

应用轻量化 FEB-YOLO 模型的 荔枝果实动态识别计数方法

李景顺^{1,2}, 刘 美², 孟亚男¹, 韩慧子³

(1. 吉林化工学院 信息与工程学院, 吉林 吉林 132022;

2. 广东石油化工学院 自动化学院, 广东 茂名 525000;

3. 香港理工大学 工程学院, 香港 999077)

摘要: 针对大场景自然环境下荔枝存在小目标、重叠和遮挡等特点, 提出一种轻量化荔枝检测模型 FEB-YOLO; 该模型基于 YOLOv8 在 C2f 模块中引入 PConv 替代部分常规卷积以实现轻量化改进, 同时融入 EMA 注意力机制提高算法的特征提取能力; 将颈部网络替换为融合 P2 特征层的 BiFPN, 增强模型对不同尺寸的跨尺度特征融合; 在回归损失函数中引入 NWD 度量, 提高模型对荔枝小目标的学习能力, 降低漏检率; 经实验测试得到 FEB-YOLO 模型的 P、R、mAP 对比原始模型分别提高 1.4%、1.6%、1.7%, 其参数量和计算量分别降低 47.3% 和 27.1%, 改进后模型占用的计算资源更少, 同时能够明显提高在复杂环境下的识别精度; 为实现果园场景下实时估计荔枝产量, 提出了一种高效的荔枝果实动态识别计数方法, 通过将 FEB-YOLO 作为 BoT-SORT 跟踪器的目标检测器, 将 FEB-YOLO 的识别输出作为 BoT-SORT 的输入, 实现动态视频序列的跟踪计数, 最后以实例验证了该方法的有效性和可行性; 所得改进模型具有较好的鲁棒性且体积小, 可以嵌入到边缘设备中, 不仅可用于实时估计荔枝产量, 还可用于规划采摘和贮藏, 为果园资源分配提供可靠支撑。

关键词: 荔枝果实; 多目标跟踪; 产量预测; 轻量化; 目标检测

Dynamic Recognition and Counting Method for Litchi Fruit Using Lightweight FEB-YOLO Model

LI Jingshun^{1,2}, LIU Mei², MENG Yanan¹, HAN Huizi³

(1. School of Information and Control Engineering, Jilin Institute of Chemical Technology, Jilin 132022, China;

2. School of Automation, Guangdong Institute of Petrochemical Technology, Maoming 525000, China;

3. School of Engineering, The Hong Kong Polytechnic University, Hongkong 999077, China)

Abstract: In response to the characteristics of small targets, overlap, and occlusion of litchi in large natural environments, a lightweight litchi detection model FEB-YOLO is proposed. Based on YOLOv8, partial convolution (PConv) is introduced in the module C2f to replace some conventional convolutions to achieve lightweight improvement, and efficient multi-scale attention (EMA) attention mechanism is integrated to improve the feature extraction capability of the algorithm. The neck network is replaced by bidirectional feature pyramid network (BiFPN) with P2 feature layer to enhance the cross-scale feature fusion of different sizes. Normalized weighted distance (NWD) measurement is introduced into the regression loss function to improve the learning ability of the model for litchi small targets and reduce the missed detection rate. Experimental results show that compared with the original model, the P, R and mAP of the FEB-YOLO model are increased by 1.4%, 1.6% and 1.7%, respectively, and its Params and FLOPs are reduced by 47.3% and 27.1%, respectively. The improved

收稿日期:2024-08-14; 修回日期:2024-09-28。

基金项目:国家自然科学基金资助项目(62073091);广东省普通高校重点领域(新一代信息技术)专项(2020ZDZX3042)。

作者简介:李景顺(1998-),男,硕士研究生。

刘 美(1967-),女,博士,教授。

孟亚男(1969-),女,硕士,教授。

引用格式:李景顺,刘 美,孟亚男,等. 应用轻量化 FEB-YOLO 模型的荔枝果实动态识别计数方法[J]. 计算机测量与控制, 2025,33(2):229-237,261.

model has less computing resources, and can improve the recognition accuracy under complex conditions. In order to estimate litchi yield in real-time in orchard scenarios, an efficient dynamic recognition and counting method for litchi fruit is proposed. By using the FEB-YOLO model as the target detector of the BoT-SORT tracker and the recognition output of the FEB-YOLO model as the input of the BoT-SORT tracker, the dynamic video sequence tracking and counting is realized. Finally, the effectiveness and feasibility of the proposed method are verified through examples. The improved model has the features of good robustness, small size, and embedding in edge equipment, which can not only be used for real-time estimation of litchi yield, but also for planning picking and storage, providing reliable support for orchard resource allocation.

Keywords: litchi fruit; multi-target tracking; yield forecast; lightweight; object detection

0 引言

荔枝是亚热带重要的经济作物,是我国南方的特色水果之一。根据统计数据,截至 2023 年我国的荔枝种植面积已达到 790.12 万亩,总产量达到 329.43 万吨^[1]。荔枝的成熟季节通常在 5 月至 7 月,长时间暴露在高温下的荔枝会导致荔枝酸的产生,从而引起变质并影响水果的整体风味。提前预测果园荔枝产量至关重要,这种预测有助于果农有效地规划采摘和贮藏计划,并使他们能够及早制定供应战略。传统上荔枝产量的估计通过农林学家计数进行,需要专家的仔细观察以及较长的估算时间,其成本高且结果精度低。当前迫切需要建立一个能够克服上述问题的智能产量估算系统,改变目前农业生产过程中依赖专家人工感官估产判定的现状。

YOLO 系列是一种先进的单阶段目标检测算法,其检测精度高且速度快,已广泛应用于自然环境中对荔枝果实的检测与识别。文献 [2] 改进 YOLOv3-tiny 的检测头,同时采用 K-means 聚类算法对荔枝进行空间分区,并调整柱梳装置采收果实,该模型的召回率为 78.9%,精确率为 87.4%, F_1 值为 0.83。文献 [3] 利用 YOLOv5 深度卷积网络提取荔枝主茎的特征,提出一种识别主茎采摘点的方法,该模型的精确率为 84.4%,召回率为 75.2%, $AP@50$ 值为 72.5%。文献 [4] 在 C3 模块中添加块注意力机制,并采用加权框融合作为预测框的融合方法,构建 YOLOv5 荔枝检测模型,该模型的召回率为 81.1%,平均精度为 87.1%,检测速度为 25 ms。以 YOLO 为代表的目标检测算法在检测精度和速度上表现优异,可以通过修改网络结构进一步降低模型的复杂度,使之适合部署在计算能力受限的边缘设备上,其在农业领域目标检测的成功应用为荔枝产业朝着智慧化发展提供了借鉴。

目标检测负责识别图像中的对象并给出其位置,通常在每一帧中进行,而实时追踪则是在视频流中持续跟踪这些对象,利用先前检测到的位置和运动信息来预测当前对象的位置,两者结合可以实现更高效的监控与分析。BoT-SORT^[5]是目前最流行的多目标跟踪算法,通常与目标检测算法结合进行跟踪计数,已被广泛应用于

畜类养殖^[6]、车辆跟踪^[7]、品质分级^[8]等。文献 [9] 利用 YOLOv5s 与 BoT-SORT 相结合,对稻穗的发育状态进行实时跟踪,该方法为管理稻穗的合理灌溉和施肥提供了技术支撑。

荔枝在自然环境中具有随机分布、重叠和遮挡等非结构化特点,同时检测荔枝面临小目标和近景色等挑战。综合考虑这些因素,本研究提出一种改进的轻量化荔枝果实检测模型,该模型在提高识别精度和检测速度的同时,具有较小的体积,进而能节约设备资源便于部署。此外,将改进模型应用于无人机拍摄的视频序列对检测到的荔枝果实进行跟踪计数,并将动态视频的检测结果进一步可视化,为预测荔枝果实的产量作铺垫。

1 材料与方法

1.1 荔枝图像采集

本研究使用的荔枝果实图像在中国广东省茂名市荔枝现代农业产业园采集,采集时间为 2023 年 6 月 10 日至 25 日每天上午 10 点至下午 6 点。采集设备使用手持智能手机和 DJI mavic3 classic 航拍无人机。由于自然环境背景的复杂性和荔枝果实形状的多样性,采集过程中选择了不同的荔枝颜色、光线条件、拍摄角度、遮挡类别、拍摄距离等,进一步增加目标数据多样性。共采集并保存为 JPG 格式的荔枝图像 837 张。图 1 展示了部分采集的荔枝图像。

1.2 荔枝图像数据集构建

使用 Labelimg 软件对荔枝图像样本进行标注,绘制荔枝果实的最小外接矩形,外接矩形的类别属性命名为“litchi”,按照 VOC2007 的数据格式创建 TXT 格式的标签文件,包含矩形框的高度、宽度及类别信息。标注的 TXT 文件名称和图像名称保持一致。

为了增强数据集的多样性,提高模型泛化能力,对采集得到的荔枝图像数据进行数据增强操作。使用 Python 中的 OpenCV 库对原始数据集进行扩充,采用镜像、旋转、加噪声、调整亮度等方式组合,来生成各种复杂环境背景的荔枝图像。扩增后得到荔枝图像数量为 3 348 张,对应的 TXT 标签文件同时进行相应的转换,对数据集按照 8 : 1 : 1 的比例划分为训练集、验证集和测试集。

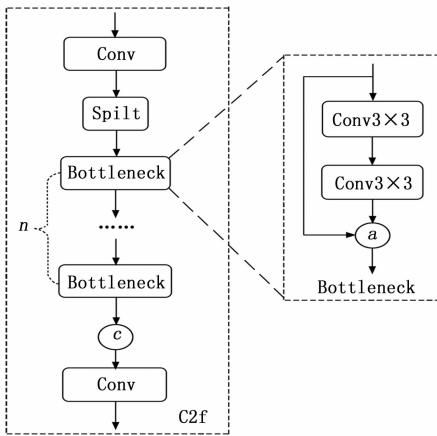


图 3 C2f 结构

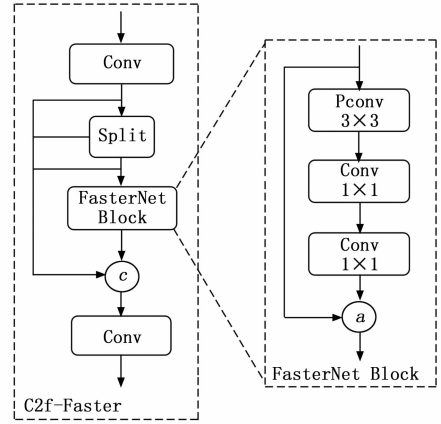


图 4 C2f-Faster 结构

计思路,在 C2f 模块的特征提取处用类继承的方法将 Bottleneck 替换为 FasterNet Block,设计 C2f-Faster 模块。由于引入 PConv 替换了部分常规卷积,缩减卷积过程中的浮点数和计算量,实现模型的轻量化设计和快速检测。C2f-Faster 模块结构如图 4 所示。

EMA 注意力机制^[14]以保留通道上的信息并减少计算成本为目的,将部分通道转换为批量维度,同时为了使每个特征组中的空间语义特征分布均匀,将通道维度分为多个子特征组;采用一种新型跨空间学习策略,通过聚合两个并行分支的输出特征,实现像素级成对关系的捕获,从而提升对像素级细节的关注;与有限接受域形成的渐进式行为相反,使用并行 3×3 和 1×1 卷积在中间特征中利用更多的上下文信息。在本研究中,为减少冗余信息的干扰,提高网络对荔枝果实特征的關注,

在 FasterNet Block 的前向传播过程中融入 EMA 模块,进而构建 C2f-Faster-EMA 模块,实现荔枝空间特征的精确提取。C2f-Faster-EMA 模块结构如图 5 所示。

本研究在原有 YOLOv8 网络的基础上,采用 C2f-Faster-EMA 模块替换骨干网络中的 C2f 模块。这一改进在前向传播过程中加入了注意力机制,使特征提取过程侧重于提高小目标检测的精度上,同时减少了骨干网络特征提取过程中的浮点运算次数。注意力机制的加入增加了模型的计算量,但有效地提高了对图像特征的提取。因此,对于 YOLOv8 颈部的 C2f 模块,没有增加 EMA 模块,而是使用 C2f-Faster 模块进行替换。

1.3.2 融合 P2 特征层的 BiFPN

BiFPN^[15]在传统的特征金字塔网络(FPN, feature pyramid network)的基础上进行了改进,该网络认为

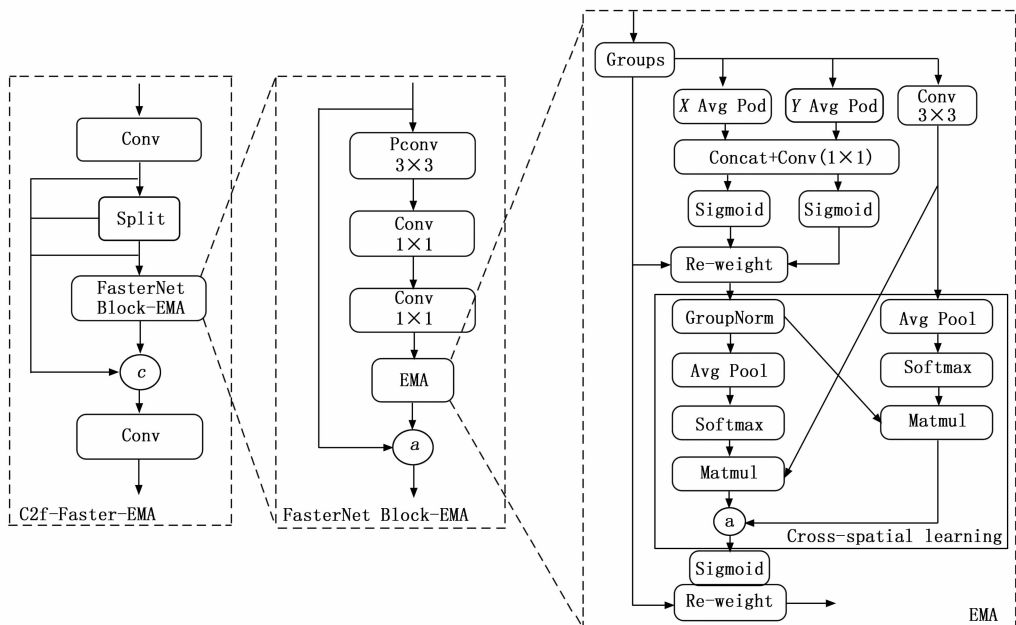


图 5 C2f-Faster-EMA 结构

高级和低级特征融合对输出特征的贡献是不相等的, 因此在特征融合中加入了权值系数。同时为了解决 PANet 中特征融合不足的问题, 提出了跨层多尺度融合, 并额外添加一条路径连接同级网络输入节点与输出节点以融合更多的特征。原始 BiFPN 结构如图 6 (a) 所示。

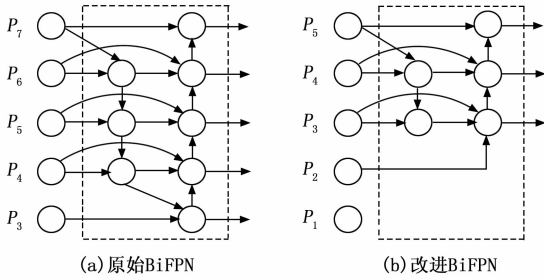


图 6 BiFPN 结构

特征融合过程中, 因为各种特征具备不同的分辨率, 它们对输出的影响也各不相同。为此, BiFPN 采用快速归一化融合模块平衡不同特征权重, 对每个输入添加一个额外的权重, 允许网络自行学习每个输入特征的重要性, 其输入与输出的关系式为:

$$O = \sum_i \frac{\omega_i}{\epsilon + \sum_j \omega_j} \cdot I_i \quad (2)$$

其中: ω_i 为输入特征 I_i 对应的学习权重, 在每个 ω_i 后添加 ReLU 激活函数使 $\omega_i \geq 0$, 设置 $\epsilon = 0.0001$ 来防止数值不稳定。通过双向跨尺度连接和快速归一化融合, 形成 BiFPN。以特征层 P_4 为例:

$$P_4^{\text{td}} = \text{Conv} \left[\frac{\omega_1 \cdot P_4^{\text{in}} + \omega_2 \cdot \text{Resize}(P_5^{\text{in}})}{\omega_1 + \omega_2 + \epsilon} \right] \quad (3)$$

$$P_4^{\text{out}} = \text{Conv} \left[\frac{\omega'_1 \cdot P_4^{\text{in}} + \omega'_2 \cdot P_4^{\text{td}} + \omega'_3 \cdot \text{Resize}(P_3^{\text{out}})}{\omega'_1 + \omega'_2 + \omega'_3 + \epsilon} \right] \quad (4)$$

其中: P_4^{td} 是自上而下路径第 4 级的中间特征, P_4^{in} 是从主干中提取的第 4 层特征, P_4^{out} 是自下而上路径第 4 级的输出特征, Resize 为上采样或下采样操作。同时为了进一步提高特征融合效率, 这里的 Conv 使用深度可分离卷积, 并且每个卷积后都添加了 BN (batch normalization) 层和激活函数。所有其他特征层都以类似的方式构建。

在大场景荔枝果实目标检测中, 图像视野的荔枝目标多且本身目标较小, 通常包含未成熟荔枝和成熟荔枝。为了识别这两个目标, 在不显著增加参数数量的情况下, 本研究将 P2 浅层特征层加入 BiFPN 中, 从而使模型获得分辨率更高的感受野特征融合, 以获取更准确的荔枝位置信息。改进 BiFPN 结构如图 6 (b) 所示。

1.3.3 NWD

文献 [16] 提出了 NWD 的新指标来度量小目标边

界框之间的相似性, 它在小目标检测方面的性能优于常规交并比 (IoU, intersection over union)。该方法将边界框建模为二维高斯分布, 中心像素具有最高的权重, 从中心向边缘的像素权重逐渐缩减, 然后计算边界框与真实值的相似度, 并进行归一化处理, 从而得到 NWD, 如式 (5) 和式 (6) 所示:

$$\omega_2^2(N_a, N_b) = \left\| \left[\left(cx_a, cy_a, \frac{\omega_a}{2}, \frac{h_a}{2} \right)^T, \left(cx_b, cy_b, \frac{\omega_b}{2}, \frac{h_b}{2} \right)^T \right] \right\|_2^2 \quad (5)$$

$$\text{NWD}(N_a, N_b) = \exp \left[-\frac{\sqrt{W_2^2(N_a, N_b)}}{C} \right] \quad (6)$$

其中: $\omega_2^2(N_a, N_b)$ 是两个边界框之间的 Wasserstein 距离, C 是数据集中的类别数, $W_2^2(N_a, N_b)$ 是一个距离测量值, N_a 和 N_b 是由 $A = (cx_a, cy_a, \omega_a, h_a)$ 和 $B = (cx_b, cy_b, \omega_b, h_b)$ 建模的高斯分布, cx, cy, ω, h 分别是边界框的坐标、宽度和高度, $\text{NWD}(N_a, N_b)$ 是经过指数形式归一化后的高斯 Wasserstein 距离。

NWD 度量可以很容易地嵌入到标签分配、非最大抑制和基于锚框的检测器的损失函数中, 以取代原始的 IoU 度量。一些研究将 NWD 度量应用于非最大抑制和损失函数中, 表明 NWD 可以有效地增强模型对小目标物体的检测性能^[17]。本研究通过在 YOLOv8 的回归损失函数中引入 NWD 度量来增强模型检测荔枝小目标的能力, 如式 (7) 和式 (8) 所示:

$$\text{Loss} = \omega_1 \cdot \text{Loss}_{\text{Objectness}} + \omega_2 \cdot \text{Loss}_{\text{Classification}} + \omega_3 \cdot \text{Loss}_{\text{Regression}} \quad (7)$$

$$\text{Loss}_{\text{Regression}} = \alpha_1 \cdot \text{CIoU} + \alpha_2 \cdot \text{NWD} \quad (8)$$

其中: ω 为各损失的权重, $\text{Loss}_{\text{Regression}}$ 由不同比例系数 α 的 CIoU 和 NWD 组成。

1.4 BoT-SORT 跟踪器

目前, 大多数多目标跟踪算法都采用高分边界框匹配的策略来识别目标, 低分数边界框被直接丢弃, 导致大量目标信息缺失。而 BoT-SORT 结合摄像机运动补偿、目标外观信息以及卡尔曼滤波器状态向量, 通过关联每个检测边界框来跟踪多个目标, 大大提高了跟踪过程中的整体性能。BoT-SORT 的核心思想是在经典 SORT 的基础上, 加入了深度学习的 Re-ID 特征增强, 并通过调优卡尔曼滤波器和轨迹管理策略来提升跟踪效果。算法的操作流程: 首先, 通过目标检测模型检测出每一帧中的目标, 得到每个目标的边界框和置信度。接着, 使用匈牙利算法将这些边界框与之前帧的目标进行关联, 确保每个目标在不同帧之间的匹配。然后, 利用卡尔曼滤波器对每个目标的状态进行预测和更新, 以提高跟踪精度。最后, 算法输出当前帧的目标轨迹和 ID, 实现对目标的连续跟踪。

目标检测网络与跟踪算法的结合可以实现多目标跟

踪功能。图 7 展示了目标检测器与 BoT-SORT 跟踪器进行对象跟踪的原理。

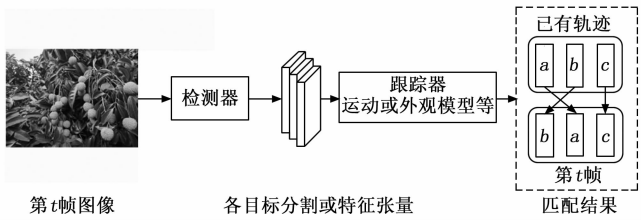


图 7 物体跟踪原理

2 实验设计与结果可视化

2.1 实验环境及评估标准

本研究实验平台的硬件配置：CPU 为 Intel (R) Xeon (R) Platinum8383C CPU@ 2.70 GHz 3.60 GHz，128 GB 运行内存；GPU 为 NVIDIA GeForce RTX 4090，24 GB 显存。软件环境：Windows 10 操作系统，Python 3.8.17 版本，Pytorch 1.12.1 框架，CUDA 12.2 版本。

使用精确率 (P, precision)、召回率 (R, recall)、平均精度均值 (mAP, mean average precision)、参数量 (Params)、计算量 (FLOPs)、每秒传输帧数 (FPS, frames per second)、 F_1 值 (F_1 Score) 作为评估模型在荔枝数据集上的检测性能指标。

2.2 训练结果及分析

将训练集 2 678 张图像输入 YOLOv8s 和 FEB-YOLO 进行训练，共进行 200 次迭代，batch-size 设为 8，比较位置损失值的收敛情况，对比结果曲线为图 8 所示。由图 8 可知，改进后模型的收敛速度比原始模型更快，说明本研究所提出的改进方法不仅能降低模型复杂度，提高检测性能，而且在收敛速度方面也具有明显优势。

为进一步验证本研究提出 3 种改进方法的有效性，对改进模型进行消融试验，将测试集 335 张图像依次输入模型进行试验，试验结果如表 1 所示。由表 1 可知，试验 2 在 YOLOv8s 的主干网络中使用 C2f-Faster-EMA 模块替换原 C2f 模块，与原模型相比，识别精确率和平均精度均值分别提高 0.8% 和 1.0%，表明 EMA 能够提高模型的特征提取能力，使其更聚焦到荔枝果实目标容易被忽略的特征信息，模型的召回率略有下降，其原因在于仅有部分通道进行卷积操作，有效减小模型权重

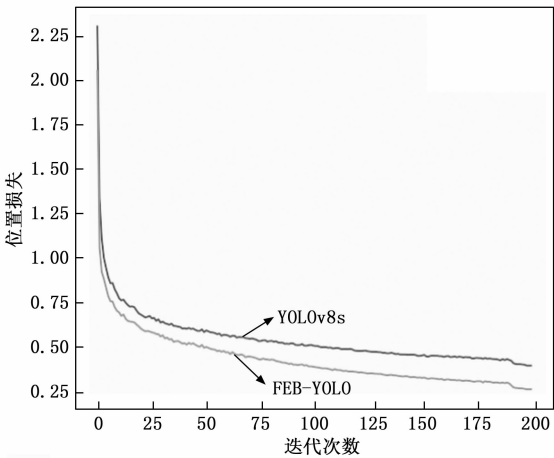


图 8 YOLOv8s 和 FEB-YOLO 训练的损失收敛对比曲线

和计算冗余的同时，丢失剩余通道可能含有的少量特征，导致召回率略有减小；试验 3 替换为融合 P2 特征层的 BiFPN 后，模型精确率和平均精度均值分别增加了 1.1% 和 0.8%，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荔枝目标分配到了更多权重；试验 4 在回归损失函数中加入 NWD 后，模型的召回率提升了 1.9%，表明 NWD 可以提高模型的拟合程度，增强模型检测小目标的能力；试验 5 同时使用 C2f-Faster-EMA 和 BiFPN 后，与原模型相比，模型精确率、召回率、平均精度均值均有提升，同时模型参数量和计算量分别降低了 47.3% 和 27.1%，FPS 提升 4.5，证明了轻量化改进的有效性；试验 6 在试验 5 的基础上增加了 NWD，结合多个模块的优势最终模型识别精确率为 89.3%，召回率为 85.5%，平均精度均值为 91.9%，与原 YOLOv8s 相比，分别提升 1.4%、1.6%、1.7%。最终模型的检测速度 FPS 为 38.9，大于满足实时性需求的 FPS 数值 24，改进模型在提升检测速度的同时，检测精度更高，模型的参数量和计算量更低，说明改进后的模型具有更好的模型性能优势。

2.3 模型改进前后检测效果对比

为了更加直观地观察改进后的模型对于解决荔枝果实重叠和遮挡问题的有效性，本试验抽取了 3 张有重叠和遮挡问题的图像，使用 Grad-CAM 绘制热力图进行可视化，通过热力图中某一区域的亮度深浅可以表现出

表 1 以 YOLOv8s 为基础模型的消融试验

序号	C2f-Faster-EMA	BiFPN	NWD	$P/\%$	$R/\%$	$mAP/\%$	$Params$	$FLOPs$	FPS
1				87.9	83.7	90.2	1.11×10^7	2.84×10^{10}	34.1
2	✓			88.7	83.3	91.2	9.70×10^6	2.46×10^{10}	35.8
3		✓		89.0	83.3	91.0	7.29×10^6	2.46×10^{10}	40.2
4			✓	87.4	85.4	90.7	1.11×10^7	2.84×10^{10}	35.2
5	✓	✓		88.7	85.1	92.1	5.86×10^6	2.07×10^{10}	38.6
6	✓	✓	✓	89.3	85.3	91.9	5.86×10^6	2.07×10^{10}	38.9

检测网络在图像中分配的权重大小。对比结果如图 9 所示。



图 9 模型热力图可视化

图 9 (b) 为 YOLOv8s 模型的热力图输出, 图 9 (c) 行为改进后模型的热力图输出。与图 9 (b) 行相比, 图 9 (c) 的荔枝目标处颜色更亮, 响应程度更高, 错误提取的叶片特征处亮度变低。这从可视化的角度进一步表明 FEB-YOLO 可以关注到更多相关区域的细节, 并将注意力集中在未被遮挡的关键部分上, 改进后的模型聚焦了更突出的荔枝目标特征, 具有更为丰富的特征提取和语义理解能力。

2.4 检测模型对比试验

将 FEB-YOLO 与 Faster R-CNN、SSD、YOLOv5s、YOLOv7-tiny、YOLOv8m 等主流目标检测网络模型进

行对比, 将测试集 335 张图像依次输入模型进行试验, 试验结果如表 2 所示。

表 2 不同模型荔枝检测结果对比

模型	$P/\%$	$R/\%$	$mAP/\%$	$Params$	$FLOPs$	$F_1/\%$
Faster R-CNN	47.6	54.8	49.7	1.37×10^8	3.70×10^{11}	50.9
SSD	80.2	32.8	45.3	2.63×10^7	6.27×10^{10}	46.5
YOLOv5s	86.3	84.1	89.5	7.01×10^6	1.58×10^{10}	85.2
YOLOv7-tiny	87.7	84.3	91.0	6.01×10^6	1.30×10^{10}	86.0
YOLOv8m	89.3	84.7	91.7	2.58×10^7	7.87×10^{10}	86.9
FEB-YOLO	89.3	85.3	91.9	5.86×10^6	2.07×10^{10}	87.3

从表 2 中得出, 对于荔枝果实目标检测场景, 一般的卷积神经网络 Faster R-CNN、SSD 的检测效果较差。相比于 YOLOv8m 模型, 改进后的模型召回率提升了 0.6 个百分点, 同时其参数量和计算量分别降低了 77.3% 和 73.7%。与 YOLOv5s、YOLOv7-tiny 轻量化模型相比, 虽然 FEB-YOLO 模型的计算量略高, 但是其参数量最小, 精确率、召回率、平均精度均值均大于这两种轻量化模型, 其中平均精度均值比 YOLOv5s、YOLOv7-tiny 分别高出 2.4%、0.9%, 同时模型参数量相较于 YOLOv5s、YOLOv7-tiny 分别减少了 16.4%、2.5%。结果表明改进后的 FEB-YOLO 相比于其他主流网络有较为明显的优势, 在具有最高的 F_1 值的同时, 减少了一定的计算量和参数量, 消耗的内存资源较少, 更适用于荔枝产量估计工作场景中的荔枝果实目标检测任务。

2.5 检测效果可视化

2.5.1 静态图像识别计数

本试验在测试集上抽取 3 张图像来对比 YOLOv8s、YOLOv8s+C2f-Faster-EMA、YOLOv8s+NWD、FEB-YOLO 的识别效果, 以验证改进模型的有效性。对比结果如图 10 所示, 图像的左上角显示计数结果。

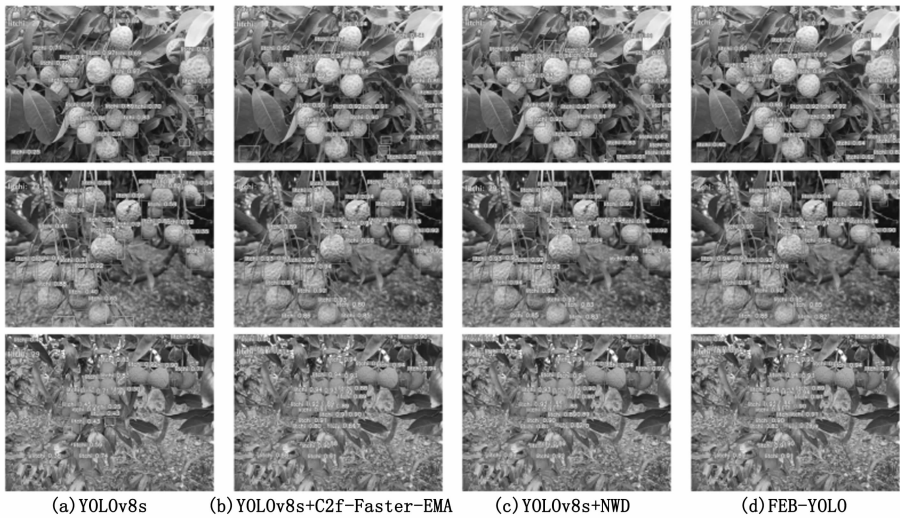


图 10 静态识别结果可视化

由图 10 所示,从上到下展示的 3 幅图像,荔枝果实逐渐成熟,表皮颜色由青变红。对比图 10 中第 1、2、3 行图像,由 YOLOv8s、YOLOv8s + C2f-Faster-EMA、YOLOv8s+NWD 的检测结果可知引入 NWD 后能够提升模型在荔枝目标较小情况下的检测效果。对比 YOLOv8s + NWD、YOLOv8s + C2f-Faster-EMA 和 FEB-YOLO 模型的检测结果可知,融入 EMA 注意力机制提高了模型对荔枝特征的关注,降低了背景中复杂因素对检测效果的影响。由此表明改进后的模型具有更好的识别能力,能有效改善漏检现象,更适用于严重遮挡、近景色以及小目标下的荔枝识别。

2.5.2 基于 FEB-YOLO+BoT-SORT 的动态视频识别计数

为了验证 FEB-YOLO 在实际荔枝果园场景中的检测效果,同时也为后续 FEB-YOLO 可以适应更多的应用场景,将提出的 FEB-YOLO 与 BoT-SORT 相结合,实现荔枝果实的动态识别计数。首先使用 FEB-YOLO 对无人机在荔枝果园中采集的动态视频进行检测,随后将目标检测得到的荔枝果实信息作为 BoT-SORT 的输入,跟踪器统计视频序列中的唯一 ID 并输出每帧的跟踪信息,最后返回得到跟踪计数结果。其中视频由 DJI mavic3 classic 无人机在距离树冠 1~2 m 的位置,保持相机垂直面向树冠,以约 0.15 m/s 的速度环绕果树一周进行拍摄,视频帧率为 30 帧/秒,原始分辨率为 1 080 P, H.265 编码。图 11 展示了 BoT-SORT 在视频帧中的跟踪结果,左上角的数字代表跟踪器匹配给每个果实的 ID。

从图 11 中可以观察到 4 种不同的计数情况。第一类为正确计数,例如 ID 为 221 和 263 的荔枝果实在 121~129 帧整个过程中都能被跟踪器准确追踪, ID 为 325 的荔枝果实在 121 帧时被确认,虽然移动到 124 帧时被遮挡暂时丢失了 ID,但在 129 帧时跟踪器重新关联将其识别为同一个果实,再次赋予同一个 ID 号。第二类情况为重复计数,例如 ID 为 229 的荔枝果实在 121 帧时被确认,但移动到 124 帧时,由于果实和树枝的遮挡,导致 ID 为 229 的果实丢失。随后,在 129 帧时,原先 ID 为 229 的果实再次被检测到,但由于消失时间超过了跟踪时间阈值,模型将其识别为新出现的果实, ID 被转变为 405,从而发生了 ID 切换现象,导致该果实被重复计数。视频序列中出现荔枝果实 ID 切换现象,原因可能是荔枝在外观或运动上非常类似,造成跟踪器干扰,导致 ID 切换。另外,跟踪器可能出现关联错误,将同一荔枝误认为是多个不同的荔枝, ID 由低变高,最终导致计数偏高。第三类情况是遗漏计数,例如图 11(d)(e)(f) 中圈内的荔枝果实,由于视野远其尺寸较小,在整个视频的起始到结束阶段,检测器和跟踪器都

未能检测到该果实并进行跟踪。第四类情况是错误计数,例如在 552 帧时, ID 为 1 222 的荔枝果实被确认, ID 为 1 211 和 1 212 框内的两个相邻荔枝果实被错误地识别为一个果实。随后,在 556 帧时, ID 为 1 222 的果实消失,检测模型将其与原先 ID 为 1 211 的果实识别为同一个果实, 3 个重叠果实产生了 ID 为 1 211 的果实框。当面对果实毗邻或重叠的情况,可能会由于跟踪器的漏识别导致遗漏计数,而当跟踪器将两个或多个不同的荔枝目标误认为是一个荔枝时,会出现计数错误,导致视频序列中的计数值低于实际值。

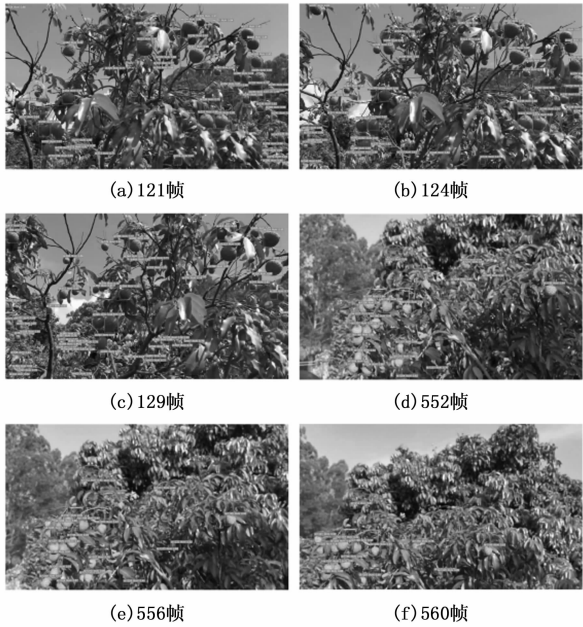


图 11 动态视频的部分跟踪结果示例

该视频最后一帧识别到的荔枝最大 ID 为 1 450,通过人工计数 3 次取平均,视频中荔枝总数约为 1 714 个,因此 FEB-YOLO + BoT-SORT 的识别计数精度为 84.6%。而利用原始的 YOLOv8s + BoT-SORT 进行识别计数时,最大 ID 是 1203,识别计数精度为 70.2%。表明改进的 FEB-YOLO 与 BoT-SORT 结合具有更高的识别计数精度,更适用于大场景环境下果园荔枝的实时检测计数。

3 讨论

对比前人改进模型的荔枝果实识别实验,基于 YOLOv8s 改进的 FEB-YOLO 模型具有更高的 P 、 R 及 mAP 。对比结果如表 3 所示。

表 3 实验结果对比

模型	$P/\%$	$R/\%$	$mAP/\%$
YOLOv3-tiny ^[2]	87.4	78.9	74.6
YOLOv5s ^[3]	84.4	75.2	72.5
YOLOv5s ^[4]	88.8	81.1	87.1
FEB-YOLO	89.3	85.3	91.9

由表3可知,相比文献[2]改进的YOLOv3-tiny网络,FEB-YOLO在 P 、 R 、 mAP 上分别高出1.9、6.4、17.3个百分点;相比文献[3]提出的基于YOLOv5s深度卷积的荔枝主茎采摘点识别算法,FEB-YOLO在 P 、 R 、 mAP 方面均具有绝对优势;相比文献[4]改进的YOLOv5s网络,FEB-YOLO的 P 、 R 、 mAP 均具有优势。由此可知,本研究提出的FEB-YOLO模型更适用于复杂环境下的荔枝识别。

在计数方面,文献[18]通过改进的YOLOv3实现了番茄静态图像的计数功能,但由于静态图像仅能捕捉番茄的局部且存在遮挡,无法实现对番茄产量的估计。文献[19]利用无人机拍摄果树周围图像并基于SIFT图像匹配构建完整的果树表面全景图,通过YOLOX对图像进行识别与计数来预测产量,该方法构建的表面全景图存在严重遮挡的情况。文献[20]对视频中的每帧图像进行统计,实现了对鸡蛋和鸡的计数功能,由于无法区分视频中重复出现的相同目标,存在严重的重复计数问题。本研究提出了一种简单高效的计数方法,通过无人机对整棵树上的荔枝进行3D拍摄,采集视频数据,然后利用目标检测网络对视频序列进行荔枝识别,最后将识别的荔枝果实信息作为跟踪器的输入,完成荔枝的识别计数。该方法能有效减少荔枝遮挡的情况,从而避免漏检或重复计数。

4 结束语

本研究针对大场景荔枝检测存在小目标、随机分布、重叠和遮挡等难题,提出了基于YOLOv8改进的轻量化荔枝目标检测模型。引入PConv替换C2f模块中的部分常规卷积进行轻量化改进,同时融入EMA注意力机制提高模型对荔枝特征的关注;将融合P2特征层的BiFPN替换原模型中的PANet,双向融合不同网络层之间的特征信息,增强模型的特征融合能力;集成NWD到回归损失函数中,在不增加模型计算复杂度的同时,改善复杂环境下小目标的错检和漏检问题。

改进后的FEB-YOLO模型在测试集上的 P 、 R 、 mAP 分别为89.3%、85.5%、91.9%,比YOLOv8s原始模型分别提高了1.4%、1.6%、1.7%。同时,改进后的模型的参数量和计算量分别为 5.86×10^6 和 2.07×10^{10} ,与原模型相比分别降低了47.3%和27.1%,表明本研究提出的改进模型检测精度更高,同时模型复杂度更低;提出的FEB-YOLO模型与主流目标检测网络Faster R-CNN、SSD、YOLOv8m相比,具有一定的优势, mAP 最高能提升46.6%。与轻量化模型YOLOv5s、YOLOv7-tiny相比,FEB-YOLO具有最小的参数量,同时 P 、 R 、 mAP 均大于这两种轻量化模型。此外,对比前人改进模型的荔枝识别实验数据可知,本研究提出的FEB-YOLO模型能在复杂环境中高

效识别荔枝果实,为未来的自动化采摘提供支撑。

为实现自然果园环境下荔枝果实的产量估计,设计了应用轻量化FEB-YOLO模型的荔枝果实动态识别计数方法。通过结合FEB-YOLO目标检测网络与BoT-SORT多目标跟踪器对果树上的荔枝进行检测和跟踪,实现基于3D视频序列的果实计数,最终得到检测跟踪的漏检率约为15.4%,比原始的YOLOv8s+BoT-SORT降低了14.4%,能够更有效地实现荔枝的识别计数。本研究提出的动态识别计数方法不仅能助力未来果园的产量预测,还可用于指导早期施肥用量及规划后续采摘活动,为果园智慧管理提供可靠支持。

参考文献:

- [1] 陈厚彬,苏钻贤,杨胜男. 2023年全国荔枝生产调查与形势分析[J]. 中国热带农业, 2023(3): 13-22.
- [2] LI C, LIN J Q, LI B Y, et al. Partition harvesting of a column-comb litchi harvester based on 3D clustering[J]. Computers and Electronics in Agriculture, 2022, 197: 106975.
- [3] QI X K, DONG J S, LAN Y B, et al. Method for identifying litchi picking position based on YOLOv5 and PSPNet[J]. Remote Sensing, 2022, 14(9): 2004.
- [4] XIE J X, PENG J J, WANG J X, et al. Litchi detection in a complex natural environment using the YOLOv5-Litchi model[J]. Agronomy, 2022, 12(12): 10253-10265.
- [5] AHARON N, ORFAIG R, BOBROVSKY B Z. BoT-SORT: Robust associations multi-pedestrian tracking[EB/OL]. (2022-07-07) [2024-05-06]. <https://arxiv.org/abs/2206.14651>.
- [6] BEWLEY A, GE Z, OTT L, et al. Simple online and real-time tracking[C] //2016 IEEE International Conference on Image Processing (ICIP). IEEE, 2016: 3464-3468.
- [7] YANG Y Q, PI D N, WANG L Y, et al. Based on improved YOLOv8 and Bot SORT surveillance video traffic statistics[J/OL]. (2024-03-27) [2024-05-06]. <https://doi.org/10.21203/rs.3.rs-4161504/v1>.
- [8] XU J J, LU Y Z. Prototyping and evaluation of a novel machine vision system for real-time, automated quality grading of sweetpotatoes[J]. Computers and Electronics in Agriculture, 2024, 219: 852-878.
- [9] ZHOU Q Y, GUO W, CHEN N, et al. Analyzing nitrogen effects on rice panicle development by panicle detection and time-series tracking[J]. Plant Phenomics, 2023, 5: 36-48.
- [10] LIU X, PENG H, ZHENG N, et al. Efficientvit: memory efficient vision transformer with cascaded group attention[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 14420-14430.

(下转第261页)