPPF 模块增强特征融合能力, 丰富特征提取的语义信息; 最后, 引入参数共享策略并实现轻量级检测头 LDetect, 满足了低功耗嵌入式设备部署需求。试验结果表明, 优化后模型的平均检测精度达到 97.2%, 与原模型相比, 检测精度提高了 7.4 个百分点; 模型计算量降低了 44.8%; 模型参数数量减小了 25.8%; 部署在边缘计算设备 Jetson Nano 4GB 上, 检测帧率达到了 48.3 帧/s。该研究所提出的方法能有效的解决复杂背景下果园李子的智能化检测, 有助于促进李子智能化采摘技术的发展。

关键词: 目标检测; YOLOv8; 轻量化; 李子; LDetect

doi: 10.11975/j.issn.1002-6819.202408158

中图分类号: S24

文献标志码: A

文章编号: 1002-6819(2025)-01-0154-07

张冬妍, 陈诺, 张淇, 等. 基于改进 YOLOv8s 的轻量级果园李子检测方法[J]. 农业工程学报, 2025, 41(1): 154-160.

doi: 10.11975/j.issn.1002-6819.202408158 <http://www.tcsae.org>

ZHANG Dongyan, CHEN Nuo, ZHANG Qi, et al. Detecting plum fruits in orchard using lightweight improved YOLOv8s[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2025, 41(1): 154-160. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.202408158 <http://www.tcsae.org>

0 引 言

李子在中国已有多年的种植历史^[1], 其果实富含多种维生素、氨基酸以及微量元素, 具有美容养颜、促进消化、增强人体免疫力的功能^[2], 具有较高的营养价值和药用价值^[3]。目前李子的采摘主要依赖人工, 存在操作强度大、效率低、周期长、劳动力短缺等问题^[4], 无法在最佳采摘时期进行李子收获限制了李子产业的发展。

随着传统农业朝着智能化发展^[5], 越来越多的人将信息技术与传统机械相结合, 开发采摘机器人, 实现精准农业^[6]。水果的准确检测是采摘机器人在复杂环境果园采摘的必然要求。目前, 国内外对果实检测的研究已经取得了一定的进展^[7]。BAI 等^[8]对番茄的形状、纹理和颜色特征进行提取和组合, 实现番茄的准确识别。YU 等^[9]利用颜色和纹理特征训练随机森林二分类模型来识别荔枝果实。WAN 等^[10]根据菠萝的形状、颜色和质地对提取的特征进行分析以识别菠萝冠部。上述方法均为传统的机器学习, 该方法主要依赖于对特征的人工选择, 虽然可以完成水果的检测任务, 但检测精度易受光照不均匀、水果遮挡、枝叶之间重叠等影响, 不能满足应用场景的要求^[11]。

与传统机器学习方法相比, 基于深度学习的目标检

测在农业领域的应用显著提高了复杂环境中物体检测的精度, 解决了传统目标检测方法中所存在的问题^[12]。典型的两阶段算法包括 R-CNN^[13-16]系列引入了候选区域生成网络, 该网络能够生成图像中的候选目标区域, 并通过分类器和回归器进行目标检测, 从而实现较高的检测精度。但因检测速度慢, 不能满足实际场景果实检测模型的边缘计算设备部署条件^[17]。

相对于两阶段目标检测算法, 单阶段检测算法不需要生成候选区域、速度更快。目前常用的单阶段检测算法是 YOLO 系列, 该算法通过卷积神经网络预测目标的类别和位置^[18], 更适合应用于果实智能采摘系统。李天华等^[19]提出了基于 YOLOv4 的成熟番茄目标检测方法, 平均检测精度为 94.77%。TANG 等^[20]开发了一种基于 YOLOv4-tiny 模型的油茶果实检测模型, 实现了油茶果实的特征提取以及检测, 检测精度达到 92.1%。张日红等^[21]为解决人工菠萝催化操作效率低等问题, 提出了基于改进 YOLOv4 的菠萝苗芯模型, 种植密集与稀疏条件的检测精度分别为 94.7% 和 95.5%。NAN 等^[22]基于 YOLOv3 开发出了 WGB-YOLO 网络, 解决了光照变化、树枝遮挡等因素对火龙果检测精度的影响, 检测精度达到了 86.0%, 在植物果园中表现出良好的检测结果。

综上所述, 现有方法极大地推进了目标检测的发展, 但对于复杂环境中果园李子检测方面, 未考虑李子果型小且分布密集、存在枝叶和果实遮蔽等问题。同时现有方法存在检测精度低和模型复杂度高问题, 无法满足低能耗边缘计算设备的部署要求。因此轻量化、高精度的检测模型研究较为重要, 本研究提出了一种基于

收稿日期: 2024-08-21 修订日期: 2024-12-02

基金项目: 国家自然科学基金项目 (32202147)

作者简介: 张冬妍, 博士, 副教授, 研究方向为人工智能应用。

Email: nefuzhdzdy@nefu.edu.cn

YOLOv8s 的轻量级李子检测模型。通过对基线模型的主干网络、颈部网络以及检测头进行改进,从整体上提高模型检测准确性和速度,以满足机械化收获过程中果园李子目标检测要求。

1 材料与方法

1.1 数据集建立

本研究选择芙蓉李作为研究对象,芙蓉李图像采集于黑龙江省齐齐哈尔市某一果园。本次图像采集使用的设备为小米 10S 安卓智能手机,像素为 2340×1080 。由于芙蓉李背景复杂,检测精度易受光照强度因素等影响,数据集将从不同的光线条件、拍摄角度、拍摄距离等进行采集。针对实际果园中可能遇到的多种恶劣条件,数据集涵盖了强光、弱光、阴天、雨天、雾天、图像模糊以及风天等典型场景,共收集了 1 220 幅具有代表性的图像样本。多样化的数据集增强了模型对不同天气条件的适应能力,从而提升了李子检测模型在复杂采摘环境下的鲁棒性,部分数据集展示如图 1 所示。



图 1 李子采集图像示例

Fig.1 Example of plum acquisition image

使用开源数据标注软件 LabImg 进行人工数据集标注^[23],生成 xml 文件,数据集按照 8:1:1 的比例随机划分训练集、验证集以及测试集。

1.2 评价指标与参数设置

在目标检测任务中,不同评价指标可以从不同角度反应检测算法的性能。为了更全面的评估本研究模型的检测性能,采用精确率 (P)、召回率 (R)、平均精确率均值 (mAP)、参数量 ($Param$) 以及浮点运算次数 ($FLOPs$) 等指标,进行综合评价。

本研究模型的训练和测试过程在 Windows11 操作系统上实现的。相关配置为 Intel(R) Core(TM) i9-12900H

CPU 和 NVIDIA GeForce RTX 3070 GPU。软件环境基于模型在 Pytorch1.13.0、Python3.8 以及 CUDA11.7 搭建的深度学习框架。初始学习率为 0.001,图像数据输入尺寸为 640×640 像素,训练迭代次数为 300。

2 网络模型的改进

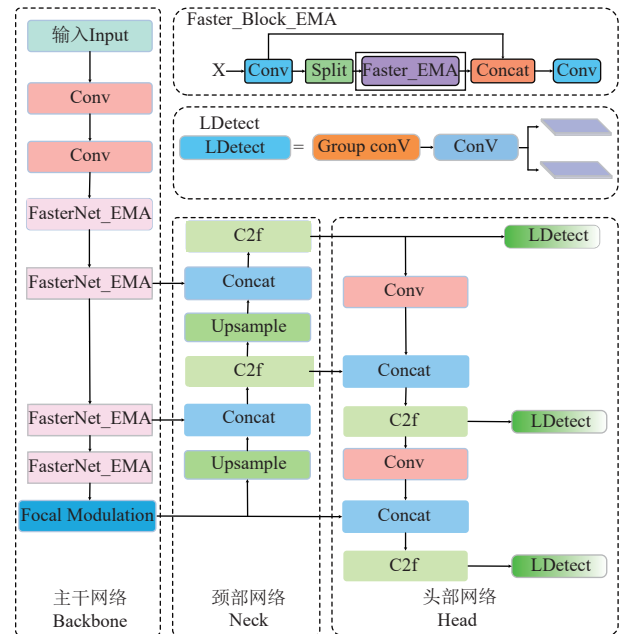
在复杂的果园环境中智能化采摘李子需要保持高精度,并且模型的复杂度需满足低消耗、低功率的边缘计算设备的部署要求。本研究基于改进 YOLOv8s 设计了一种高检测精度、低模型复杂度的算法,主要有以下有效改进与优化:

1) 首先,通过融合注意力机制的方法,采用自设计的 Faster-EMA 模块对 YOLOv8s 主干网络进行优化,提高模型对李子的特征提取能力。

2) 其次,采用焦点调制网络替代原模型的空间金字塔池化 (SPPF),保持模型复杂度的同时,进一步提高了检测精度。

3) 最后,通过共享参数的策略,引入组卷积 (group conv) 实现检测头的轻量化,降低了模型的计算复杂度和参数量、加快检测速度以适应边缘计算设备的资源限制。

改进后的网络模型结构图如图 2 所示。



注: Conv 为卷积模块,Concat 为特征连接模块,Upsample 为上采样模块,LDetect 为检测头,Focal Modulation 为改进的金字塔池化模块,X 代表输入。

Note: Conv refers to the convolution module, Concat refers to the feature concatenation module, Upsample refers to the upsampling module, LDetect refers to the detection head, Focal Modulation refers to the improved spatial pyramid pooling module, X represents input.

图 2 改进 YOLOv8s 网络结构图

Fig.2 Improved YOLOv8s network architecture diagram

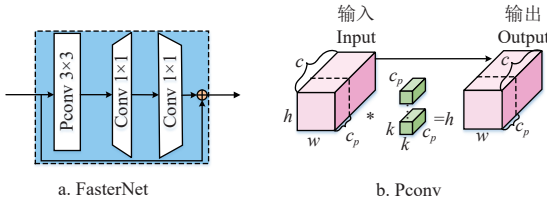
2.1 主干网络的优化

李子具有形状小、生长密集、容易被果实或枝叶遮挡的特点^[24]。当 YOLOv8s 的主干网络进行特征提取时存在小目标的定位能力弱、信息特征提取不充分的问题。

因此,本研究提出了Faster_EMA模块对YOLOv8s的主干网络进行优化。本文将EMA(Efficient multi-scale attention)多尺度注意力机制^[25]与FasterNet Block进行融合,利用两者优点提高网络特征提取能力、降低模型计算量。

EMA注意力机制采用 3×3 内核与 1×1 分支并行放置的方式,利用2D全局平均池化在 3×3 分支中编码全局空间信息, 1×1 分支在通道特征联合激活前调整为目标维度,从而生成第二个空间注意力图。最后,每组内输出的特征图通过聚合两个生成的空间注意力权重值进行计算。EMA注意力机制的这种工作方式使空间语义特征在每个特征组中均匀分布,能够充分获取不同尺度的特征信息并提高李子的检测精度。

FasterNet Block采用一种创新的PConv卷积,如图3a所示,在FasterNet Block中,一个PConv层后边跟着两个PWConv层,将归一化和激活层放在中间以保持特征多样性并实现低延迟。PConv利用特征图中的冗余性,只对部分输入通道应用常规卷积并进行空间特征提取,不触及其余通道。如图3b所示,本研究采用卷积率 $r=c_p/c$ 为1/4,模型的计算量仅为常规卷积的1/16,这种设计方式使得网络能够以更低的计算成本提取高质量的空间特征,从而显著减少计算冗余和内存访问。



注: c 表示输入特征图的通道数, c_p 表示参与卷积通道数, h 、 w 分别表示特征图的高和宽, k 表示卷积核大小, $*$ 表示分离操作。
Note: c represents the number of channels of the input feature map, c_p represents the number of convolutional channels involved, h and w represent to the height and width of the feature map, and k denotes the size of the convolution kernel, $*$ represents the separation operation.

图3 FasterNet与Pconv模块结构图

Fig.3 FasterNet and Pconv block structure diagram

本研究将在FasterNet Block中提取特征经过卷积层后,融入EMA注意力机制。FasterNet Block对输入图像进行特征提取,通过PConv从输入图像中提取多尺度空间特征,并保持较低的计算开销。借助EMA注意力机制对卷积输出进行加权,自动调整不同尺度信息的权重,增加模型对果园李子特征的敏感度。多尺度特征图的加权组合进一步提升了模型在处理不同尺度信息时的能力。这种融合方式能够将来自不同尺度的特征信息进行综合,从而增强模型空间语义信息提取能力和模型在复杂场景下进行果园李子检测的能力。

2.2 空间金字塔池化改进

多尺度池化操作虽然能够提升模型的多尺度特征处理能力,但会显著增加计算量,并且对小目标检测的效果有限,部署在低功耗的边缘嵌入式设备上,计算复杂度更加明显。因此将Neck部分的SPPF替换为焦点调制模块Focal Modulation^[26],该模块的核心包括焦点上下文

信息获取、门控聚合、逐元素仿射变换三部分,通过堆叠深度卷积层编码不同范围视觉的上下文信息,捕捉从近到远的视觉特征,能保持对局部细节的敏感性,增强对全局结构的理解。Focal Modulation模块采用两阶段聚合过程,如图4所示,首先输入一个特征图 $X \in \mathbf{R}^{h \times w \times c}$,其中 h 和 w 分别表示特征图的高度和宽度, c 表示特征图通道数, Z^0 代表新特征空间的输入。将其投影到具有线性层为 $f_z(\cdot)$ 的新特征空间如式(1):

$$Z^0 = f_z(X) \in \mathbf{R}^{h \times w \times c} \quad (1)$$

其次利用卷积核大小为 k^m 的深度卷积(DWConv)对 $m-1$ 层输入进行处理得到第 m 层输出,接着通过GeLU激活函数进行非线性变换实现 L 个深度卷积堆栈获得上下文信息 Z^m 的分层表示如式(2):

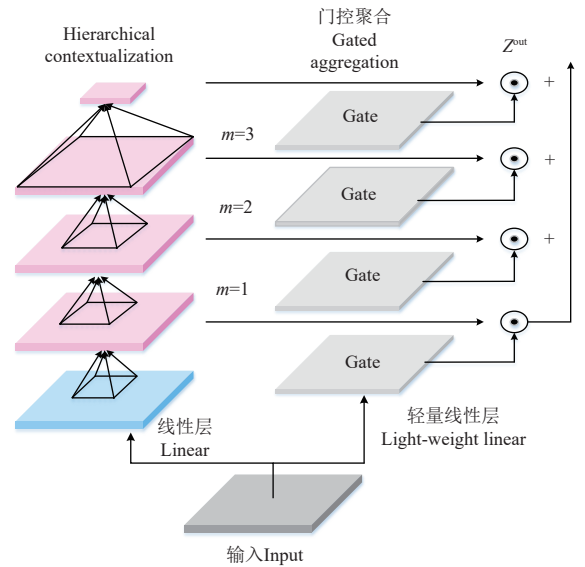
$$Z^m = f_a^m(Z^{m-1}) \triangleq \text{GeLU}(\text{DWConv}(Z^{m-1})) \in \mathbf{R}^{h \times w \times c} \quad (2)$$

通过分层上下文获得 L 张特征图并对第 L 层特征图进行全局平均池化,最终得到 $L+1$ 张特征图,捕获不同粒度级别的上文。式中 $m \in \{1, 2, \dots, L\}$ 代表焦点层, f_a^m 代表第 m 层的上下文信息函数。

接下来利用线性层计算空间门控权重 G ,通过元素乘法执行加权和,利用门控机制控制每个不同级别聚合的数量,以获得与输入 X 大小相同的输出特征如式(3):

$$Z^{\text{out}} = \sum_{m=1}^{L+1} G^m \odot Z^m \in \mathbf{R}^{h \times w \times c} \quad (3)$$

式中 $G^m \in \mathbf{R}^{h \times w \times 1}$,是门控权重 G 在第 m 个焦点层中用于实现对该层特征图进行加权调整的特征切片,实现门控聚合, \odot 代表乘法关系。



注: Z^{out} 代表门控聚合后输出的特征。
Note: Z^{out} represents for features output after gated aggregation.

图4 上下文聚合原理图

Fig.4 Context aggregation principle diagram

实现图中所述的门控聚合后,Focal Modulation模块最后利用另一个线性层 $h(\cdot)$ 函数对聚合的输出特征图进行

线性变换实现焦点调制器 M 如式 (4)：

$$M = h(Z^{\text{out}}) \in \mathbf{R}^{h \times w \times c} \quad (4)$$

Focal Modulation 模块的三个核心步骤相结合能够更有效地聚焦于当前小目标检测任务的关键信息，减少不必要的信息处理以及解决计算时间和内存的限制问题，这种精细的调整机制允许果园李子检测模型更好地适应复杂的视觉场景，提高对细节语义信息的捕捉能力，从而提升模型在果园李子检测任务中的特征提取能力。

2.3 轻量化检测头

YOLOv8s 的检测头采用的是解耦头，其模型运算量占整个模型的 41%。解耦头有两个分支，每一个分支都是由两个 3×3 卷积以及一个 1×1 卷积组成，提取特征图后，需要遍历每一个卷积才能实现目标检测，增加了模型的复杂度和计算参数。

因此针对以上问题本研究设计了一个轻量级检测头 LDetect 代替解耦头。首先引入 3×3 的组卷积替换常规的 3×3 卷积。组卷积通过较少的参数量和计算量生成大量特征图，从而能够提取更多的语义信息。其次采用共享卷积参数的策略，将基线模型检测头的两个分支共享同一个组卷积和 1×1 卷积。这种优化方式在保证检测精度的同时，降低模型参数量和复杂度，有利于边缘计算设备的部署。

3 结果与分析

3.1 主流模型对比试验结果

为了验证本研究所提出的模型在果园李子检测方面具有优势，将 Faster-RCNN、YOLOv4-Tiny、YOLOv5s、YOLOv7n、YOLOv8s 共 5 种主流模型与本研究改进后的模型 (Improved YOLOv8s) 在相同条件下进行对比试验，结果如表 1 所示。

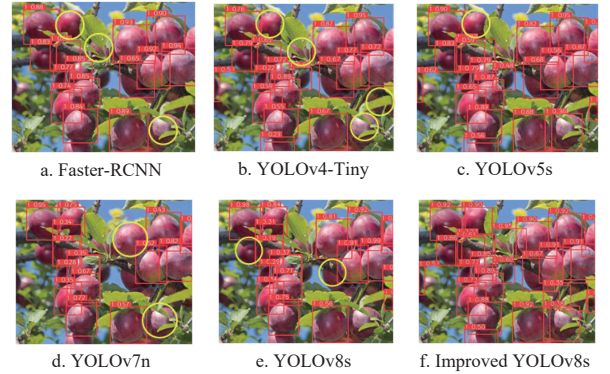
表 1 主流模型对比结果

Table 1 Mainstream model comparison test results

模型 Model	平均精确率 Mean average precision mAP/%	精确率 Precision P/%	召回率 Recall R/%	运算符号点数 Floating point operations FLOPs/G	参数量 Parameters/ 10^6
Faster-RCNN	85.5	82.7	72.2	121.4	60.3
YOLOv4-Tiny	86.6	81.6	72.4	3.2	5.9
YOLOv5s	90.1	88.4	82.2	15.8	13.7
YOLOv7n	80.3	85.6	82.2	105.1	71.3
YOLOv8s	89.8	87.6	86.6	8.7	6.2
Improved YOLOv8s	97.2	95.3	90.5	4.8	4.6

如表 1 结果分析可知，与 Faster-RCNN、YOLOv4-Tiny、YOLOv7n 相比，本研究采用的基线模型 YOLOv8s 在各项评价指标上表现优越，mAP 分别提高了 4.3、3.2、9.5 个百分点。与 YOLOv5s 相比，YOLOv8s 在计算量和参数量方面均低于 YOLOv5s，表明了其在保持较高精度的同时，具备更高的计算速度和更低的模型复杂度。本研究改进后的模型 (Improved YOLOv8s) 与基线模型 YOLOv8s 相比较，mAP、 P 以及召回率分别提高了 7.4、7.7、3.9 个百分点，模型计算量降低了 44.8%，模型参数

量减小了 25.8%。与轻量级检测模型 YOLOv4-Tiny 相比较，本研究模型计算量略有增加，但 mAP 却提高了 10.6 个百分点，模型参数量减小了 22.0%，更好的平衡了检测精度与计算复杂度之间的关系。图 5 为 6 种模型的检测效果，可以直观看到综合不同检测模型性能对比，本研究所提出的模型在检测精度、计算复杂度方面均表现优异，表明该模型更适合用于果园李子检测系统，能够满足实际应用需求。



注：圆圈代表漏检目标，框是预测框。

Note: The circle represents the missed target and the box is the prediction box.

图 5 不同模型可视化对比结果

Fig.5 Comparison results of different model visualizations

3.2 消融试验结果

为了验证每一处改进模块在果园李子检测上的有效性，以 YOLOv8s 为基础网络，对每一处优化进行消融试验，结果如表 2 所示。

表 2 YOLOv8s 消融试验结果

Table 2 The results of YOLOv8s ablation experiments

Faster-EMA	Focal Modulation	LDetect	mAP/%	P/%	R/%	FLOPs/G	参数量 Parameters/ 10^6
×	×	×	89.8	87.6	86.6	8.7	6.2
√	×	×	93.2	91.6	90.7	7.1	5.5
√	√	×	96.9	94.7	92.2	7.2	5.8
√	√	√	97.2	95.3	90.5	4.8	4.6

根据表 2 试验结果分析可知，与基线模型 YOLOv8s 相比，采用融合 EMA 注意力机制的 FasterNet Block 模块作为模型主干网络，mAP、 P 以及召回率分别提升了 3.4、4.0、4.1 个百分点；计算量下降了 18.4%；模型参数量减少了 11.3%，这表明 Faster-EMA 主干网络提升了模型对小目标的特征提取能力和平衡了模型计算量，使模型更能适应复杂环境下果园李子检测。基于前面的改进引入 Focal Modulation 代替 SPPF 的方法使 mAP、 P 以及召回率再次提升 3.7、3.1、1.5 个百分点，这种较大幅度的提升主要是因为 Focal Modulation 模块采用焦点上下文、门控聚合、逐元素仿射变换三部分结合的方式，本研究利用该模块分步门控聚合的功能，在保证果园李子检测模型计算量和复杂度一定的条件下，增强了模型对空间上下文信息的感知能力和目标定位能力。轻量化检测头 LDetect 的引入使模型计算量再次下降了 33.3%，模型参数量再次减少了 20.7%，表明轻量化检测头适用于实际场景李子检测设备的部署。

3.2.1 主干网络对比试验结果

YOLOv8s 的主干网络在特征提取时, 存在对小目标定位能力弱, 特征提取不充分导致检测小目标效率低下的问题。为了提高网络对果园李子的定位能力, 采取轻量网络 FasterNet Block 融合 EMA 注意力机制的方式。将 8 种主流主干网络与本研究自设计融合注意力机制主干网络 Faster-EMA 相比较, 结果如表 3 所示。

表 3 YOLOv8s 主干网络对比结果

Table 3 YOLOv8s backbone network comparison results

主干网络 Backbone	mAP/%	P/%	R/%	FLOPs/G	参数量 Parameters/10 ⁶
Darknet-53	89.8	87.6	86.6	8.7	6.2
Convnextv2	84.8	86.5	86.0	14.1	11.5
Repvit	89.7	91.3	87.9	19.1	14.0
EfficientFormer	85.4	89.4	86.3	8.2	39.5
Swintransformer	89.3	89.5	88.6	79.1	60.5
MobileNet	88.7	90.2	86.8	22.6	11.7
EfficientDet	89.6	91.6	87.4	7.2	4.9
GhostNet	86.7	88.5	82.6	5.1	3.9
Faster-EMA	93.2	91.6	90.7	7.1	5.5

由表 3 结果分析可知, 本研究的 Faster-EMA 主干网络与基线模型的主干网络 Darknet-53 相比较, mAP 提高了 3.4 个百分点, 模型计算量和模型参数量分别减小了 18.4%、11.3%。表明采用注意力机制 EMA 融合 FasterNet Block 方法可以解决小目标定位能力弱、模型计算量较大的问题。与经典轻量化主干网络 MobileNet 和 EfficientDet 相比, 本研究主干网络模型的计算量与参数量最低, mAP 分别高出 4.5、3.6 个百分点。与 GhostNet 相比, 虽然模型计算量和参数量略大, 但在增加小幅度计算量的同时检测精度提高了 6.5 个百分点, 保证了复杂场景下果园李子的准确性。

3.2.2 检测头对比试验结果

为了解决基线模型检测头参数量大、训练速度慢的问题, 本研究采用了共享参数的思想, 利用组卷积优化原模型的检测头, 结果如表 4 所示。

表 4 检测头对比结果

Table 4 Comparison results of detection head

检测头 Detect	mAP/%	P/%	R/%	FLOPs/G	参数量 Parameters/10 ⁶
Detect	89.8	87.6	86.6	8.7	6.2
Detect_PConv	89.4	87.5	86.9	4.6	4.3
Detect_RepConv	89.5	88.6	83.7	7.5	7.6
LDetect	92.6	92.3	89.7	4.7	4.3

结果表明, 本研究分别采用 PConv^[27]、RepConv 以及 Group Conv 对检测头进行改进, Group Conv 改进的检测头的方式平均检测精度提高幅度最大、模型计算量和参数量较低, 可以更好地平衡检测精度和模型计算量、复杂度之间的关系。与基线模型 YOLOv8s 的检测头相比较, 引用轻量化检测头 LDetect 的果园李子检测模型计算量下降了 46.0%、参数量减小了 30.6%, 检测精度、预测精度以及召回率分别提升了 2.8、4.7、3.1 个百分点。综上性能指标对比分析, 本研究的轻量化检测头在模型部署能力和小目标检测能力上都具有很大的优势, 在复杂环境下果园李子检测方面具有准确性高、计

算速度快的特点。

3.3 边缘设备部署试验结果

智能农业采摘嵌入式设备计算能力和功耗都是有限的, 为了说明本研究所提出的模型在实际应用有一定的可行性, 采用 NVIDIA Jetson Nano 4GB 作为边缘部署设备, 选择 TensorRT 推理库进行模型加速实现低功率、高效地果园李子检测。YOLOv8s 和改进的 YOLOv8s 模型训练权重文件不利于实际的推理部署, 因此为了进一步满足实际检测需求, 将当前的权重文件转化为 onnx 文件进行 TensorRT 加速部署。

由设备部署结果分析, 改进前后的模型部署在 Jetson Nano 4GB 上的每秒帧数 (FPS) 分别是 25.8 和 48.3 帧/s。改进的 YOLOv8s 高于基线模型 22.5 帧/s; 从推理耗时的角度, 改进前后模型推理时间分别为 48.3、19.7 ms, 改进的 YOLOv8s 推理时耗下降了 59.2%。综上所述, 改进的 YOLOv8s 模型推理速度优于基线模型, 在计算能力有限的边缘计算设备上, 成本与推理时耗方面有优势。

4 结 论

本研究针对准确、快速检测果园李子问题, 采用 YOLOv8s 作为基线模型进行优化, 提出了一种基于改进 YOLOv8s 的轻量级李子检测模型。该模型有效解决了复杂环境果园李子受枝叶果实遮蔽、环境和背景影响等问题。试验结果表明, 与 Faster-RCNN、YOLOv4-Tiny、YOLOv5s、YOLOv7n、YOLOv8s 主流检测模型相比, 本研究的模型平均精度均值分别提高了 11.7、10.6、7.1、16.9、7.4 个百分点, 改进的 YOLOv8s 达到 97.2%。与基线模型相比, 模型计算量和参数量分别下降 44.8%、25.8%。将本研究模型部署到 Jetson Nano 4GB 嵌入式设备上每秒帧数达到 48.3 帧/s, 每张数据集的推理时耗下降了 59.2%, 推理速度较快。消融试验结果表明, 在面对复杂环境, 改进的 YOLOv8s 模型保证高检测精度的同时实现了便捷式边缘计算设备的部署。

[参 考 文 献]

- [1] 李可, 林籽汐, 刘佳, 等. 基于主成分分析和聚类分析的李子果实品质综合评价[J]. 食品工业科技, 2024, 45(8), 293-300.
LI Ke, LIN Zixi, LIU Jia, et al. Comprehensive evaluation of plums quality based on principal component analysis and cluster analysis[J]. Science and Technology of Food Industry, 2024, 45(8), 293-300. (in Chinese with English abstract)
- [2] 周美, 王瑜, 李青, 等. 李子薏苡仁复合酵素中营养成分及香气成分研究[J]. 食品科技, 2020, 45(7): 309-316.
ZHOU Mei, WANG Yu, LI Qing, et al. Study on the nutritional and aroma components of plum and coix seed composite enzyme[J]. Food Science and Technology, 2020, 45(7): 309-316. (in Chinese with English abstract)
- [3] 董振山, 张欣敏, 农怀园, 等. GC-MS 法分析李子提取物中

- 挥发性成分[J]. 中国食品添加剂, 2020, 31(10): 105-108.
- DONG Zhenshan, ZHANG Xinmin, NONG Huaiyuan, et al. Analysis of volatile components of plum extract by GC-MS[J]. China Food Additives, 2020, 31(10): 105-108. (in Chinese with English abstract)
- [4] MA B, HUA Z, WEN Y, et al. Using an improved lightweight YOLOv8 model for real-time detection of multi-stage apple fruit in complex orchard environments[J]. Artificial Intelligence in Agriculture, 2024, 11: 70-82.
- [5] 李茂, 肖洋铁, 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): 201-209.
- 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): 201-209. (in Chinese with English abstract)
- [6] ZHANG B, XIA, Y, WANG R, et al. Recognition of mango and location of picking point on stem based on a multi-task CNN model named YOLOMS[J]. Precision Agriculture, 2024: 25(2): 1454-1476.
- [7] 孙俊, 吴兆祺, 贾忆琳, 等. 基于改进 YOLOv5s 的果园环境葡萄检测[J]. 农业工程学报, 2023, 39(18): 192-200.
- SUN Jun, WU Zhaoqi, JIA Yilin, et al. Detecting grape in an orchard using improved YOLOv5s[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2023, 39(18): 192-200. (in Chinese with English abstract)
- [8] BAI Y, MAO S, ZHOU J, et al. Clustered tomato detection and picking point location using machine learning-aided image analysis for automatic robotic harvesting[J]. Precision Agriculture, 2023, 24(2): 727-743.
- [9] YU L, XIONG J, FANG X, et al. A litchi fruit recognition method in a natural environment using RGB-D images[J]. Biosystems Engineering, 2021, 204: 50-63.
- [10] WAN N S R, MUHAMMAD A H, MEGAT S A M A, et al. Automated image identification, detection and fruit counting of top-view pineapple crown using machine learning[J]. Alexandria Engineering Journal, 2022, 61(2): 1265-1276.
- [11] FAN Y, CAI Y, YANG H. A detection algorithm based on improved YOLOv5 for coarse-fine variety fruits[J]. Journal of Food Measurement and Characterization, 2024, 18(2): 1338-1354.
- [12] 岳凯, 张鹏超, 王磊, 等. 基于改进 YOLOv8n 的复杂环境下柑橘识别[J]. 农业工程学报, 2024, 40(8): 152-158.
- YUE Kai, ZHANG Pengchao, WANG Lei, et al. Recognizing citrus in complex environment using improved YOLOv8n[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(8): 152-158. (in Chinese with English abstract)
- [13] CHU P, LI Z, LAMMERS K, et al. Deep learning-based apple detection using a suppression mask R-CNN[J]. Pattern Recognition Letters, 2021, 147: 206-211.
- [14] LI J, ZHU Z, LIU H, et al. Strawberry R-CNN: Recognition and counting model of strawberry based on improved faster R-CNN[J]. Ecological Informatics, 2023, 77: 102210.
- [15] MACÍAS-MACÍAS M, SÁNCHEZ-SANTAMARIA H, GARCIA C J, et al. Mask R-CNN for quality control of table olives[J]. Multimedia Tools and Applications, 2023, 82(14): 21657-21671.
- [16] SIRICHAROEN P, YOMSATIEANKUL W, BUNSRI T. Fruit maturity grading framework for small dataset using single image multi-object sampling and Mask R-CNN[J]. Smart Agricultural Technology, 2023, 3: 100130.
- [17] 谭厚森, 马文宏, 田原, 等. 基于改进 YOLOv8n 的香梨目标检测方法[J]. 农业工程学报, 2024, 40(11): 178-185.
- TAN Housen, MA Wenhong, TIAN Yuan, et al. Improved YOLOv8n object detection of fragrant pears[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(11): 178-185. (in Chinese with English abstract)
- [18] 林森, 刘美怡, 陶志勇. 采用注意力机制与改进 YOLOv5 的水下珍品检测[J]. 农业工程学报, 2021, 37(18): 307-314.
- LIN Sen, LIU Meiyi, TAO Zhiyong. Detection of underwater treasures using attention mechanism and improved YOLOv5[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(18): 307-314. (in Chinese with English abstract)
- [19] 李天华, 孙萌, 丁小明, 等. 基于 YOLO v4+HSV 的成熟期番茄识别方法[J]. 农业工程学报, 2021, 37(21): 183-190.
- LI Tianhua, SUN Meng, DING Xiaoming, et al. Tomato recognition method at the ripening stage based on YOLO v4 and HSV[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(21): 183-190. (in Chinese with English abstract)
- [20] TANG Y, ZHOU H, WANG H, et al. Fruit detection and positioning technology for a *Camellia oleifera* C. Abel orchard based on improved YOLOv4-tiny model and binocular stereo vision[J]. Expert Systems with Applications, 2023, 211: 118573.
- [21] 张日红, 区建爽, 李小敏, 等. 基于改进 YOLOv4 的轻量化菠萝苗心检测算法[J]. 农业工程学报, 2023, 39(4): 135-143.
- ZHANG Rihong, Ou Jianshuang, LI Xiaomin, et al. Lightweight algorithm for pineapple plant center detection based on improved an YOLOv4 model[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2023, 39(4): 135-143. (in Chinese with English abstract)
- [22] NAN Y, ZHANG H, ZENG Y, et al. Intelligent detection of Multi-Class pitaya fruits in target picking row based on WGB-YOLO network[J]. Computers and Electronics in Agriculture, 2023, 208: 107780.
- [23] LIN Y, HUANG Z, LIANG Y, et al. AG-YOLO: A rapid citrus fruit detection algorithm with global context fusion[J]. Agriculture, 2024, 14(1): 114.

- [24] WANG L, ZHAO Y, LIU S, Li Y, et al. Precision detection of dense plums in orchards using the improved YOLOv4 model[J]. *Frontiers in Plant Science*, 2022, 13: 839269.
- [25] OU Y D, HE S, ZHANG G, et al. Efficient multi-scale attention module with cross-spatial learning[C]// In ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes 2023.
- [26] YANG J, LI C, DAI X, et al. Focal modulation networks[C]// *Advances in Neural Information Processing Systems*, 2022: 35, 4203-4217.
- [27] CHEN J, KAO S, HE H, et al. Run, don't walk: Chasing higher FLOPS for faster neural networks[C]// In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Vancouver 2023.

Detecting plum fruits in orchard using lightweight improved YOLOv8s

ZHANG Dongyan , CHEN Nuo , ZHANG Qi , WU Chenxu , ZHANG Lanxiang

(College of Computer and Control Engineering, The University of Northeast Forestry, Harbin 150040, China)

Abstract: Plum fruit is often requiring timely harvesting during specific seasons, due to the significant nutritional and culinary value. However, it is still challenging to accurately detect the plums in real-world orchard environments, such as the shading from foliage, and the overlapping of fruits. These influencing factors have also posed the higher demands on intelligent harvesting, in terms of speed, accuracy and real-time performance. In this study, an efficient and reliable fruit detection model was proposed to fully meet the specific needs of the plums in the complex orchard environments using an enhanced and lightweight version of YOLOv8s. Firstly, the backbone network (named Faster-EMA) was developed to reduce the overall complexity of the model with the high detection accuracy. The architecture of the backbone network was optimized to more effectively extract the critical features from the input images, even when the plums were occluded by branches or surrounded by other fruits. The optimal network structure was then achieved for the accurate detection with the few computational resources. Secondly, Focal Modulation was introduced to replace the spatial pyramid pooling (SPPF) module in original YOLOv8s. The multi-scale features were integrated to detect the plums at different sizes under varying environmental conditions. The semantic information was also captured during feature extraction. The key aspects of the plums were then focused (such as the shape and color), rather than the less relevant background elements. The feature fusion mechanism was significantly enhanced the overall performance of the model. Thirdly, the parameter-sharing strategy (LDetect) was introduced to implement the lightweight detection head. The parameters were shared across different branches of the detection head, in order to maintain the high detection performance while significantly reducing the number of parameters and computational complexity. This lightweight detection head was designed specifically to meet the deployment requirements of low-power embedded devices, such as edge computing platforms. The efficiency of the model was particularly advantageous for the real-time applications, indicating the rapid detection and decision making. The experiment was also validated the effectiveness of the improved model. In terms of the average detection accuracy, the mean average precision (mAP) reached an impressive 97.2%, which was a 7.4 percentage point over the baseline YOLOv8s model. Additionally, the optimal model was reduced the computational load, with a 44.8% decrease in floating point operations (FLOPs), and a 25.8% reduction in the number of model parameters. The more efficient model was achieved with the processing time and memory usage. The detection frame rate was achieved 48.3 frames per second, when the improved model was deployed on the Jetson Nano 4GB (low-power edge computing device), thus enabling real-time detection even in resource-constrained environments. In conclusion, the lightweight model can be expected to effectively detect the plums under the complex orchard backgrounds, environmental variations and occlusions. Both high accuracy and computational efficiency were achieved to incorporate the Faster-EMA, Focal Modulation, and the lightweight detection head (LDetect). The successful deployment of this model on the edge computing devices can represented the significant step toward to the intelligent plum harvesting. The finding can provide a viable solution to the automated fruit picking in dynamic orchard environments. The valuable insights can also offer for the real-time detection in agricultural robotics and precision farming.

Keywords: target detection; YOLOv8; lightweight optimization; plum; LDetect