doi:10.6041/j.issn.1000-1298.2024.04.019

基于改进 COF－YOLO v8n 的油茶果静态与动态检测计数方法

王金鹏　何　萌　甄乾广　周宏平

（南京林业大学机械电子工程学院，南京 210037）

摘要：针对自然环境下油茶果存在严重遮挡、近景色、小目标等现象，使用 YOLO 网络存在检测精度低、漏检现象严重等问题，提出对 YOLO v8n 网络进行改进。首先使用 MPDIOU 作为 YOLO v8n 的损失函数，有效解决因为果实重叠导致的漏检问题；其次调整网络，向其中加入小目标检测层，使网络能够关注小目标油茶以及被树叶遮挡的油茶；最后使用 SCConv 作为特征提取网络，既能兼顾检测精度又能兼顾检测速度。改进 COF－YOLO v8n 网络精确率、召回率、平均精度均值分别达到 97.7%、97%、99%，比未改进的 YOLO v8n 分别提高 3.2、4.8、2.4 个百分点，其中严重遮挡情况下油茶检测精确率、召回率、平均精度均值分别达到 95.9%、95%、98.5%，分别比 YOLO v8n 提高 4.0、9.1、4.6 个百分点。因此改进后COF－YOLO v8n 网络能够明显提高油茶在严重遮挡、近景色、小目标均存在情况下的识别精度，减小油茶的漏检。此外，模型能够实现动、静态输入条件下油茶果计数。动态计数借鉴 DeepSORT 算法的多目标跟踪思想，将改进后 COF－YOLO v8n 的识别输出作为 DeepSORT 的输入，实现油茶果实的追踪计数。所得改进模型具有很好的鲁棒性，且模型简单可以嵌入到边缘设备中，不仅可用于指导自动化采收，还可用于果园产量估计，为果园物流分配提供可靠借鉴。

关键词：油茶果；机器视觉；COF－YOLO v8n；计数；产量估计

中图分类号：S126　　**文献标识码**：A　　**文章编号**：1000-1298(2024)04-0193-11　　**OSID**：

Camellia oleifera Fruit Static and Dynamic Detection Counting Based on Improved COF－YOLO v8n

WANG Jinpeng　HE Meng　ZHEN Qianguang　ZHOU Hongping

(*College of Mechanical and Electronic Engineering*, *Nanjing Forestry University*, *Nanjing* 210037, *China*)

Abstract: Aiming at the problems of severe occlusion, close view and small target *Camellia oleifera* in *Camellia oleifera* fruit, the original YOLO v8n network was selected to improve the phenomenon of low detection accuracy and serious missed detection phenomenon by using the original YOLO network. MPDIOU was used as the loss function of YOLO v8n to effectively solve the problem of missed detection caused by fruit overlap. Adjusting the network and adding a small target detection layer to it, so that the network can pay attention to small target *Camellia oleifera* and *Camellia oleifera* obscured by leaves; SCConv was used to replace the C2f in the original YOLO v8n, so that the network can balance both detection accuracy and detection speed. The *P*, *R* and mAP of the improved COF－YOLO v8n network reached 97.7%, 97% and 99% respectively, which were 3.2 percentages, 4.8 percentages and 2.4 percentages higher than *P*, *R* and mAP of the unimproved YOLO v8n, among which the *P*, *R* and mAP of *Camellia oleifera* reached 95.9%, 95% and 98.5% under severe occlusion, respectively, which was 4.0 percentages, 9.1 percentages and 4.6 percentages higher than that of the original YOLO v8n. The COF－YOLO v8n network can significantly improve the recognition accuracy of *Camellia oleifera* under the conditions of severe occlusion, close vie, and small targets. In addition, the model can realize the counting of *Camellia oleifera* under dynamic and static input conditions. Dynamic counting draws on the multi-target tracking idea of DeepSORT algorithm, and took the recognition output of COF－YOLO v8n as the input of DeepSORT to realize the recognition and counting of *Camellia oleifera* fruits, and used the

收稿日期：2023－08－11　修回日期：2023－09－13

基金项目：国家林业和草原局应急科技项目(202202－3)、江苏省农业科技自主创新基金项目(CX(22)3099)、江苏省现代农机装备与技术推广项目(NJ2021－18)和江苏省重点研发计划项目(BE20211016－2)

作者简介：王金鹏(1979—)，男，副教授，博士，主要从事智能化采收研究，E-mail：jpwang@ njfu.edu.cn

reduced resolution *Camellia oleifera* data to simulate the target situation in the field environment and restored the real picking environment. The resulting improved model had good robustness and simple model can be embedded in the robotic arm, which can not only be used to guide future automated harvesting, but also for yield estimation of orchards, providing reliable reference for orchard logistics distribution.

Key words: *Camellia oleifera* fruit; machine vision; COF - YOLO v8n; count; production estimates

0 引言

油茶是一种亚热带常绿灌木,在许多国家作为油料植物被广泛种植,不仅是一种重要的经济林物种,而且还具有丰富的不饱和脂肪酸[1],营养价值高。而油茶果实密度大,闭塞严重,油茶收获需要大量的人力,采收效率低下[2],因此油茶果的机械化采收势在必行。

基于机器视觉的深度学习算法被广泛应用于科学研究中,包括缺陷检测[3]、病虫害检测[4]、植物表型[5]、除草管理[6]等。目前,果蔬目标检测主要采用两种方法,一种是基于颜色、纹理[7-8]、边缘[9]等的传统图像检测和分割技术。然而,浅层特征只能在有限的场景中检测目标,这些方法虽然提供相对较快的处理速度,但是往往缺乏泛化性和鲁棒性,不适合复杂背景下油茶的识别。而使用深度学习策略可以解决复杂环境下的作物检测问题。IQBAL 等[10]使用 Inception v3 对 8 个芒果品种进行自动分类和分级,考虑到颜色、大小、形状和纹理等特征,该方法分类准确率达到 99.2% 和分级准确率 96.7%,具有良好的检测效果,且模型轻量化易于部署。WANG 等[11]使用 YOLO v7 算法识别小米辣,通过使用可变形卷积使模型轻量化同时不降低检测速度,可以满足边缘部署条件。CONG 等[12]提出一种基于 Swin Transformer 注意力机制的改进 Mask RCNN,使用 UNet3 + 提高 Mask 头和 Mask 分割质量,在叶片遮挡情况下有效分割不同类别的青椒,但是这种一阶段检测模型检测速度慢,不能满足实时性要求。

目前的研究应用中,目标检测为目标的实时追踪奠定基础,目标检测的精度决定目标追踪器的追踪效果。多目标追踪指在事先不知道目标数量的情况下,对视频中的多个目标进行检测并分配 ID 进行轨迹跟踪,其中每个目标都有不同的 ID,以便后续进行轨迹预测、精确查找等。DeepSORT 是目前最流行的多目标跟踪算法,一般与深度学习框架相结合,最先应用于人和车辆检测计数方面,目前已被广泛应用于果园产量估计[13-14]、禽类养殖[15]、动物保护[16]、船舰实时检测追踪[17]等。ZHOU 等[18]使用 YOLO v5s 模型结合 DeepSORT 实现对单个稻穗成长发育的定位追踪,为获取每棵稻穗成长发育的表型变换以及指定合理施肥策略奠定基础。

机器识别的精确度是实现自动化采收、产量估计等的重要前提。油茶果识别不仅面临颜色多变、近景色、小目标等困境,还存在生长环境复杂的问题,比如多遮挡、不均匀光照等。综合考虑这些影响因素,如何提高果园油茶的识别效果以及满足网络模型轻量化要求使之便于部署是本研究的重点。此外,为了将单幅图像的检测结果进一步可视化,以及为预测油茶果的产量提供理论支持,对 COF - YOLO v8n 模型检测到的油茶果进行计数。本文对 YOLO v8n 网络进行 3 方面的改进,使网络在复杂田间环境下具有更高的鲁棒性,且模型足够轻量使之便于实现边缘部署。

1 材料与方法

1.1 图像获取

本研究以油茶果为实验对象。油茶果实通常为球形,与叶片的形状对比明显;其品种多,存在红褐色、青色等不同颜色;果实通常粘连成长,密度大。由于油茶的外形颜色与经济林物种冬枣等相似,因此基于油茶的研究有望推广到其他种类的果实中。

油茶数据集在江苏省南京市江宁区的油茶果种植基地获取。拍照设备为 ZED - Mini 相机,采集天气分为晴天和阴天,采集时间段分别为 07:00—11:00、13:00—18:00,获取得到不同光照强度、不同颜色的油茶图像共 788 幅。图 1 为多样性数据集。

1.2 数据处理

1.2.1 数据扩充

数据扩充是在训练前对整个数据集进行预处理,目的是增加数据集的大小和多样性。使用 OpenCV 库对原始数据集进行扩充。为模仿运动抖动导致的图像模糊,对图像添加椒盐、高斯噪声;为模仿相机在一瞬间捕捉到不同角度的图像,对原始数据进行仿射、平移、翻转等变换。数据增强后共计 3 566 幅图像,使用 Pycharm 中的 LabelImg 对图像进行标注,分为严重闭塞(0,油茶果被遮挡部分大于整个果实的 1/4)和轻微闭塞或不闭塞(n,油茶果被遮挡的部分小于整个果实的 1/4)2 类,严重闭



(a) 不同光照条件下的油茶

(b) 不同颜色的油茶

图 1　多样性油茶数据集

Fig. 1　Diversity *Camellia oleifera* datasets

塞总计 6 778 个油茶标签，轻微闭塞或不闭塞有 6 989 个标签，油茶类别标签占比约为 1∶1。数据集按照 8∶2划分为训练集和验证集。图 2 为增强后图像。



(a) 原图　(b) 翻转+噪声　(c) 仿射变换+填充　(d) 平移+噪声

图 2　数据扩充

Fig. 2　Data augmentation

1.2.2　数据增强

数据在进行训练时，网络会通过随机色调增强、饱和度增强、亮度增强、水平翻转、Mosiac、Mixup 等变换生成多样化的训练样本，增加数据的多样性，提高模型的泛化能力。数据增强应用于每个训练迭代时，通过随机对训练样本进行变换，使模型每次看到的样本有所不同。表 1 为训练过程中数据增强的超参数。

表 1　训练过程中数据增强方法

Tab. 1　Data augmentation method during training

数据增强方法	图像调整程度
色调增强	0.015
饱和度增强	0.7
亮度增强	0.4
水平翻转	0.5
Mosaic 数据增强	0.9
Mixup 数据增强	0.2

1.3　YOLO v8 模型

YOLO v8 是 2023 年 Ultralytics 公司开源的新网络，集成目标检测、关键点检测、实例分割、实时追踪等任务。相比 YOLO v5，最大的改进是采用基于中心检测算法的无锚框（Anchor－Free）检测方式，直接检测物体的中心区域和边界信息，将分类和回归解耦为 2 个子网，因无需手动设计锚框从而减小了由于选择锚框数量和尺寸而导致的干扰，能有效提高目标检测的准确性，尤其对小目标更加友好，因为不同尺寸的锚框数量一般不平衡，所以可能会导致小目标的检测效果差，而 Anchor－Free 这种无锚框设计解决了这种数量不平衡问题。使用 Decoupled－Head 作为最终检测头，Decoupled－Head 包含 2 个并行的分支，每个分支都包含 2 个3×3 的卷积层，其中一个分支用于预测分类，另一个分支用于预测回归和目标置信度，将分类和回归任务进行分开预测，从而减少这两任务之间的冲突，理论上能够提高模型性能，图 3a 为 Decoupled－Head 的结构图；此外 C3 模块改为 C2f 模块，同时将上采样之前的 1×1 卷积去除，将 Backbone 不同阶段输出的特征直接输入上采样操作，C2f 综合 C3 模块以及 YOLO v7 中的 ELAN 优点，具有丰富的梯度信息流的同时还保证轻量化，图 3b 为 C2f 结构图。

因为 YOLO v8n 网络是 YOLO v8 系列中最小的网络，更易满足轻量化要求，本文模型在 YOLO v8n 的基础上进行改进。

1.4　YOLO v8n 模型改进

1.4.1　轻量化网络重建

模型轻量化方便完成边缘部署，从而利于实现后续的采摘工作。本文利用 SCConv[19]（Spatial and channel reconstruction convolution）使 COF－YOLO

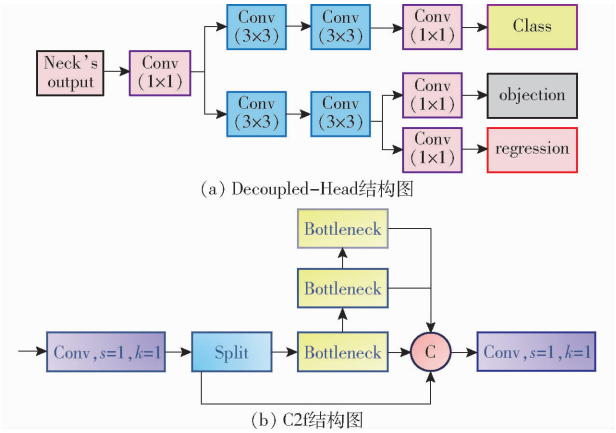


图 3 YOLO v8n 的改进部分示意图

Fig. 3 Improvement schematic of YOLO v8n

v8n 模型轻量化。SCConv 是 2023 年 LI 团队开发的一种轻量化卷积，由 SRU(Spatial reconstruction unit)和 CRU(Channel reconstruction unit)两部分串联组成，通过减少标准卷积中的空间和通道特征冗余来降低模型参数量和浮点运算量(FLOPs)。

SRU 通过分离-重建 2 部分抑制空间冗余，获得空间细化特征 X^w。图 4 为其原理图。

首先将输入特征 X 进行组归一化(GN)得到输出特征 X_{out} 为

$$X_{out} = \mathrm{GN}(X) = \gamma \frac{X - u}{\sqrt{\sigma^2 + \varepsilon}} + \beta \tag{1}$$

式中 u——X 的均值 ε——正常数

γ、β——学习参数

σ——X 的标准差

将随机初始化参数 γ 归一化后得到通道权重 W_γ 为

$$W_\gamma = \frac{\gamma_i}{\sum_{j=1}^{C} \gamma_j} \tag{2}$$

式中 i、j——索引变量

C——通道数量

将 W_γ 与 X_{out} 相乘的值映射到(0,1)之间，使用阈值进行门控。大于阈值的 W_γ 视为有用特征设置为 W_1，小于阈值的 W_γ 视为无用特征设置为 W_2。W 的计算公式为

$$W = \mathrm{Gate}(\mathrm{sigmoid}(W_\gamma(X_{out}))) \tag{3}$$

式中 Gate——门控函数

W——通过门控同时得到的富信息权重 W_1 和弱信息权重 W_2

使用交叉重建方式在空间维度上增强有用特征抑制冗余特征，从而获得空间细化特征 X^w。

$$X_1^w = W_1 \otimes X \tag{4}$$

$$X_2^w = W_2 \otimes X \tag{5}$$

$$X^{w1} = X_{11}^w \oplus X_{22}^w \tag{6}$$

$$X^{w2} = X_{21}^w \oplus X_{12}^w \tag{7}$$

$$X^w = X^{w1} \cup X^{w2} \tag{8}$$

式中 X_{11}^w、X_{21}^w——从 X_1^w、X_2^w 中进一步分离得到的丰富信息

X_{12}^w、X_{22}^w——X_1^w、X_2^w 中剩下的较不丰富的信息

X_1^w——信息量大的特征

X_2^w——信息量小的特征

X^{w1}、X^{w2}——交叉重构特征

所得空间细化特征 X^w 在通道上还是冗余的，采用 CRU 进行分割-转换-融合策略减少通道冗余。图 5 为 CRU 的实现原理图。

使用分割将特征进一步分为富特征提取器 X_{up} 和浅特征提取器 X_{low}，X_{up} 采用高效卷积运算 GWC(Group-wise convolution)结合 PWC(Group-wise convolution)替代 $K \times K$ 卷积，以较少的计算成本提取代表性特征 Y_1，X_{low} 用廉价的 1×1 PWC 操作来生成具有浅层隐藏细节的特征映射得到 Y_2，作为 X_{up} 特征提取器的补充，对 Y_1 和 Y_2 进行融合，即可得到具

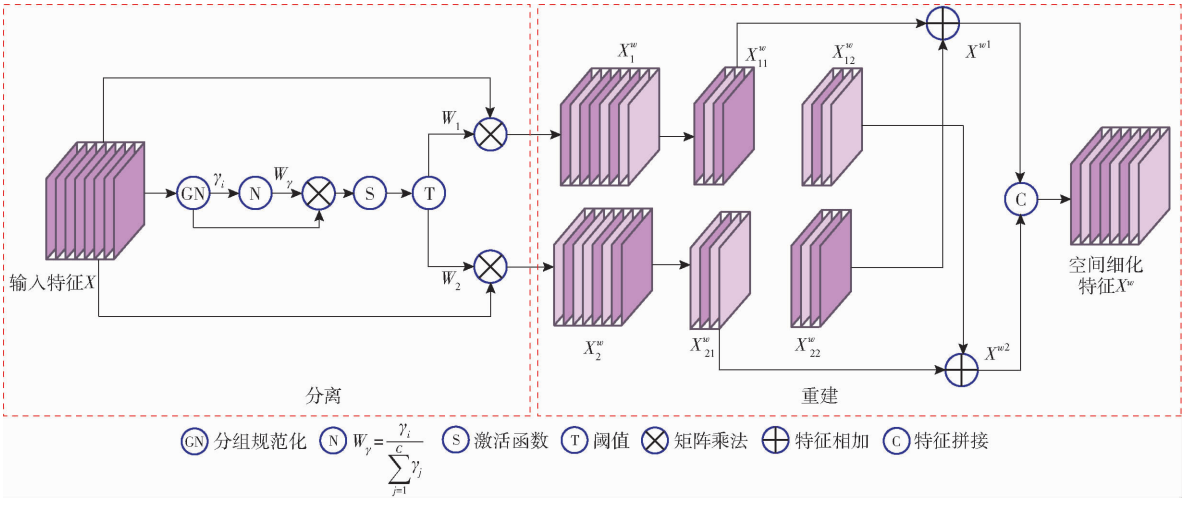


图 4 SRU 模型原理图

Fig. 4 SRU model principle diagram

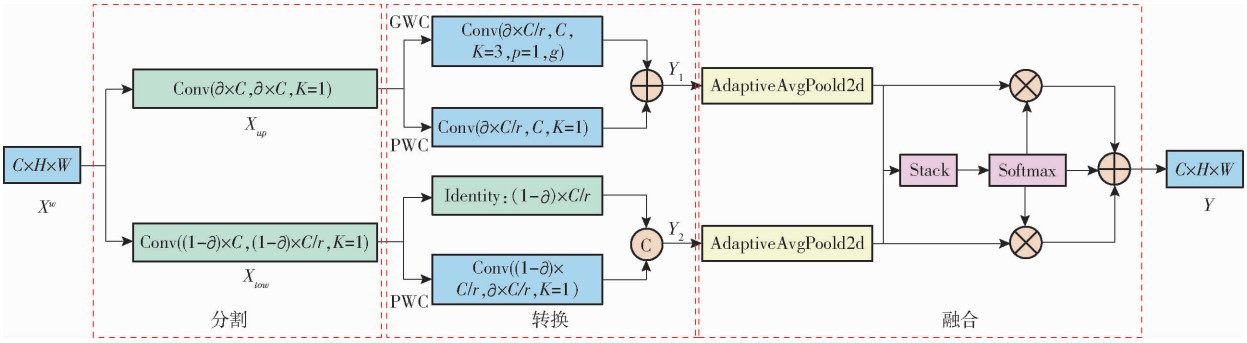


图 5　CRU 的实现原理图

Fig. 5　CRU implementation diagram

有更多有效信息的 Y。因此输入特征 X 经过 SRU 和 CRU 之后既能减少信息冗余又能提取出有效信息。

1.4.2　MPDIOU 改进

由于现有大多数损失函数在不同预测结果下具有相同的值，降低了边界框回归的收敛速度和准确性，受水平矩形的几何特征启发，SILIANG 等[20]设计一种基于最小点距离的新型损失函数，使用 MPDIOU 可以直接预测边界框与实际标注框之间的左上角和右下角点之间的距离。具体计算公式为

$$MPDIOU = \frac{A \cap B}{A \cup B} - \frac{d_1^2}{w^2 + h^2} - \frac{d_2^2}{w^2 + h^2} \tag{9}$$

其中

$$d_1^2 = (x_1^B - x_1^A)^2 + (y_1^B - y_1^A)^2 \tag{10}$$

$$d_2^2 = (x_2^B - x_2^A)^2 + (y_2^B - y_2^A)^2 \tag{11}$$

式中　A、B——两个任意图形

w、h——输入图像宽度、高度

(x_1^A, y_1^A)、(x_2^A, y_2^A)——图形 A 的左上和右下点坐标

(x_1^B, y_1^B)、(x_2^B, y_2^B)——图形 B 的左上和右下点坐标

d_1、d_2——图形 A、B 的左上角点之间的欧氏距离和右下角点之间的欧氏距离

MPDIOU 简化了两个边界框之间的相似性比较，帮助算法选择最合适的边界框来准确定位目标。使用 MPDIOU 可以有效解决果实重叠导致的检测框失真现象，有效降低油茶果漏检问题。

1.4.3　小目标检测层添加

当处理小目标油茶图像识别问题时，面临油茶直径在 15～45 mm 范围内、低像素覆盖率，以及可能被树叶、枝干和其他果实遮挡等问题。此外，输入图像尺寸对小目标油茶识别也产生了干扰，因为过大的图像会导致缩放，使小目标的识别更加困难。为了应对这些挑战，对 YOLO v8n 网络进行优化，并添加一个专门处理小目标的预测头。这个预测头将允许网络更加关注像素较小的油茶实例，从而提高小目标油茶识别准确性。当输入尺寸为 640 像素×640 像素的特征图时，YOLO v8n 输出特征图尺寸为 40 像素×40 像素、80 像素×80 像素、20 像素×20 像素，调整网络结构后，添加一个 160 像素×160 像素的小目标检测层。由于 YOLO v8n 是基于无锚框进行检测的，所以不再需要使用 K-means 算法对锚框进行重新设计。图 6 为改进 COF－YOLO v8n 模型结构图。

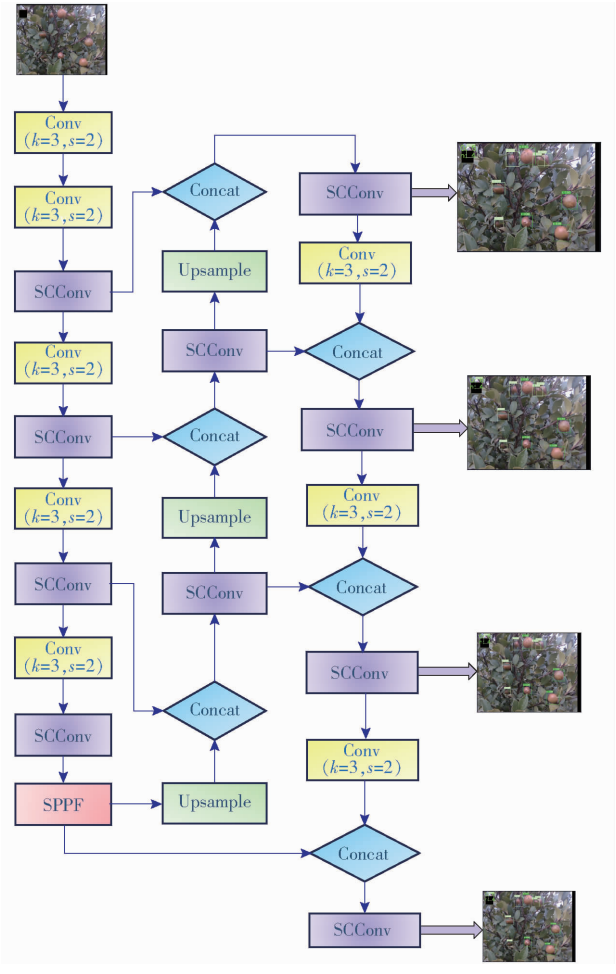


图 6　改进 COF－YOLO v8n 网络结构图

Fig. 6　Improved COF－YOLO v8n network structure diagram

1.5　DeepSORT 追踪器

DeepSORT[21]是对 SORT[22]的改进，用于目标

跟踪,是一种基于行人再识别数据集离线训练的深度学习模型。在实时目标跟踪中,提取目标的特征进行最近邻匹配,提高了遮挡情况下的目标跟踪效果。同时也减少了目标 ID 跳变的问题。DeepSORT 算法的核心思想仍然是传统的单假设跟踪方法,使用递归卡尔曼滤波和逐帧数据关联。流程操作如下:①读取当前目标检测帧的位置和各检测帧图像块的深度特征。②根据帧的置信度对帧进行滤波,去除置信度低的帧和特征。③随后对检测帧进行非极值抑制,消除一个目标的多帧。④利用卡尔曼滤波预测目标在当前帧中的位置。

追踪器一般都是和目标检测网络结合实现多目标的追踪功能。图 7 为目标检测器和 DeepSORT 结合实现物体追踪的原理图。

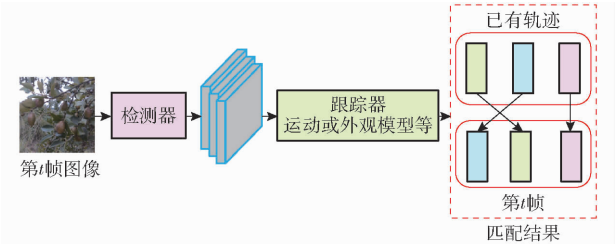


图 7 物体追踪原理图

Fig. 7 Object tracking principle diagram

1.6 网络模型训练

1.6.1 环境配置

本研究模型训练和测试均在同一环境下运行,计算机的 GPU 为 NVIDIA GeForce RTX3070Ti,CPU 为 AMD Ryzen 7 5800X 8 核处理器,操作系统为 Windows 11,使用 PyTorch 1.10、Python 3.9 和 CUDA 12.1。

1.6.2 模型评价指标

为了评估模型性能,使用精确率(Precision,P)、召回率(Recall,R)、平均精度均值(mAP)作为评价指标。

2 实验与结果可视化

2.1 轻量化网络设计

为了验证本文使用的 SCConv 的有效性,对比近两年的轻量化 SOTA(State of the art)模型,选用 Efficientformer[23]、ODConv[24](Omni-dimensional dynamic convolution)、Fasternet[25]、EfficientViT[26] 轻量化网络进行对比实验,结果如表 2 所示。

表 2 轻量化网络对比结果

Tab. 2 Comparison results of lightweight network

模型	精确率/%	召回率/%	平均精度均值/%	参数量	浮点运算量
YOLO v8n + Efficientformer	95.1	96.3	98.1	5 254 768	1.24×10^{10}
YOLO v8n + ODConv	93.2	93.6	93.8	3 074 842	5.8×10^{9}
YOLO v8n + EfficientViT	93.7	92.4	94.2	4 011 686	9.5×10^{9}
YOLO v8n + Fasternet	93.6	94.4	94.4	2 452 000	7.1×10^{9}
YOLO v8n + SCConv	95.3	96.1	97.6	2 752 278	7.3×10^{9}
YOLO v8n	94.5	92.2	96.6	3 011 238	8.2×10^{9}

由表 2 可知,基于动态卷积的 ODConv 具有比较小的参数量以及最小的浮点运算量,检测方面,召回率比 YOLO v8n 增加 1.4 个百分点,但是精确率、平均精度均值分别比 YOLO v8n 降低 1.3、2.8 个百分点;Fasternet 具有最少的参数量以及较低的浮点运算量,召回率比 YOLO v8n 增加 2.2 个百分点,但精确率、平均精度均值分别比 YOLO v8n 降低 0.9、2.2 个百分点;Efficientformer 相比于 YOLO v8n 在精确率、召回率、平均精度均值上均有提升,分别提升 0.6、4.1、1.5 个百分点,但是参数量以及计算复杂度比较大;而 SCConv 具有较少的参数量以及比较小的计算复杂度,且在精确率、召回率、平均精度均值方面的表现均比 YOLO v8n 好,分别高 0.8、3.9、1.0 个百分点。综上所述,SCConv 具有更好的表现能力,在减少模型复杂度的同时,能够提高模型的识别效果,因此利用 SCConv 替代 YOLO v8n 的 C2f 是合理的。

使用 SCConv 替换 YOLO v8n 中的 C2f,实验证明可以有效降低参数量和浮点运算量,表 3 为添加 SCConv 后的参数量以及计算量。

表 3 轻量化模型参数

Tab. 3 Parameters of lightweight models

模型	参数量	浮点运算量
YOLO v8n	3 011 238	8.20×10^{9}
YOLO v8n + SCConv	2 752 278	7.30×10^{9}
YOLO v8n + Multiple	10 863 144	3.81×10^{10}
YOLO v8n + Multiple + SCConv	9 318 904	3.36×10^{10}

注:Multiple 表示加入小目标检测层。

2.2 IOU 对比实验

YOLO v8n 默认使用的损失函数是 CIOU,将 CIOU 改为 EIOU、DIOU、MPDIOU 损失函数后,在验

证集上进行对比实验，如图 8 所示，MPDIOU 比 CIOU 在验证集上表现更好，类别损失更低。

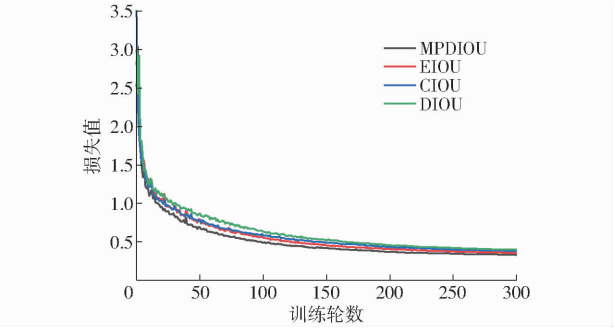


图 8　不同 IOU 在验证集上的类别损失值

Fig. 8　Class loss values for different IOU thresholds on validation set

2.3　消融实验

分别使用 MPDIOU、Multiple、SCConv 改进 YOLO v8n 网络，所得消融实验结果如表 4 所示。表中状态 all 表示油茶果未被遮挡。

表 4　消融实验结果

Tab. 4　Results of ablation experiments　%

模型	状态	精确率	召回率	平均精度均值
YOLO v8n	all	94.5	92.2	96.6
	0	91.9	85.9	93.9
	n	97.1	98.5	99.4
YOLO v8n + MPDIOU	all	95.8	96.4	98.3
	0	93.0	93.8	97.1
	n	98.5	98.9	99.4
YOLO v8n + Multiple	all	96.2	96.0	98.4
	0	93.9	94.4	98.1
	n	98.5	97.6	98.7
YOLO v8n + SCConv	all	95.3	96.1	97.6
	0	92.7	93.9	96.8
	n	97.9	98.3	98.4
YOLO v8n + MPDIOU + Multiple	all	96.0	95.5	98.3
	0	93.8	93.7	97.3
	n	98.2	97.3	99.3
YOLO v8n + MPDIOU + SCConv	all	94.2	95.8	97.9
	0	92.2	93.6	96.5
	n	96.2	98.0	99.3
YOLO v8n + Multiple + SCConv	all	97.1	96.2	98.7
	0	95.4	94.6	97.9
	n	98.8	97.8	99.5
COF－YOLO v8n	all	97.7	97.0	99.0
	0	95.9	95.0	98.5
	n	99.5	98.9	99.5

由表 4 可得，YOLO v8n 对不闭塞或轻微闭塞情况下的油茶检测效果较好，严重闭塞情况下油茶的检测效果却不理想。但当向 YOLO v8n 中加入 MPDIOU、Multiple、SCConv 后，网络整体的参数均有提升。加入 MPDIOU 后在严重闭塞时精确率、召回率、平均精度均值分别提高 1.1、7.9、3.2 个百分点；加入 Multiple 后，对于严重闭塞的精确率、召回率、平均精度均值分别提高 2.0、8.5、4.2 个百分点；当加入 MDPIOU、SCConv 后，严重闭塞情况下的精确率、召回率、平均精度均值分别提高 0.3、7.7、2.6 个百分点；同时加入 Multiple 和 SCConv 时，在严重闭塞时的精确率、召回率、平均精度均值分别提高 3.5、8.7、4 个百分点；同时加入 Multiple 和 MPDIOU 后在严重闭塞时的精确率、召回率、平均精度均值分别提高 1.9、7.8、3.4 个百分点。

2.4　YOLO 模型对比实验

为了验证 COF－YOLO v8n 的检测优势，将其与未经改进的 YOLO v5s、YOLO v7-tiny、YOLO v8n 网络进行对比，结果如表 5 所示。

表 5　YOLO 模型对比实验结果

Tab. 5　Experimental comparison results of YOLO model　%

模型	状态	精确率	召回率	平均精度均值
YOLO v5s	all	92.6	90.3	94.7
	0	87.9	87.0	90.8
	n	97.3	93.6	98.6
YOLO v7-tiny	all	95.9	94.4	96.7
	0	93.8	91.4	95.2
	n	98.0	97.4	98.2
YOLO v8n	all	94.5	92.2	96.6
	0	91.9	85.9	93.9
	n	97.1	98.5	99.4
COF－YOLO v8n	all	97.7	97.0	99.0
	0	95.9	95.0	98.5
	n	99.5	98.9	99.5

由表 5 可知，YOLO v7-tiny 相比 YOLO v5s 和 YOLO v8n 具有更高的精确率、召回率、平均精度均值，因为 YOLO v7-tiny 网络较深，能对特征进行深度提取，但 YOLO v7-tiny 在验证集上的推理时间比较长，不利于模型的部署以及实时检测。COF－YOLO v8n 在整个数据集上的精确率、召回率、平均精度均值比表现较好的 YOLO v7-tiny 分别高 1.8、2.6、2.3 个百分点，严重闭塞情况下的精确率、召回率、平均精度均值分别比 YOLO v7-tiny 高 2.1、3.6、3.3 个百分点。

由于轻微闭塞或者无闭塞情况下模型的识别效果均较好，只对严重闭塞情况下油茶进行 $P-R$ 曲线可视化，如图 9 所示。

2.5　检测结果可视化

2.5.1　静态图像识别计数

将 YOLO v8n 与 YOLO v8n＋SCConv、YOLO v8n＋MPDIOU、COF－YOLO v8n 在验证集上的检测效果

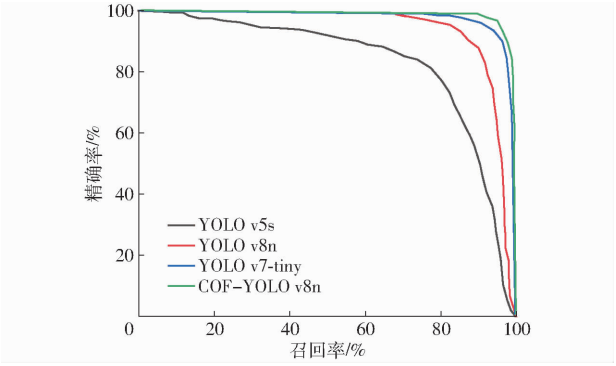
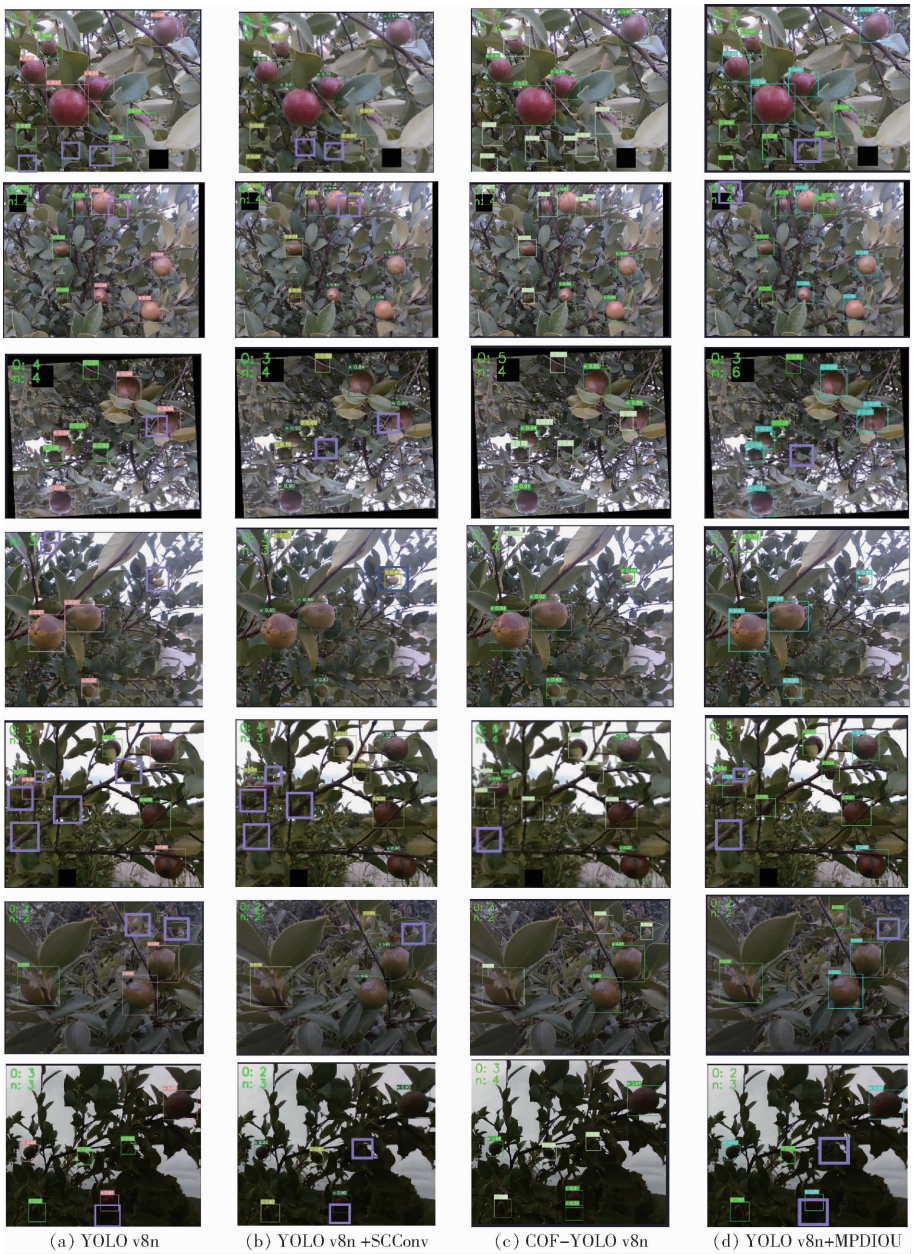


图 9 验证集在不同模型上的 $P-R$ 曲线

Fig. 9 $P-R$ curves of validation set on different models

进行对比,来验证改进模型的有效性。在检测到油茶果的同时,算法还能实现对单幅图像不同类别的油茶果实进行计数,计数结果显示在可视化图像的左上角。图 10 中紫色框表示漏检,蓝色框表示误检。

图 10 从上到下展示的 7 幅可视化图像的光照强度逐渐减弱,油茶的颜色也由深变浅。为了区分不同模型的检测结果,设置目标检测框颜色随机生成;为了防止目标框宽度太宽导致检测失真,根据特征图大小设置检测框的宽度。针对图 10 中第 2、3、4 行图像,对比 YOLO v8n、YOLO v8n + SCConv 和 YOLO v8n + MPDIOU 的可视化检测结果,可得出加入 MPDIOU 后能改善果实重叠情况下油茶果的识别效果。对比图 10 中的所有图像,比较 YOLO v8n + MPDIOU、YOLO v8n + SCConv 和 COF - YOLO v8n 模型可得,加入小目标层之后的模型对树枝、树叶遮挡、近景色以及小目标油茶具有更好的识别能力。由此表明改进后的模型能够更好地适应小目标、近



(a) YOLO v8n (b) YOLO v8n +SCConv (c) COF-YOLO v8n (d) YOLO v8n+MPDIOU

图 10 检测结果可视化

Fig. 10 Detection visualizations

景色、严重闭塞、不同光照下的油茶果，能够有效提高油茶的识别精度，有效改善漏检现象。

2.5.2　基于 COF－YOLO v8n＋DeepSORT 的动态视频识别计数

将动态视频作为 COF－YOLO v8n 的输入，再将得到的油茶果信息作为 DeepSORT 输入，每个检测到的油茶果都将被分配唯一的追踪号，并在不同的视频结果中关联相同的水果数据，得到的动态检测结果如图 11 所示。图中蓝色框表示跳动的油茶 ID，绿色框表示漏检。



图 11　COF－YOLO v8n＋DeepSORT 的油茶跟踪结果

Fig. 11　COF－YOLO v8n＋DeepSORT realized tracking of *Camellia oleifera* fruit

视频由人手持相机以速度约 0.15 m/s 拍摄获得，视频帧率（FPS）为 29 f/s，即每帧图像出现的时间约为 34.5 ms，而目标检测时间总共为 14.8 ms，目标匹配以及目标状态更新需要 13.4 ms，将 DeepSORT 追踪器设置为每间隔 10 f 抽取一幅图像，抽取的相邻 2 幅图像有 70% 左右的油茶重复，每个油茶在视频中持续存在约 345 ms，远远大于目标检测和追踪所需时间，因此能够满足输入视频的实时检测要求。但是追踪存在较明显的漏检现象，如绿色框所示。此外，同一幅图像中出现了油茶果 ID 切换现象，一方面由于目标追踪器可能会出现关联错误，将 2 个或多个不同的目标油茶误认为是同一个油茶，从而使目标从较高的 ID 跳到较低的 ID，另一方面由于部分油茶在外观或运动上非常相似，导致追踪器可能发生混淆，从而导致 ID 切换，导致最终计数偏少。视频最后一帧检测到油茶的最大 ID 是 798，而手工计数 3 次取平均值得到视频中总的油茶数目约为 987 个，即目标检测追踪精度约为 81.9%。使用 YOLO v8n＋DeepSORT 追踪到的油茶的最大 ID 为 658，目标检测追踪的精度约为 67%。因此 COF－YOLO v8n 配合 DeepSORT 能更好地实现油茶的追踪计数。

3　讨论

对比前人改进的油茶果识别实验，基于 YOLO v8n 改进的 COF－YOLO v8n 模型具有更高的精确率、召回率以及平均精度均值，表 6 为对比结果。

由表 6 可得，改进的 YOLO v8n 模型相比其它改进模型具有更高的识别精确率以及平均精度均值。相比一阶段算法 Mask R－CNN，COF－YOLO v8n 在精确率、召回率、平均精度均值方面分别提高 8.28、4.14、9.15 个百分点；相比周宏平等[2]提出的基于多源图像融合的油茶识别算法，COF－YOLO v8n 在精确率、召回率、平均精度均值方面处于绝对优势；相比宋怀波等[27]改进的 YOLO v5s 网络，COF－YOLO v8n 识别精确率以及平均精度均值均具有优势，召回率相对低 1.38 个百分点，分析原因可能是将油茶分为严重闭塞和轻微闭塞，虽然在轻微闭塞或无闭塞情况下油茶的召回率达到 98.9%，但当严重闭塞和轻微闭塞油茶的标签数量相当的情况下，油茶整体的召回率会被严重闭塞的油茶拉低。综合可得，COF－YOLO v8n 模型能够实现复杂情况下油茶的识别，能为未来自动化采收提供支持。

表 6　实验结果对比

Tab. 6　Comparison of experimental results with predecessors　%

模型	精确率	召回率	平均精度均值
YOLO v5s[27]	90.73	98.38	94.00
YOLO v5s[2]	94.80	94.60	98.40
Mask R－CNN[28]	89.42	92.86	89.85
COF－YOLO v8n	97.70	97.00	99.00

此外，在计数方面，赵春江等[29]也能实现鸡和鸡蛋计数功能，对视频中的每帧图像进行计数，但是视频图像之间存在目标重复出现，算法却并不能区分相同目标，因此会出现严重的重复计数现象；WANG 等[30]使用改进的 YOLO v3 实现静态输入的计数，但是并不能真正实现番茄产量预估，因为一幅图像只是拍摄番茄的局部且存在遮挡；ZHANG 等[31]通过无人机拍摄苹果树全景图，使用 YOLOX 对经过处理后的全景图像实现识别计数，从而估计单棵苹果树的产量，但果实闭塞情况难以避免，因此会出现产量少估现象。而本文提出的计数方法成本低，易于实现，不仅仅可以对输入的图像进行计数，

还可以实现对 3D 视频中出现的所有油茶进行实时跟踪,通过对整棵树上的油茶进行 3D 拍摄,可以解决油茶闭塞问题,避免油茶少计数或重复计数的情况,可以帮助实现果园的产量估计。

4 结论

(1)通过添加小目标检测层,使严重闭塞情况下的精确率、召回率、平均精度均值分别提高 2.0、8.5、4.2 个百分点,从验证集上的检测效果看,网络对小目标油茶果以及被树枝、树叶遮挡的油茶果具有优秀的识别能力;通过添加 MPDIOU 损失函数替代 YOLO v8n 的 CIOU 损失函数,从检测效果上看,网络对因果实重叠而导致的油茶漏检有明显的改善,很大程度上降低了漏检率;加入 MPDIOU 后相比 YOLO v8n 在严重遮挡情况下召回率提高 7.9 个百分点;加入 SCConv 之后使模型既能均衡精度又能均衡检测速度;改进后的 COF - YOLO v8n 整体的精确率、召回率、平均精度均值比 YOLO v8n 分别提高 3.2、4.8、2.4 个百分点,在严重遮挡情况下精确率、召回率、平均精度均值分别比 YOLO v8n 提高 4.0、9.1、4.6 个百分点。此外,对比前人改进模型的数据可得,本文提出的 COF - YOLO v8n 模型比前人的改进算法更能适应复杂环境。

(2)对于静态输入图像,COF - YOLO v8n 具有更强的泛化能力以及鲁棒性,且可对检测到的不同类别的油茶进行计数;对于 3D 视频帧图像,用 COF - YOLO v8n结合 DeepSORT 能够实现油茶的实时检测,最终得到检测检测器精度为 81.9%,比未经改进的 YOLO v8n + DeepSORT 更适合实时检测计数。

参 考 文 献

[1] ZHOU Y, TANG Y, ZOU X, et al. Adaptive active positioning of *Camellia oleifera* fruit picking points: classical image processing and YOLO v7 fusion algorithm[J]. Applied Siences, 2022, 12(24): 12959.

[2] 周宏平,金寿祥,周磊,等. 基于多模态图像的自然环境下油茶果识别[J]. 农业工程学报,2023,39(10):175 - 182.
ZHOU Hongping, JIN Shouxiang, ZHOU Lei, et al. Recognition of *Camellia oleifera* fruit in natural environment based on multimodal image[J]. Transactions of the CSAE,2023,39(10):175 - 182. (in Chinese)

[3] YANG Y, KANG H. An enhanced detection method of PCB defect based on improved YOLO v7[J]. Electronics, 2023, 12(9):2120.

[4] CHEN J, CHEN J, ZHANG D, et al. Using deep transfer learning for image-based plant disease identification[J]. Computers and Electronics in Agriculture, 2020, 173: 105393.

[5] LEE C, YANG M, TSENG H, et al. Single-plant broccoli growth monitoring using deep learning with UAV imagery[J]. Computers and Electronics in Agriculture, 2023, 207: 107739.

[6] RAI N, ZHANG Y, RAM B G, et al. Applications of deep learning in precision weed management: a review[J]. Computers and Electronics in Agriculture, 2023, 206: 107698.

[7] JANANI M, JEBAKUMAR R. Detection and classification of groundnut leaf nutrient level extraction in RGB images[J]. Advances in Engineering Software, 2023, 175: 103320.

[8] ALSHEHHI R, MARPU P R. Change detection using multi-scale convolutional feature maps of bi-temporal satellite high-resolution images[J]. European Journal of Remote Sensing, 2023, 56(1): 2161419.

[9] QUAN Y, LI M, HAO Y, et al. Tree species classification in a typical natural secondary forest using UAV-borne LiDAR and hyperspectral data[J]. Giscience & Remote Sensing, 2023, 60(1): 2171706.

[10] IQBAL H M R, HAKIM A. Classification and grading of harvested mangoes using convolutional neural network[J]. International Journal of Fruit Science, 2022, 22(1): 95 - 109.

[11] WANG F, JIANG J, CHEN Y, et al. Rapid detection of Yunnan Xiaomila based on lightweight YOLO v7 algorithm[J]. Frontiers in Plant Science, 2023, 14: 1200144.

[12] CONG P, LI S, ZHOU J, et al. Research on instance segmentation algorithm of greenhouse sweet pepper detection based on improved Mask RCNN[J]. Agronomy, 2023, 13(1): 196.

[13] HE L, FANG W, ZHAO G, et al. Fruit yield prediction and estimation in orchards: a state-of-the-art comprehensive review for both direct and indirect methods[J]. Computers and Electronics in Agriculture, 2022, 195: 106812.

[14] GAN H, LEE W S, ALCHANATIS V, et al. Active thermal imaging for immature citrus fruit detection[J]. Biosystems Engineering, 2020, 198: 291 - 303.

[15] JIANG K, XIE T, YAN R, et al. An attention mechanism-improved YOLO v7 object detection algorithm for hemp duck count estimation[J]. Agriculture, 2022, 12(10): 1659.

[16] CHEN X, PU H, HE Y, et al. An efficient method for monitoring birds based on object detection and multi-object tracking networks[J]. Animals, 2023, 13(10): 1713.

[17] JIANG Q, LI H. Silicon energy bulk material cargo ship detection and tracking method combining YOLO v5 and DeepSORT[J]. Energy Reports, 2023, 9: 151 - 158.

[18] ZHOU Q, 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: 0048.

[19] LI J, WEN Y, HE L. SCConv: spatial and channel reconstruction convolution for feature redundancy[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 6153－6162.
[20] SILIANG M, YONG X. MPDIoU: a loss for efficient and accurate bounding box regression[J]. arXiv preprint arXiv:2307.07662, 2023.
[21] WOJKE N, BEWLEY A, PAULUS D. Simple online and realtime tracking with a deep association metric[C]//2017 IEEE International Conference on Image Processing (ICIP). IEEE, 2017: 3645－3649.
[22] BEWLEY A, GE Z, OTT L, et al. Simple online and realtime tracking[C]//2016 IEEE International Conference on Image Processing (ICIP). IEEE, 2016: 3464－3468.
[23] LI Y, YUAN G, WEN Y, et al. Efficientformer: vision transformers at mobilenet speed[J]. Advances in Neural Information Processing Systems, 2022, 35: 12934－12949.
[24] LI C, ZHOU A, YAO A. Omni-dimensional dynamic convolution[J]. arXiv preprint arXiv:2209.07947, 2022.
[25] 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, 2023: 12021－12031.
[26] XIONG J T, LIN R, LIU Z, et al. The recognition of litchi clusters and the calculation of picking point in a nocturnal natural environment[J]. Biosystems Engineering, 2018, 166: 44－57.
[27] 宋怀波,王亚男,王云飞,等. 基于 YOLO v5s 的自然场景油茶果识别方法[J]. 农业机械学报,2022, 53(7): 234－242.
SONG Huaibo, WANG Ya'nan, WANG Yunfei, et al. Identification method of *Camellia oleifera* fruit in natural scene based on YOLO v5s[J]. Transactions of the Chinese Society for Agricultural Machinery,2022, 53(7): 234－242. (in Chinese)
[28] 王梁,侯义锋,贺杰. 基于 Mask－RCNN 的自然场景下油茶果目标识别与检测[J]. 中国农机化学报, 2022, 43(12): 148－154.
WANG Liang, HOU Yifeng, HE Jie. Target recognition and detection of *Camellia oleifera* in natural scene based on Mask－RCNN[J]. Chinese Journal of Agricultural Mechanization, 2022, 43(12): 148－154. (in Chinese)
[29] 赵春江,梁雪文,于合龙,等. 基于改进 YOLO v7 的笼养鸡/蛋自动识别与计数方法[J]. 农业机械学报, 2023, 54(7): 300－312.
ZHAO Chunjiang, LIANG Xuewen, YU Helong, et al. Automatic identification and counting method of caged chicken/egg based on improved YOLO v7[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(7): 300－312. (in Chinese)
[30] WANG X, VLADISLV Z, VIKTOR O, et al. Online recognition and yield estimation of tomato in plant factory based on YOLO v3[J]. Scientific Reports, 2022, 12(1): 8686.
[31] ZHANG Y, ZHANG W, YU J, et al. Complete and accurate holly fruits counting using YOLOX object detection[J]. Computers and Electronics in Agriculture, 2022, 198: 107062.

(上接第 **183** 页)

[15] TIAN Y H. Artificial intelligence image recognition method based on convolutional neural network algorithm[J]. IEEE Access, 2020, 8: 125731－125744.
[16] WANG L, ZHANG Q, QIAN Q, et al. Multispectral palm print and palm vein acquisition platform and recognition method based on convolutional neural network[J]. The Computer Journal, 2022, 65(6): 1461－1471.
[17] SHI Y, YUAN H C, XIONG C A, et al. Improving performance: a collaborative strategy for the multi-data fusion of electronic nose and hyperspectral to track the quality difference of rice[J]. Sensors and Actuators B: Chemical, 2021, 333: 129546.
[18] MEN H, LIU M, SHI Y, et al. Interleaved attention convolutional compression network: an effective data mining method for the fusion system of gas sensor and hyperspectral[J]. Sensors and Actuators B: Chemical, 2022, 355: 131113.
[19] LIU C C, CHU Z J, WENG S Z, et al. Fusion of electronic nose and hyperspectral imaging for mutton freshness detection using input-modified convolution neural network[J]. Food Chemistry, 2022, 385: 132651.
[20] CERVINI C, VERHEECKE-VAESSEN C, HE T T, et al. Improvements within the peanut production chain to minimize aflatoxins contamination: an Ethiopian case study[J]. Food Control, 2022, 136: 108622.
[21] LI X, WANG W, HU X, et al. Selective kernel networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 510－519.
[22] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132－7141.
[23] WANG Q, WU B, ZHU P, et al. ECA－Net: efficient channel attention for deep convolutional neural networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11534－11542.
[24] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision(ECCV), 2018: 3－19.
[25] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Rrecognition, 2021: 13713－13722.