

基于改进 YOLO v8s 的自然环境下荔枝果实检测方法

谢家兴^{1,2} 吕振东¹ 陈绍楠¹ 卢美怡¹ 高鹏¹ 王卫星³ 毛霏雨¹ 李君⁴

(1. 华南农业大学电子工程学院(人工智能学院), 广州 510642;

2. 广东省农情信息监测工程技术研究中心, 广州 510642;

3. 华南农业大学珠江学院, 广州 510900; 4. 华南农业大学工程学院, 广州 510642)

摘要: 随着智慧农业技术的推广, 针对自然环境下荔枝果实目标尺寸小、遮挡现象严重与背景复杂导致的检测精度较低以及实时检测设备的资源限制等问题, 提出了基于改进 YOLO v8s 的轻量化荔枝果实检测方法。采用轻量级网络 MobileNetV3 作为主干网络, 减小模型参数量; 引入卷积局部注意力模块 (Convolutional local attention module, CLAM), 同时在通道和空间上提高模型在自然环境下对荔枝果实特征提取能力, 并引入残差学习理念, 将通过注意力模块前后的特征进行加权融合, 保证了模型对原始图像的特征学习, 提高模型在复杂环境下的检测稳定性; 将部分卷积层替换为深度可分离卷积, 并进行多尺度特征融合; 采用 αEIoU 作为损失函数, 提高目标检测边界框纵横比收敛速度, 减小重叠果实漏检率。试验结果表明, 改进 YOLO v8s 在试验数据集上精确率和平均精度分别达到 91.75% 和 79.07%, 相较原始模型分别提升 17.29、14.75 个百分点, 同时参数量减至 5.488×10^6 , 较原始模型降低 50.7%。与主流轻量级模型 YOLO v5s、EfficientNetV2、YOLO v7-Tiny、YOLO v9s、YOLO v10s 与 YOLO 11s 相比, 本文模型在精确率、召回率和平均精度均更优, 为现代果园环境下移动设备的荔枝检测及测产提供了技术参考。

关键词: 荔枝果实检测; 自然环境; CLAM; 轻量化; YOLO v8s

中图分类号: TP391.41; S24

文献标识码: A

文章编号: 1000-1298(2026)08-0226-09

OSID:



Method for Detecting Litchi Fruits in Natural Environment Based on Lightweight Improved YOLO v8s

XIE Jiaxing^{1,2} LÜ Zhendong¹ CHEN Shaonan¹ LU Meiyi¹ GAO Peng¹

WANG Weixing³ MAO Feiyu¹ LI Jun⁴

(1. College of Electronic Engineering (College of Artificial Intelligence), South China Agricultural University, Guangzhou 510642, China

2. Engineering Research Center for Monitoring Agricultural Information of Guangdong Province, Guangzhou 510642, China

3. Zhujiang College, South China Agricultural University, Guangzhou 510900, China

4. College of Engineering, South China Agricultural University, Guangzhou 510642, China)

Abstract: With the promotion of smart agriculture technology, there are issues such as low detection accuracy and resource limitations of real-time detection devices due to the small size of litchi fruits, severe occlusion, and complex background in natural environments. A lightweight litchi fruit detection method was proposed based on the improved YOLO v8s. Firstly, a lightweight network, MobileNetV3, was adopted as the backbone network to reduce the model parameters. On this basis, the convolutional local attention module (CLAM) was introduced to enhance the model's feature extraction ability for litchi fruits in natural environments both in channels and space. The concept of residual learning was also introduced to fuse the features before and after the attention module through weighted addition, ensuring the model's feature learning from the original image and improving the detection stability in complex environments. Secondly, some convolutional layers were replaced with depthwise separable convolutions, and multi-scale feature fusion was performed. Finally, to address the issues of small fruit size and severe

收稿日期: 2024-12-27 修回日期: 2025-02-20

基金项目: 广州市科技计划重点研发项目(2025B03J0100)、广东省现代农业产业技术体系创新团队项目(2024CXTD19-11)、国家现代农业产业技术体系项目(CARS-32-09)、“百千万工程”开展新一轮农村科技特派员重点派驻任务项目(KTP20240199)和大学生创新创业训练计划项目(202410564079)

作者简介: 谢家兴(1979—),男,副教授,博士,主要从事农情信息获取与智慧农业研究,E-mail: xjx1998@scau.edu.cn

通信作者: 李君(1978—),男,教授,博士,主要从事农业电气化与自动化研究,E-mail: autojunli@scau.edu.cn

occlusion, the α EIoU was used as the loss function to accelerate the convergence speed of the aspect ratio of the detection bounding box and reduce the missed detection rate of overlapping fruits. Experimental results indicated that the improved YOLO v8s achieved an accuracy rate of 91.75% and a detection precision of 79.07% on the experimental dataset, which were 17.29 and 14.75 percentage points higher than that of the original model, respectively. At the same time, the number of parameters was significantly reduced to 5.488×10^6 , a decrease of 50.7% compared with that of the original model. Compared with mainstream lightweight models such as YOLO v5s, EfficientNetV2, YOLO v7 - Tiny, YOLO v9s, YOLO v10s, and YOLO 11s, this model demonstrated advantages in terms of precision, recall, and detection accuracy, providing a technical reference for litchi detection and yield estimation in modern orchard environments using mobile devices.

Key words: litchi fruit detection; natural environment; CLAM; lightweight; YOLO v8s

0 引言

中国荔枝种植面积、产量和产值均居世界首位, 2024 年 3 月底, 全国荔枝种植面积达到 $5.234 2 \times 10^5 \text{ hm}^2$, 其中广东种植面积为 $2.688 5 \times 10^5 \text{ hm}^2$ ^[1]。荔枝是广东最重要的经济作物之一, 但随着社会产业结构调整 and 工业化进程推进, 从事农业生产的人数持续减少, 未来农业的自动化和智能化将成为主要生产方式^[2], 以适应不断变化的农业生产需求。中国大多数地区主要依靠人工收获荔枝, 荔枝成熟的季节性强, 快速完成大规模荔枝收获需要大量劳动力。应用荔枝采摘机器人有助于解决劳动力不足和大规模种植所带来的挑战, 可以显著减少生产成本, 减轻农业人口外流对生产力的不利影响。因此, 利用计算机视觉技术对荔枝进行检测, 实现荔枝果园产量估算, 指导农户及时、选择性、分批开展采摘工作具有重要意义。其中, 荔枝检测是实现荔枝果园产量估算和智能采摘的前提^[3]。

在荔枝目标检测研究中, 传统方法依赖于荔枝果实颜色、几何形状、纹理等多种特征。郭艾侠等^[4]提出二次阈值分割识别荔枝不同部位, 熊俊涛等^[5]用 Otsu 算法、模糊 C 均值聚类算法和 Hough 圆拟合方法精准检测荔枝果实, YU 等^[6]利用深度图像的颜色与纹理训练随机森林, 完成对红、绿荔枝识别。

随着计算机性能和深度学习算法的快速发展, 通过计算机视觉和运用深度学习的方法对果实进行检测成为主流。传统果实目标识别检测是通过结合果实的纹理和颜色特征, 使用支持向量机进行二分类^[7]。但这种方法特征提取过程复杂耗时, 单纯的纹理和颜色特征易受噪声干扰, 且其自适性差、稳定性不足^[8], 并不适合在现代化果园进行实时检测。相较于传统识别方法, 基于深度学习的目标检测方法显著特点在于其多层次特征表达^[9], 这种方法对模型特征提取更完善、准确度更高, 解决了传统方法

不能满足现代化果园实时检测的需求。基于深度学习的主流目标检测算法根据有无候选框生成阶段分为双阶段目标检测算法和单阶段目标检测算法 2 类^[10]。双阶段目标检测算法虽通常在准确性上表现更好, 但由于算法检测速率较低, 难以达到实时检测要求^[11], 因此在自然环境对果实的检测中更倾向于使用单阶段目标检测算法, 如 YOLO 系列、SSD 系列与 EfficientDet^[12] 等, 且在番茄^[13]、冬枣^[14]、苹果^[15]、柑橘^[8]、猕猴桃^[16] 等上已有一定研究。而在荔枝果实的单阶段目标检测研究中, XIE 等^[17]提出了改进的荔枝检测模型 YOLO v5 - Litchi, 其增加了注意力机制、更改损失函数以及替换了主干模型, 实现了平均交并比和平均像素精度分别为 90.28% 和 94.95%, 帧率为 19.83 f/s 的试验效果。LIANG 等^[18]基于 YOLO v3 对夜间自然环境中的荔枝果实进行检测, 然后根据荔枝果实的边界框确定果茎的感兴趣区, 最后基于 U - Net 对果梗进行逐个分割, 从而实现夜间荔枝果实和果梗检测。基于深度学习的荔枝果实检测研究虽已取得一定成果, 但由于复杂环境下的不同光照、遮挡以及重叠等因素, 再加上嵌入式设备上的资源限制, 要求网络轻量化, 距离实现现代果园中的高精度实时检测, 仍面临不少挑战。

为实现在嵌入式设备等移动端上的荔枝果实目标检测, 本文提出一种基于 YOLO v8s 的轻量化改进方法, 通过采集不同数量、不同光照、不同成熟度以及不同距离条件下的荔枝图像构建荔枝数据集进行训练, 以实现在减小参数量的同时提高模型检测精度, 以期在现代果园环境下荔枝智能采摘装备提供技术参考。

1 试验材料

1.1 图像数据采集

主要在 2 个不同试验果园(华南农业大学热带亚热带特色果树资源圃中的荔枝标准果园和广东省荔枝文化博览园)进行荔枝数据采集, 采集时间为

2023 年 5—7 月和 2024 年 5—7 月。采集设备包括 iphone 14 型手机和大疆 Mavic Air2 型无人机。由于不同品种与不同成熟度的荔枝果实有一定颜色、纹理与尺寸差距,而不同距离与不同光照影响模型能提取到的具体纹理与整体外形两大特征。为提高模型对荔枝果实不同尺寸、不同颜色、整体外形和表面纹理的特征学习,提高模型对荔枝检测的稳定性和实际应用场景的适用性,采集包括各类主流荔枝品种,如桂味、糯米糍、妃子笑等在不同成熟度、不同距离以及不同光照情况下的图像,在 2 个试验果园共采集荔枝图像 1 981 幅。随后根据荔枝成熟度,将图像分为未成熟青果与已成熟红果 2 类;根据图像拍摄距离分为近景(小于 1.5 m)与远景(大于 1.5 m) 2 类;根据光照情况分为顺光、侧光与背光 3 类。最后为使数据集涵盖不同程度的遮挡和重叠情况,根据单幅图像中荔枝数量将图像分为较少(小于 10 个果实)、中等(10 ~ 50 个果实)和较多(大于 50 个果实)3 类。荔枝果实图像数据集组成如表 1 所示,其中部分图像如图 1 所示。

表 1 荔枝果实数据集组成
Tab.1 Composition of litchi fruit dataset

分类依据	分类情况	数量/幅
拍摄距离	远	960
	近	1 021
果实成熟度	青果	645
	红果	1 336
光照情况	顺光	1 094
	侧光	468
	背光	419
果实数量	较少	813
	中等	742
	较多	426



图 1 部分荔枝图像示例

Fig.1 Some sample images of litchi

1.2 数据集制作

使用 Labelimg 对采集数据进行手动标注,标注格式为 YOLO。其中未成熟荔枝青果标签数为 6

841 个,成熟荔枝红果标签数量为 15 983 个,共计标签 22 824 个。采集不同成熟度、不同拍摄距离、不同光照情况与不同数量的荔枝果实图像构成数据集,使得训练后模型可在荔枝种植的各时期、各检测距离以及各种光照情况下实现实时检测计数,方便果园数字化管理。得到标签后,对数据集按比例 8:1:1 划分为训练集、测试集、验证集,随后放入模型中训练,且在训练中使用了 YOLO v8 的在线数据增强功能,通过随机变换生成不同视角、尺寸、亮度的新样本,提高了模型对不同尺寸的适应性,并且缓解了红青果和不同光照情况下图像数据不均衡的问题。

2 模型改进

2.1 YOLO v8s 目标检测算法

YOLO v8 是 Ultralytics 公司推出的 YOLO(You only look once)系列^[19]的第 8 个版本,基本上继续沿用 YOLO v5^[20]的网络框架。

YOLO v8 在 Backbone 上同样使用 CSPDarknet 架构,但对 CSPDarknet 进行了改进,用 C2f 模块替代了 C3 模块。C2f 模块借鉴 YOLO v7^[21]的 ELAN 思想,通过并行更多的梯度流分支,在保证轻量化的同时获得更加丰富的梯度流信息。YOLO v8 引入 SPPF 结构,其速度与 SPP 相同,但感受野更优,且从原先的耦合头变成了解耦头,将分类和检测头分离,并行提取类别特征和位置特征,然后各用 1 层 1×1 卷积完成分类和定位任务。这种设计提高了目标检测准确性,并加快了网络收敛速度。YOLO v8 系列提供了 Nano、Small、Meduim、Large 和 Extra Large 等多种尺度模型,以满足不同硬件平台和应用场景需求。考虑到需兼顾针对小目标的高精度检测与移动采摘设备的资源限制问题,选择使用 Small 版本,即 YOLO v8s。

2.2 YOLO v8s 算法改进

2.2.1 CLAM 注意力机制

原始 YOLO v8s 在自然环境中检测荔枝果实面临的挑战主要为荔枝尺寸较小与背景复杂导致的特征提取困难,以及荔枝果实数量较多导致模型计算资源消耗较大。为解决这些问题,在主干网络中引入卷积局部注意力模块(CLAM)结构。

CLAM^[22]由 2 个主要的子模块组成,分别为卷积通道注意力模块(Convolutional channel attention module,CCAM)和局部空间注意力模块(Local spatial attention module,LSAM)。CLAM 结构如图 2 所示。

CCAM 通过全局平均池化和全局最大池化捕捉输入特征图中每个通道的全局空间特征。2 种池化

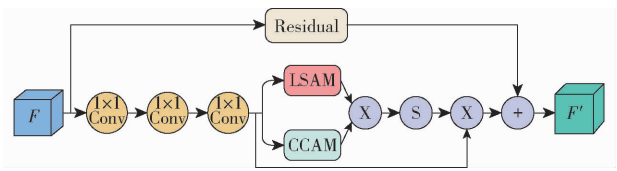


图 2 CLAM 结构图

Fig. 2 CLAM structure diagram

方式分别提供了不同的角度来观察特征,平均池化关注整体的平均趋势,而最大池化关注最显著特征。LSAM 则关注输入特征图中空间上相邻像素之间的相关性。通过局部最大池化和局部平均池化,LSAM 能够捕捉到特征图在局部邻域内的空间分布情况。通过并行地应用 CCAM 和 LSAM,CLAM 能够同时获取通道和空间 2 个维度的注意力,从而使得模型能够更加关注对于检测任务最为关键的特征和区域。

此外,在自然环境实际检测中,往往会受到光照、遮挡等多种环境因素影响,如果完全依赖注意力机制,模型可能对这些变化更为敏感,从而导致检测不稳定。因此,在 CLAM 机制中引入残差学习的思想,将注意力机制前后的特征按照不同的权值进行加权相加。通过引入残差学习,在获得通过注意力机制得到的特征的同时,保留了原始图像特征信息,特别是在复杂背景中,保持对荔枝果整体结构理解,提高了模型稳定性。

2.2.2 Focal- α EIoU

针对自然环境下荔枝果实检测中由于背景样本过于丰富导致荔枝作为正样本与作为负样本的背景不平衡问题,引入 Focal Loss^[23] 的交叉熵损失函数。Focal Loss 公式为

$$FL(P_i) = -\alpha_i(1 - P_i)^\gamma \ln(P_i) \tag{1}$$

式中 P_i ——模型预测概率(对于正类为 P_i ,对于负类为 $1 - P_i$)

α_i ——类别平衡因子,平衡正负样本权重

γ ——聚焦参数,减少易分类样本权重

Focal Loss 为难以分类的样本赋予更高权重,从而使模型在训练过程更关注那些被错误分类或难以检测的荔枝果实。这种关注能力提升,有助于提高模型在荔枝果实检测中的精确率和召回率,特别是在存在遮挡或光照变化情况下,模型能够更好地捕捉到果实特征。

此外,还将损失函数改进为 α EIoU,与 Focal Loss 加权融合。考虑到在荔枝检测任务中,果实通常尺寸不一,且重叠严重。传统的 IoU 损失只关注预测框与真实框重叠部分,可能会因为重叠区域较大而导致梯度较小,影响训练效果。EIoU^[24] 通过引入边框距离的惩罚机制,使得模型在预测框与

真实框中心距离较远时,会受到更大的损失,从而迫使模型在训练过程中更快速地调整预测框位置,减少定位误差,且更准确地判断重叠目标的边界和位置,减少误检和漏检情况,对于提高模型在复杂环境下的精确定位能力与重叠目标的检测性能尤其重要。在此基础上,将 IoU 项加上 α 次幂($\alpha = 2$)进一步加强重叠区域的优化。由于 IoU 在 0~1 之间变化,通过对 IoU 进行平方处理,可以动态调整 IoU 损失对定位误差的影响,特别是那些预测框与真实框有一定偏差的情况。当模型因少量重叠而漏检时,对 IoU 的平方处理会使得 EIoU 损失函数的值更大,使得模型更快速加大边界框位置的调整力度,缩小果实与背景之间的定位误差;而当模型严重重叠导致漏检,IoU 在接近 1 时进行平方处理,对原 EIoU 损失函数的影响较小,但会使得模型对这种漏检更为敏感。这种机制帮助模型在自然环境的荔枝果实检测中精确定位荔枝果实边界。 α EIoU 计算式为

$$\alpha EIoU = 1 - \left[IoU^\alpha - \left(\frac{p^2}{c^2} + \frac{p_w^2}{c_w^2} + \frac{p_h^2}{c_h^2} \right) \right] \tag{2}$$

- 式中 IoU ——预测框和真实框交并比
 α ——控制 IoU 权重的参数
 p ——预测框和真实框中心点之间距离
 c ——最小外接矩形对角线长度
 p_w, p_h ——预测框和真实框宽度、高度差
 c_w, c_h ——最小外接矩形宽度、高度

2.2.3 MobileNetV3 主干网络改进

将 YOLO v8s 主干网络替换为 MobileNetV3,降低模型参数量和提高检测效率。MobileNeV3 作为一种轻量级网络^[25],具有较少的参数量和较低的计算成本,能够在保证模型性能的同时大幅提升检测速度,有效解决了模型计算开销过大的问题,在嵌入式或移动设备等资源受限的环境下尤为关键。

MobileNetV3 能够加快检测速度的主要原因在于其轻量级的架构设计,特别是 MobileNetV3 特有的 Bneck 结构,结构如图 3 所示。

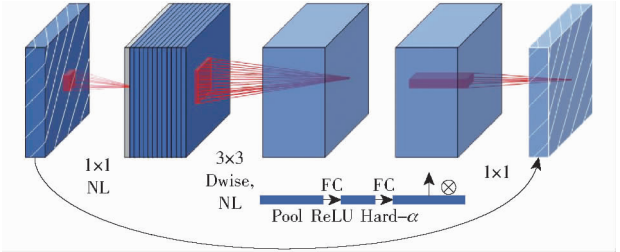


图 3 Bneck 结构

Fig. 3 Bneck structure

输入特征经过 1×1 卷积进行升维,接着,使用 3×3 深度可分离卷积进行特征提取。这种卷积将

标准卷积分解为 2 个独立操作:深度卷积和逐点卷积。深度卷积负责对每个输入通道进行独立的卷积操作,而逐点卷积则通过 1×1 卷积整合深度卷积的输出。在深度可分离卷积之后,引入 SE 模块调整每个通道权重。这个模块首先对特征图进行全局平均池化,得到每个通道的统计信息;然后,通过 2 个全连接层(或 1×1 卷积层)学习每个通道的权重;最后将这些权重应用原始特征图的对应通道上,以增强重要特征并抑制不重要特征。经 SE 模块处理,特征图再次通过 1×1 卷积进行降维,以匹配下一层的输入通道数。这样,MobileNetV3 大幅减少了计算量和参数量,相比于传统卷积网络,能在保持相似精度的情况下显著提高速度。MobileNetV3 - Large 结构参数如表 2 所示。

将主干网络替换为 MobileNetV3,并在主干网络的特征提取后,通过 CLAM 注意力机制以及深度可分离卷积(Depthwise separable convolution, DSC)^[26]进行多尺度特征融合,有助于缓解荔枝果实检测中图像样本尺寸不一和背景复杂导致检测精度较低的问题。在特征融合阶段,深度可分离卷积可以通过逐点卷积实现跨通道的特征整合,将不同通道的特征信息进行线性组合,从而生成新的特征表示,并能

有效降低计算的复杂度。改进 YOLO v8s 整体结构如图 4 所示。

表 2 MobileNetV3 结构参数

Tab.2 MobileNetV3 network structure parameters

输入	结构, 卷积核	扩展层 的输出 通道数	输出 通道 数	SE 注 意力机 制模块	非线性 激活 函数	步长
$224 \times 224 \times 3$	Conv2d		16		HS	2
$112 \times 112 \times 16$	Bneck, 3×3	16	16		RE	1
$112 \times 112 \times 16$	Bneck, 3×3	64	24		RE	2
$56 \times 56 \times 24$	Bneck, 3×3	72	24		RE	1
$56 \times 56 \times 24$	Bneck, 5×5	72	40	√	RE	2
$28 \times 28 \times 40$	Bneck, 5×5	120	40	√	RE	1
$28 \times 28 \times 40$	Bneck, 5×5	120	40	√	RE	1
$28 \times 28 \times 40$	Bneck, 3×3	240	80		HS	2
$14 \times 14 \times 80$	Bneck, 3×3	200	80		HS	1
$14 \times 14 \times 80$	Bneck, 3×3	184	80		HS	1
$14 \times 14 \times 80$	Bneck, 3×3	184	80		HS	1
$14 \times 14 \times 80$	Bneck, 3×3	480	112	√	HS	1
$14 \times 14 \times 112$	Bneck, 3×3	672	112	√	HS	1
$14 \times 14 \times 112$	Bneck, 5×5	672	160	√	HS	2
$7 \times 7 \times 160$	Bneck, 5×5	960	160	√	HS	1
$7 \times 7 \times 160$	Bneck, 5×5	960	160		HS	1
$7 \times 7 \times 160$	Conv2d		960			

注:√表示采用此模块。

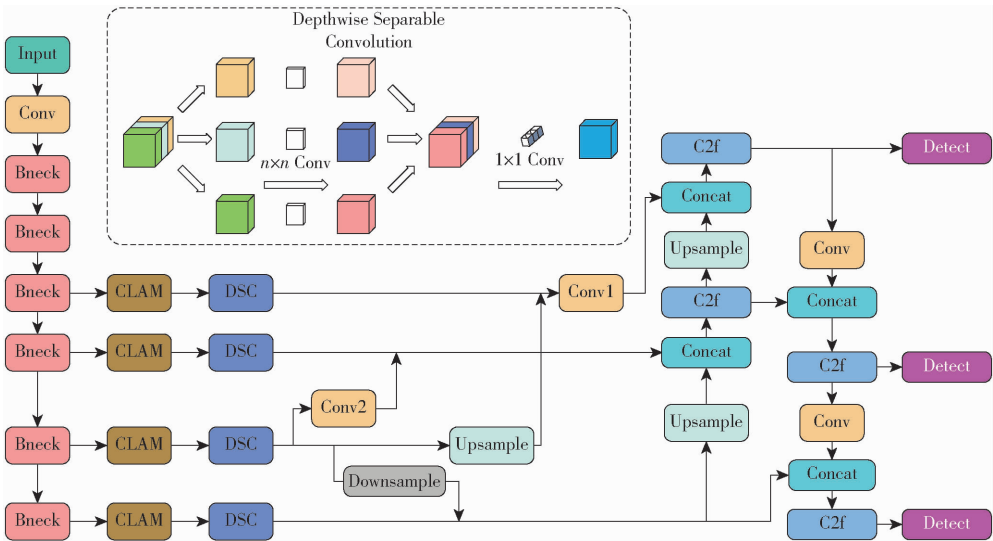


图 4 改进 YOLO v8s 结构图

Fig.4 Improved YOLO v8s network structure

3 试验与结果

3.1 试验环境

试验运行环境为 Windows 10(64 位)操作系统,训练基于 PyTorch 深度学习框架,编程语言为 Python 3.8, GPU 为 NVIDIA Geforce RTX3090, CPU 为 Intel(R) Core(TM) i7-6800K, CUDA 11.0 版本。

3.2 参数设置

将图像样本的 80% 用于训练,10% 作为验证集

帮助改进模型训练效果,剩余 10% 作为测试集对模型最终泛化能力与检测效果进行验证。并对数据集设置 50% 的概率进行 Mosaic 数据增强。在模型训练过程中,初始学习率为 0.01,配合优化器 SGD 调整学习率。考虑到图像尺寸较大且为增加模型的泛化能力,批次大小与训练轮次分别设置为 4、100。训练中设置 IoU 为 0.5,预测时根据训练所得 F1 分数曲线将分类阈值(Threshold)设为 0.3,非极大值抑制(Nonmaximum suppression, NMS)设为 0.5。

3.3 评价指标

采用平均精度 (Average precision, AP)、浮点运算量、精确率 (Precision)、召回率 (Recall)、参数量、检测帧数 (Frames per second, FPS) 对算法进行评估。

3.4 结果与分析

3.4.1 不同模型试验结果

按照模型评估指标对部分主流网络进行训练结果对比与分析,结果如表 3 所示。

表 3 不同网络模型试验结果
Tab.3 Experimental results from different network models

模型	精确率/%	召回率/%	平均精度/%	检测帧率/(f·s ⁻¹)	参数量	浮点运算量
YOLO v8s	74.46	61.80	64.32	46.53	1.114 × 10 ⁷	1.790 × 10 ¹¹
YOLO v5s	81.59	35.39	49.60	55.90	7.064 × 10 ⁶	6.591 × 10 ¹⁰
YOLO v7 - Tiny	85.87	48.98	58.59	7.21	6.227 × 10 ⁶	1.386 × 10 ¹⁰
EfficientNetV2	77.56	17.37	21.50	17.28	3.828 × 10 ⁶	4.737 × 10 ⁹
YOLO v9s	79.80	63.94	69.30	19.23	9.743 × 10 ⁶	3.961 × 10 ¹⁰
YOLO v10s	77.30	59.72	65.70	75.74	8.067 × 10 ⁶	2.822 × 10 ¹⁰
YOLO 11s	78.71	62.50	68.20	83.33	9.428 × 10 ⁶	2.150 × 10 ¹⁰
改进 YOLO v8s	91.75	71.14	79.07	41.12	5.488 × 10 ⁶	6.818 × 10 ¹⁰

由表 3 可看出,在针对荔枝果实的检测任务中,改进 YOLO v8s 相比原模型,精确率、召回率、平均精度分别提高 17.29、9.34、14.75 个百分点。相比主流轻量级网络 YOLO v5s、YOLO v7 - Tiny、EfficientNetV2^[27]、YOLO v9s^[28]、YOLO v10s^[29] 与 YOLO 11s^[30],改进 YOLO v8s 精确率上有不错的优势;在召回率与平均精度上,改进 YOLO v8s 更优,体现了改进网络在荔枝果实检测中对漏检错检问题的良好改善。轻量化改进模型不仅检测精

确率更优,更重要的是参数量与浮点运算量相比原模型大量减少,且检测帧率没有较大幅度下降,可更好满足现代化农业中智能采摘装备的设备要求。

3.4.2 不同光照情况下检测结果

为验证在自然环境下不同光照情况对荔枝果实检测的影响,在采集的原始图像中,选取荔枝红果与青果在顺光、侧光、背光 3 种情况下的图像各 100 幅进行测试,结果如表 4 所示。

表 4 不同光照情况下检测结果
Tab.4 Detection results under different lighting conditions

光照情况	模型	荔枝果实类型	实际果实数量/个	检测数目/个	检测误差/个	平均精度/%	精确率/%	召回率/%
顺光	YOLO v8s	红果	1 395	1 060	335	69.62	77.18	65.03
		青果	2 061	1 497	564	65.23	74.79	62.30
	改进 YOLO v 8s	红果	1 395	1 296	99	80.17	92.63	72.61
		青果	2 061	1 845	216	78.97	91.82	71.55
侧光	YOLO v8s	红果	2 084	1 434	650	66.65	75.09	63.29
		青果	2 073	1 288	785	65.91	74.37	62.91
	改进 YOLO v8s	红果	2 084	1 755	329	75.44	93.41	69.60
		青果	2 073	1 694	379	74.31	93.48	69.14
背光	YOLO v8s	红果	2 176	1 622	554	68.07	76.63	63.81
		青果	2 263	1 647	616	67.51	75.74	62.22
	改进 YOLO v8s	红果	2 176	1 896	280	75.23	93.79	70.50
		青果	2 263	1 958	305	75.26	93.73	71.04

由表 4 可知,不同光照情况对自然环境下荔枝果实检测影响较大。在顺光中,模型能够更好地根据荔枝纹理、颜色等特征进行检测,因此平均精度最高,检测性能最好。在侧光和背光光照情况下,模型更多地只能通过提取荔枝整体形状特征进行检测,检测性能也明显降低。而青果颜色与检测时背景颜

色相近,易造成漏检,影响荔枝果实检测精度。此对比试验也体现了不同光照情况、不同果实类型的图像对模型训练的重要意义,能够有效提高模型检测泛化性与稳定性。

通过模型改进前后检测效果对比可以看到,改进 YOLO v8s 在不同光照情况下,对红果和青果的

检测性能都有明显提升。

3.4.3 消融试验

为验证本文方法针对荔枝检测数据集检测的有效性,设计消融试验:原始 YOLO v8s 模型、在原模

型上仅在特征提取部分添加 CLAM 注意力机制、在前两点基础上进行损失函数的改进、在上述基础上替换主干网络(即改进模型)。结果如表 5 所示。

由表 5 可知,YOLO v8s 模型对荔枝果实的检测

表 5 不同改进模块消融试验结果

Tab.5 Different improvement modules ablation experiment results

CLAM 模块	α EIOU 损失函数	MobileNetV3 特征融合网络	精确率/%	召回率/%	平均精度/ %	检测帧率/ ($f \cdot s^{-1}$)	参数量	浮点运算量
×	×	×	74.46	61.80	64.32	46.53	1.114×10^7	1.790×10^{11}
√	×	×	80.91	65.36	72.54	25.66	1.151×10^7	1.876×10^{11}
√	√	×	84.81	69.70	75.27	25.84	1.151×10^7	1.876×10^{11}
√	√	√	91.75	71.14	79.07	41.12	5.488×10^6	6.818×10^{10}

注:√表示采用此模块,×表示不采用此模块。

平均精度仅为 64.32%,且浮点运算量与参数量都较大。针对荔枝果实检测任务中模型特征提取不充分的问题,加入 CLAM 注意力机制,使得模型精确率和平均精度都有明显提升。随后损失函数的改进帮助模型更聚焦于荔枝位置的准确性,有效提高了模型对重叠荔枝的检测精度。本文方法准确率、检测精度均有明显提升,精确率、召回率、平均精度达到 91.75%、71.14%、79.07%,与 YOLO v8s 模型相比,分别提高 17.29、9.34、14.75 个百分点;参数量

和浮点运算量也有明显下降,为 5.488×10^6 和 6.818×10^{10} ,与 YOLO v8s 模型相比,分别降低 50.7%和 61.9%。模型改进前后对荔枝果实的检测对比如图 5 所示,其中红框为模型预测框,蓝框为原始模型漏检错检的荔枝果实。由图 5 可看到,YOLO v8s 模型对荔枝这种小目标的检测效果存在众多漏检,尤其是与背景颜色接近的青果更为严重,在遮挡、重叠部分难以检测。改进 YOLO v8s 模型不仅参数量有明显减少,对重叠荔枝、遮挡荔枝及青



图 5 部分荔枝检测效果
Fig.5 Partial litchi detection effect

果的检测准确率均有明显提升。

4 结论

(1)针对自然环境下荔枝果实目标检测问题,提出一种改进 YOLO v8s 的轻量化检测方法。采用轻量级网络 MobileNetV3 作为主干网络,减小模型参数量;在此基础上,引入 CLAM 和残差学习的理念,提高模型在自然环境下对荔枝果实的特征提取能力,以及在复杂环境下检测的稳定性;将部分卷积层替换为深度可分离卷积,进行多尺度特征融合;采用 α EIoU 作为损失函数,提高目标检测边界框纵横

比的收敛速度,减小重叠果实的漏检率。

(2)在荔枝果实检测任务中,改进 YOLO v8s 模型检测帧中达到 41.2 f/s,精确率、召回率、平均精度达到 91.75%、71.14%、79.07%,与 YOLO v8s 模型相比,分别提高 17.29、9.34、14.75 个百分点,在保障模型检测速度的同时提升精度。

(3)改进 YOLO v8s 模型参数量和浮点运算量为 5.488×10^6 和 6.818×10^{10} ,与 YOLO v8s 模型相比,分别降低 50.7% 和 61.9%,在显著降低计算资源需求的同时,仍保持高精度实时检测能力,为荔枝果实智能化识别提供了轻量级、可部署的解决方案。

参 考 文 献

[1] 陈厚彬,杨胜男,苏钻贤,等. 2024 年全国荔枝生产形势分析与管理建议[J]. 中国热带农业, 2024(3): 8-20.
CHEN Houbin, YANG Shengnan, SU Zuanxian, et al. Analysis of the national litchi production in 2024 and management suggestions[J]. Chinese Tropical Agriculture, 2024(3): 8-20. (in Chinese)

[2] ZHOU Hongyu, WANG Xing, AU W, et al. Intelligent robots for fruit harvesting: recent developments and future challenges[J]. Precision Agriculture, 2022, 23(5): 1-52.

[3] 李景顺,刘美,陈嘉升,等. 基于深度学习的荔枝识别技术研究进展[J]. 中南农业科技, 2024, 45(8): 238-246.

[4] 郭艾侠,邹湘军,邹海鑫. 基于二次阈值的荔枝果与结果母枝图像识别[J]. 计算机工程与设计, 2014, 35(2): 557-561.
GUO Aixia, ZOU Xiangjun, ZOU Haixin. Recognition on image of litchi fruits and their main fruit bearing based on twin-threshold method[J]. Computer Engineering and Design, 2014, 35(2): 557-561. (in Chinese)

[5] 熊俊涛,林睿,刘振,等. 夜间自然环境下荔枝采摘机器人识别技术[J]. 农业机械学报, 2017, 48(11): 28-34.
XIONG Juntao, LIN Rui, LIU Zhen, et al. Visual technology of picking robot to detect litchi at nighttime under natural environment[J]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(11): 28-34. (in Chinese)

[6] YU Lianyi, XIONG Juntao, FANG Xueqing, et al. A litchi fruit recognition method in a natural environment using RGB-D images[J]. Biosystems Engineering, 2021, 204: 50-63.

[7] 宋怀波,尚钰莹,何东健. 果实目标深度学习识别技术研究进展[J]. 农业机械学报, 2023, 54(1): 1-19.
SONG Huaibo, SHANG Yuying, HE Dongjian. Review on deep learning technology for fruit target recognition[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(1): 1-19. (in Chinese)

[8] 毕松,高峰,陈俊文,等. 基于深度卷积神经网络的柑橘目标识别方法[J]. 农业机械学报, 2019, 50(5): 181-186.
BI Song, GAO Feng, CHEN Junwen, et al. Detection method of citrus based on deep convolution neural network[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(5): 181-186. (in Chinese)

[9] 张政旭,庞为光,谢文静,等. 面向实时应用的深度学习研究综述[J]. 软件学报, 2020, 31(9): 2654-2677.
ZHANG Zhengkui, PANG Weiguang, XIE Wenjing, et al. Deep learning for real-time applications: a survey[J]. Journal of Software, 2020, 31(9): 2654-2677. (in Chinese)

[10] 刘颖,刘红燕,范九伦,等. 基于深度学习的小目标检测研究与应用综述[J]. 电子学报, 2020, 48(3): 590-601.
LIU Ying, LIU Hongyan, FAN Jiulun, et al. A survey of research and application of small object detection based on deep learnin[J]. Acta Electronica Sinica, 2020, 48(3): 590-601. (in Chinese)

[11] 董文轩,梁宏涛,刘国柱,等. 深度卷积应用于目标检测算法综述[J]. 计算机科学与探索, 2022, 16(5): 1025-1042.
DONG Wenxuan, LINAG Hongtao, LIU Guozhu, et al. Review of deep convolution applied to target detection algorithms[J]. Computer Science and Exploration, 2022, 16(5): 1025-1042. (in Chinese)

[12] TAN M, PANG R, QUOC V L, et al. EfficientDet: scalable and efficient object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020: 10778-10787.

[13] 何斌,张亦博,龚健林,等. 基于改进 YOLO v5 的夜间温室番茄果实快速识别[J]. 农业机械学报, 2022, 53(5): 201-208.
HE Bin, ZHANG Yibo, GONG Jianlin, et al. Rapid identification of tomato fruits in greenhouse at night based on improved YOLO v5[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(5): 201-208. (in Chinese)

[14] 刘天真,滕桂法,苑迎春,等. 基于改进 YOLO v3 的自然场景下冬枣果实识别方法[J]. 农业机械学报, 2021, 52(5): 17-25.
LIU Tianzhen, TENG Guifa, YUAN Yingchun, et al. Winter jujube fruit recognition method based on improved YOLO v3 under natural scene[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(5): 17-25. (in Chinese)

[15] LIU Jingfan, LIU Zhaobing. YOLO v5s-BC: an improved YOLO v5s-based method for real-time apple detection[J]. Journal of Real-Time Image Processing, 2024, 21(3): 88.

[16] 穆龙涛,高宗斌,崔永杰,等. 基于改进 AlexNet 的广域复杂环境下遮挡猕猴桃目标识别[J]. 农业机械学报, 2019, 50(10): 24-34.
MU Longtao, GAO Zongbin, CUI Yongjie, et al. Kiwifruit detection of far-view and occluded fruit based on improved AlexNet[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(10): 24-34. (in Chinese)

[17] XIE Jiaxing, PENG Jiajun, WANG Jiaxin, et al. Litchi detection in a complex natural environment using the YOLO v5-Litchi model[J]. Agronomy, 2022, 12(12): 3054.

[18] LIANG Cuixiao, XIONG Juntao, ZHENG Zhenhui, et al. A visual detection method for nighttime litchi fruits and fruiting stems[J]. Computers & Electronics in Agriculture, 2020, 169: 105192.

- [19] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016:779–788.
- [20] LIU Yu, LI Xiukun, FAN Yiming, et al. Classification of peanut pod rot based on improved YOLO v5s[J]. *Frontiers in Plant Science*, 2024, 15:1364185.
- [21] YANG Zhengqiang, XIE Ruonan, LIU Linyue, et al. Dense – YOLO v7: improved real-time insulator detection framework based on YOLO v7[J]. *International Journal of Low-Carbon Technologies*, 2024, 19: 157–170.
- [22] CHEN Sibao, WEI Qingsong, WANG Wenzhong, et al. Remote sensing scene classification via multi-branch local attention network[J]. *IEEE Transactions on Image Processing*, 2021, 30: 99–109.
- [23] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[J]. *Proceedings of the IEEE International Conference on Computer Vision*, 2017:2980–2988.
- [24] ZHANG Yifan, REN Weiqiang, ZHANG Zhang, et al. Focal and efficient IOU loss for accurate bounding box regression[J]. *Neurocomputing*, 2022, 506: 146–157.
- [25] ANDREW H, MARK S, GRACE C, et al. Searching for MobileNetV3[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019:1314–1324.
- [26] FRANÇOIS C. Xception: deep learning with depthwise separable convolutions[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2017:1800–1807.
- [27] TAN M, LE Q. EfficientNetV2: smaller models and faster training[C]//International Conference on Machine Learning. PMLR, 2021:10096–10106.
- [28] 蒙永合, 雷振华, 王振东, 等. 基于改进 YOLO v9s 的变电设施渗漏油高效检测及模型部署[J]. *物联网技术*, 2025, 15(19): 132–135, 139.
MENG Yonghe, LEI Zhenhua, WANG Zhendong, et al. Efficient detection and model deployment of leakage oil in power facilities based on improved YOLO v9s[J]. *Internet of Things Technology*, 2025, 15(19): 132–135, 139. (in Chinese)
- [29] 胡文豪, 蒋亚军, 吴文龙, 等. 基于改进 YOLO v10s 的镀锌铁板缺陷检测算法研究[J]. *中国机械*, 2025(27): 50–56.
HU Wenhao, JIANG Yajun, WU Wenlong, et al. Research on defect detection algorithm for galvanized iron sheets based on improved YOLO v10s[J]. *Machine China*, 2025(27): 50–56. (in Chinese)
- [30] WANG Dingran, TAN Jiaoheng, HONG Wangcai, et al. SDS – YOLO: an improved vibratory position detection algorithm based on YOLO v11[J]. *Measurement*, 2025, 244: 116518.

(上接第 202 页)

- [31] VASWANI A. Attention is all you need[J]. *Advances in Neural Information Processing Systems*, 2017, 30: 5998–6008.
- [32] HAN K, WANG Y, CHEN H, et al. A survey on vision transformer[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 45(1): 87–110.
- [33] DU J, ZHANG Y, WANG P, et al. Enhancing winter wheat yield estimation with a CNN-transformer hybrid framework utilizing multiple remotely sensed parameters[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2025, 63: 4405213.
- [34] 崔佳旭, 杨博. 贝叶斯优化方法和应用综述[J]. *软件学报*, 2018, 29(10): 3068–3090.
CUI Jiaxu, YANG Bo. Survey on Bayesian optimization methodology and applications[J]. *Journal of Software*, 2018, 29(10): 3068–3090. (in Chinese)
- [35] 李亚茹, 张宇来, 王佳晨. 面向超参数估计的贝叶斯优化方法综述[J]. *计算机科学*, 2022, 49(6A): 86–92.
LI Yaru, ZHANG Yulai, WANG Jiachen. Survey on Bayesian optimization methods for hyper-parameter tuning[J]. *Computer Science*, 2022, 49(6A): 86–92. (in Chinese)
- [36] LUNDBERG S M, LEE S I. A unified approach to interpreting model predictions[J]. *Advances in Neural Information Processing Systems*, 2017, 30:4768–4777.
- [37] 王鹏新, 杜江莉, 张悦, 等. 基于遥感多参数和 CNN – Transformer 的冬小麦单产估测[J]. *农业机械学报*, 2024, 55(3): 173–182.
WANG Pengxin, DU Jiangli, ZHANG Yue, et al. Yield estimation of winter wheat based on multiple remotely sensed parameters and CNN – Transformer[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2024, 55(3): 173–182. (in Chinese)
- [38] WON S, BAE S H, KIM S T. Effects of mixed sample data augmentation on interpretability of neural networks[J]. *Neural Networks*, 2025, 190: 107611.
- [39] MA Y, CHEN S, ERMEN S, et al. Transfer learning in environmental remote sensing[J]. *Remote Sensing of Environment*, 2024, 301: 113924.
- [40] LI J, LIU X, LI Y, et al. Combining filling and fusion strategies for generating synthetic daily Landsat time series image on Google Earth Engine[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2025, 63: 5401415.
- [41] HASSIJA V, CHAMOLA V, MAHAPATRA A, et al. Interpreting black-box models: a review on explainable artificial intelligence[J]. *Cognitive Computation*, 2024, 16(1): 45–74.
- [42] MINH D, WANG H X, LI Y F, et al. Explainable artificial intelligence: a comprehensive review[J]. *Artificial Intelligence Review*, 2022, 55(5): 3503–3568.
- [43] PAUDEL D, BOOGAARD H, DE WIT A, et al. Machine learning for large-scale crop yield forecasting[J]. *Agricultural Systems*, 2021, 187: 103016.
- [44] FENG P, WANG B, LI L D, et al. Dynamic wheat yield forecasts are improved by a hybrid approach using a biophysical model and machine learning technique[J]. *Agricultural and Forest Meteorology*, 2020, 285: 107922.