

DOI: 10.13733/j.jcam.issn.2095-5553.2025.06.019

邵明明, 张立萍, 郑威强, 等. 基于改进YOLOv7的复杂环境下香梨和果梗的采摘识别[J]. 中国农机化学报, 2025, 46(6): 128-135

Shao Mingming, Zhang Liping, Zheng Weiqiang, et al. Picking recognition of fragrant pear and fruit stem based on improved YOLOv7 [J]. Journal of Chinese Agricultural Mechanization, 2025, 46(6): 128-135

基于改进 YOLOv7 的复杂环境下香梨与果梗的 采摘识别*

邵明明, 张立萍, 郑威强, 吕文涛, 雷佳乐

(新疆大学机械工程学院, 乌鲁木齐市, 830017)

摘要:为解决遮挡、光照、密集程度等复杂生长环境造成香梨采摘识别成功率不高及夹取采摘使果实破损的问题,提出一种基于YOLOv7的改进模型与一种剪抓相辅的采摘方式;制作基于真实复杂采摘环境下的香梨与果梗的数据集;增加BiFormer注意力机制提高模型对小目标的检测精度及效率;增加损失函数WIoU提升目标检测锚框的质量。经试验测试,改进的模型对样本数据集的平均精度均值 mAP 达到89.9%,比原模型提升1.9%,少特征的果梗平均精度 AP 值提高3.0%。与SSD、Faster R-CNN、Deformable DETR、YOLOv7模型对比, mAP 值与检测速度均有较大的提升。结果表明,提升后的模型能较好完成对香梨与果梗的检测,模型识别速度快,能够满足实时检测需求。

关键词:香梨与果梗;注意力机制;损失函数;复杂环境;图像识别

中图分类号:S24 **文献标识码:**A **文章编号:**2095-5553(2025)06-0128-08

Picking recognition of fragrant pear and fruit stem based on improved YOLOv7

Shao Mingming, Zhang Liping, Zheng Weiqiang, Lü Wentao, Lei Jiale

(College of Mechanical Engineering, Xinjiang University, Urumqi, 830017, China)

Abstract: To address the challenges of low success rate in picking and recognizing fragrant pears, caused by a complex growth environment with factors such as occlusion, illumination, and density, as well as fruit damage during harvesting, an improved model based on YOLOv7 and a cutting-and-grasping picking method were proposed. A dataset was created for fragrant pears and fruit stems, reflecting real-world complex picking conditions. The BiFormer attention mechanism was incorporated to enhance the detection accuracy and efficiency of the model, particularly for small targets. Additionally, the WIoU loss function was introduced to improve the quality of anchor frame, thereby enhancing target detection precision. The experimental results showed that the mAP of the improved model for the sample dataset reached 89.9%, which was a 1.9% increase over the original model. The AP value for detecting fruit stems, which have fewer features, increased by 3.0%. When compared to SSD, Faster R-CNN, Deformable DETR, and YOLOv7, both the mAP and detection speed showed significant improvements. These findings indicate that the enhanced model can more effectively detect pears and fruit stems while maintaining fast recognition speeds, making it suitable for real-time detection.

Keywords: fragrant pear and fruit stem; attention mechanism; loss function; complex environment; image recognition

0 引言

新疆库尔勒地区日照时间长,昼夜温差大,极大促进了香梨生长和糖分储存^[1],当地盛产的香梨皮薄肉细,

酥脆爽口,汁多甘甜,香味浓郁,耐贮藏^[2]。新疆香梨种植面积和产量位居国内香梨产量的前列。然而,目前当地的香梨采摘仍以传统的人力采摘为主,人工采摘成本高、速度慢,进一步导致经济效益低,是制约库尔勒香梨

收稿日期:2023年7月10日 修回日期:2023年12月24日

* 基金项目:国家自然科学基金项目(52265039);新疆维吾尔自治区重点研发项目计划(2022B02028-4)

第一作者:邵明明,男,1992年生,河南周口人,硕士研究生;研究方向为机器视觉。E-mail: 1078722937@qq.com

通讯作者:张立萍,女,1979年生,宁夏青铜峡人,博士,硕导;研究方向为现代制造装备技术等。E-mail: Zhlp@xju.edu.cn

走出去、阻碍增产增益的棘手因素,同时也是束缚新疆瓜果产业化发展的关键因素^[3]。

近几年,人工智能领域的科学技术进入高速发展的新阶段,机器代替人工的情况已经出现在多个领域。开发机器人进行自主式采摘任务的关键是率先完成对香梨和果梗的识别与定位。随着这几年深度卷积网络的迅速发展与不断地开源,国内外学者将目标检测网络运用到果蔬的采摘中,果蔬的检测、识别、定位、采摘等研究取得了较大进展。

Hameed等^[4]通过改进Mask-RCNN算法提升在光照强度等非结构性复杂环境下模型对水果识别的准确度与定位精度,然而Mask-RCNN作为双阶段的目标检测方式,其检测的速度慢导致检测效率较低,使得其在果蔬的实时检测中的性能并不太高;Zhang等^[5,6]通过使用Faster R-CNN算法模型对苹果进行检测识别,识别的准确率与效率均取得较大提升,同时也提升了在复杂环境下小目标物的检测效果,但因Faster R-CNN算法在模型训练时需要大量的数据集做支撑,在少样本或样本不丰富的情况下训练模型的检测效果有待提升。近几年,一阶段的目标检测算法发展迅猛,具有检测速度快、精度高等特点,因此被广泛应用于果蔬采摘的目标检测、工程应用等实际场景中。Tian等^[7]采用改进的YOLOv3模型算法对苹果进行实时检测,较好地解决了光照变化、枝叶遮挡及果实密集造成的相互遮挡等复杂因素影响下识别率低的问题。王新彦等^[8]对YOLOv5算法的检测模型进行改进,有效解决了遮挡环境下割草机器人识别检测精度低及速度慢的问题,并且在不影响检测效果的情况下显著降低了检测模型的大小,为后续的模式部署创造了有利的条件。Liu等^[9]将改进后的YOLOv5模型应用到苹果检测上。Kalfas等^[10]利用YOLOv7检测模型对火龙果进行检测,取得了较好的检测效果,但检测的果实与背景颜色差异较大,对于果实颜色和背景颜色相近的香梨来说,迁移使用效果不太理想。宋怀波等^[11]将YOLOv7进行轻量化、图像增强、通道融合等方面的优化,较好地克服了小目标、目标与背景接近、目标密集等因素的干扰,提升模型对苹果的检测效果。张三林等^[12]运用YOLOv5对核桃进行检测,并借助机械臂完成分拣,其分拣精准率高,但因其分拣背景较为单一,与复杂采摘背景差异较大,其数据集制作思路具有显著的参考价值。

基于此,本文建立复杂自然环境下香梨与果梗的数据集,选用YOLOv7作为检测网络模型,并根据建立的数据集样本特性对算法模型进行改进优化,以期香梨的智能设备多种自主化采摘方式研究提供技术支持。

1 材料与方法

1.1 试验数据获取

试验主要以新疆库尔勒香梨为数据样本,制作1个满足香梨采摘识别与检测的数据集。为达到真实的采摘环境,试验数据集图片由库尔勒香梨果园的实地拍摄和网上搜集2种途径获得。图像尺寸为4 624像素×2 136像素,取得能够满足试验条件的图像合计4 240张,图片的保存格式是JPG。图1是获取的在复杂生长环境下不同强度光照、不同遮挡、不同密集程度的香梨及果梗的数据样本。

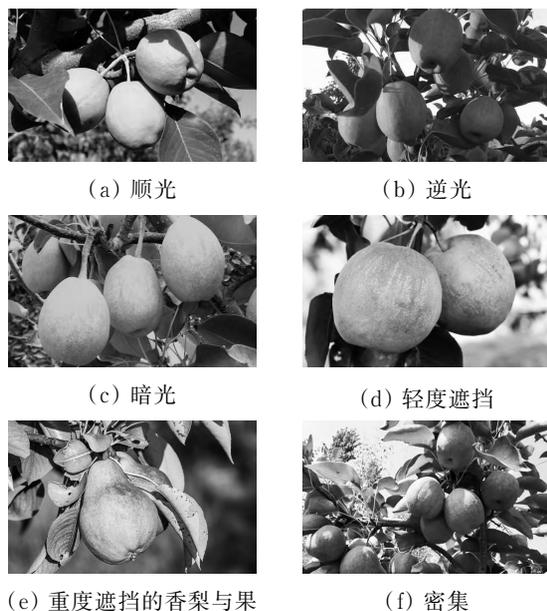


图1 复杂采摘环境下香梨与果梗样本数据

Fig. 1 Sample data of fragrant pear and fruit stem under complex picking environment

1.2 数据标注与分析

根据剪抓相辅助的采摘方式,确定梨(pear)、果梗(stalk)两类标注标签,利用图像标注工具Labelimg对采集的样本图像进行手工标注,并对不同环境下的照片按照8:1:1的比例随机划分成模型的训练集3 392张、验证集424张、测试集424张。其中枝叶遮挡、光照强度等因素是导致模型获取香梨及果梗特征信息偏差的主要因素,并造成目标漏检误检,进而降低模型的检测性能。香梨的采摘主要是对梗折断或剪切,抓取梨拖拽或旋转使得果与树分离,完成采摘的目的。因此,在识别检测的过程中只需对香梨与梨梗2类目标进行检测识别与位置获取。

2 网络模型与改进

2.1 YOLOv7网络检测模型

YOLO算法是近几年迅速发展起来的一阶段典

型目标检测算法,因其较快的运行速度和较高的检测精度而被广泛应用在实时检测的工作当中,根据 YOLOv7 在公开数据集 MSCOCO 上取得的测试结果可知,其检测精度与速度均超越同时期的其他 YOLO 系列模型,因而也是选用 YOLOv7^[13]作为对香梨与果梗进行目标检测的主要理由。YOLOv7 网络结构如图 2 所示,包含 3 个模块,分别是 Input 层、Backbone 层和 Head 层。

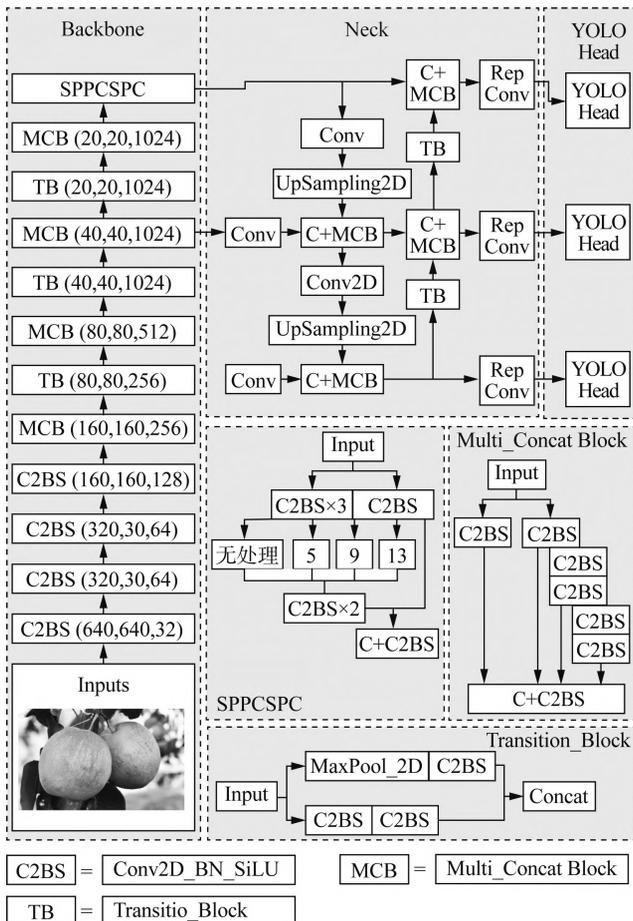


图 2 YOLOv7 网络结构

Fig. 2 YOLOv7 network structure

Input 层作为图片信息的输入层,主要的功能是将信息进行预处理后输入到下一模块的 Backbone 层; Backbone 层又叫提取特征层,主要任务是用 3 种不同大小的锚框从图像上获取目标的信息特征,网络模型是由 51 层(0~50 层)不同的卷积模块组合而成,其最大的亮点是使用了高效的 E-ELAN 网络架构,可以提高算法的检测精度与效率,Backbone 层输出 3 个不同尺度下的图像特征信息,其所在位置为检测模型特征层的 24 层、37 层以及 50 层; Head 层主要是将 Backbone 模块输出的特征信息进行处理,生成有目标位置的边界框,同时预测目标类别并带有与之对应的置信度信息,该模块主要由 SPPCPC 层、Conv 层(若干)、MPCConv 层与 REPconv 层构成,3 种不同尺寸的

特征图分别在 Head 模块的第 75 层、88 层与 101 层生成,在经过重参数化结构 REPconv 层处理后输出 3 个不同尺度的预测结果锚框,便于后续算法模型的部署。

2.2 网络模型改进

2.2.1 BiFormer 注意力机制

注意力机制是嵌入在机器学习模型中的一种特殊的结构机制,可用于模型的自动学习和求解所输入数据对输出数据的作用大小,进一步获得相应的权重系数,同时以“实时加权”的方式对目标区域进行强化,同时弱化作为背景的干扰区域的结构机制。当前的注意力机制主要分为硬注意力机制(局部注意力机制)、软注意力机制(全局注意力机制)以及自注意力机制(内注意力机制)。硬注意力机制是对所有的输入特征配置非 0 即 1 的权重系数,硬注意力机制只关注感兴趣的特征信息,对不感兴趣特征信息直接舍弃掉。因其在复杂环境下的数据集表现不佳,因此并未在 YOLOv7 中进行部署。软注意力机制是将输入的特征图配置 0~1 的权重,对特征信息做出全局考虑,主要有通道注意力机制、空间注意力机制以及融合二者的混合注意力机制,但是对小目标的香梨与果梗的识别精度与速度并不够突出。与以上 2 种注意力方式不同的是,自注意力机制根据输入的特征图之间的相互作用并进行复杂的矩阵计算来获得各特征图的权重系数。自注意力机制这种不依赖于卷积网络,更加关注每个像素特征之间的相互作用关系,更适合特征信息不足的香梨与果梗这样的小目标识别,这也是关注自注意力机制的原因之一。

自注意力机制源自 Transformer 模型,是该模型的核心模块,突出的优势是能够帮助网络获取上下文长距离间的相互关系,但缺点是内存占用大、计算成本过高。基于此,借助双层路由的一种新的动态稀疏注意力机制应运而生即 BRA^[14],BRA 模型在对输入的香梨与果梗的图片进行处理时的结构及原理如图 3 所示,可以看到,该模型是将所收集的前 k 个相关窗口中的键值对和借助稀疏性操作直接优化掉所认为最不相关区域,仅保留少部分相关联的路由区域,之后,在这些相关的路由区域的联合作用下通过借助具备动态的查询感知稀疏性能的细粒度 token-to-token 注意力,更灵活地实现计算分配和内容感知。同时,运行方式有助于缩减参数量与计算量。

基于 BRA 模块,BiFormer 提出一种通用视觉转换方式。如图 4 所示,其在延续大部分 vision transformer 架构设计框架的同时,借用四层金字塔模型结构,即 32 倍下采样。BiFormer 在香梨与果梗的数据集图片输入的第一阶段应用的是重叠块融合嵌入,在第二到第四阶段应用块合并模块用来降低输入空间特征图的分辨

率,同时增加模型通道数获取更多的特征信息,随后利用 BiFormer 块进行连续的特征变换。特别注意的是,在每一个模块的初始均是借用深度卷积网络来完成特征的相对位置信息的隐式编码。最后则依次使用 BRA 模块和扩展率为 2 的多层感知机(MLP)模块,分别应用在交叉位置关系建模和对每个位置的嵌入。因为 BiFormer 利用查询自适应的方式只关注相关特征并进行标记,不会分散其他不相关特征标记的注意力,所以 BiFormer 模型具备较好的运算性能和较高的计算效率。因此,在 YOLOv7 中添加 BiFormer 注意力机制。

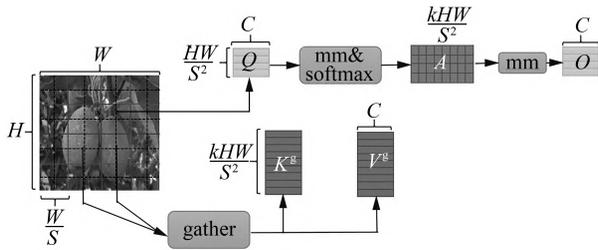


图 3 Bi-Level Routing Attention 注意力机制网络架构及原理

Fig. 3 Bi-Level Routing Attention attention mechanism network architecture and principle

注: H 表示输入图像的高; W 表示输入图像的宽; S 表示下采样因子,用于将图像尺寸缩小为 $H/S \times W/S$,降低计算量; K^k 表示键向量, K^k 与查询交互生成注意力; Q 表示查询向量,用于计算注意力权重; A 表示注意力矩阵,通过 Softmax 处理后生成,表征特征间关联权重注意力矩阵; O 表示输出,经矩阵乘法 mm 运算后的最终特征结果; C 表示通道数,描述特征图的通道维度; V^v 表示值向量,聚合特征信息。

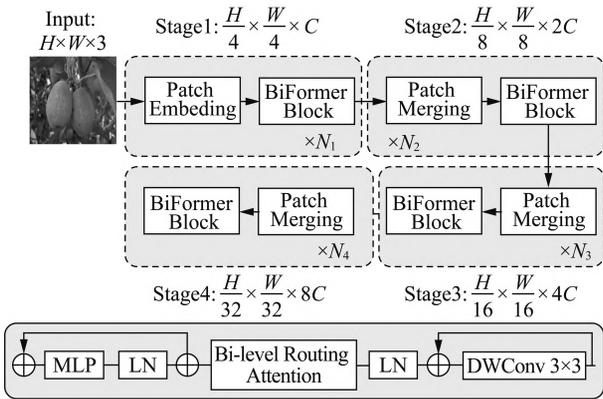


图 4 BiFormer 注意力机制网络架构及原理

Fig. 4 BiFormer attention mechanism network architecture and principle

2.2.2 WIoU 损失函数

YOLOv7 算法模型的损失函数用于更新梯度的损失,是延续 YOLOv5 损失函数对 3 个损失项进行加权求和,分别是目标分类损失 (L_{cls})、目标置信度损失 (L_{obj})和锚框坐标回归损失 (L_{CioU}),如式(1)所示。

$$Loss = L_{CioU} \times 0.05 + L_{obj} \times 0.1 + L_{cls} \times 0.125 \quad (1)$$

目标置信度损失和分类损失是采用带对数的二值交叉熵损失。为更好地解决不同质量样本间的 BBR

平衡问题时未被充分利用的静态聚焦机制,Tong 等^[15]提出一种基于 IoU 的损失,动态非单调聚焦机制 (WIoU)使用“离群度”代替 IoU 对锚框的性能评估,同时提供高效的梯度增益分配策略。该策略不仅削弱了高质量锚框的竞争占比,同时也减弱了低质量样本产生的有害梯度。这使得 WIoU 能够聚焦在普通质量锚框上,能进一步提升检测器的检测性能。该方式能够进一步提升模型对特征信息不足的果梗的检测效果。

WIoU 共有 v1、v2、v3 三个版本,因只应用动态非单调聚焦机制的 WIoUv3,故只对 WIoUv3 进行介绍。WIoUv3 利用动态非单调聚焦明智的梯度增益分配策略,对高低质量样本的权值再分配进行动态实时调整,使得 WIoUv3 获得优越的性能。同时,证明动态非单调聚焦机制高效的性能。

WIoUv3 定义离群度以描述锚框的质量,其定义如式(2)所示。

$$\beta = \frac{L_{IoU}^*}{L_{IoU}} \quad \beta \in [0, +\infty) \quad (2)$$

式中: L_{IoU}^* ——当下预测锚框的 L_{IoU} 数值;

$\overline{L_{IoU}}$ ——数据集中所有预测锚框 L_{IoU} 损失的平均值。

离群度小意味着锚框质量高,为其分配一个小的梯度增益,以便使边界框回归聚焦到普通质量的锚框上。对离群度大的锚框分配较小的梯度增益,将有效防止低质量示例产生较大的有害梯度。

$$L_{WIoUv3} = rL_{WIoUv1} \quad (3)$$

$$r = \frac{\beta}{\delta \alpha^{\beta-\delta}} \quad (4)$$

$$L_{WIoUv1} = R_{WIoU} L_{WIoU} \quad R_{WIoU} \in [1, e) \quad (5)$$

$$R_{WIoU} = \exp \left[\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)^*} \right] \quad (6)$$

式中: W_g ——最小包围框的宽;

H_g ——最小包围框的高;

L_{WIoU} ——动态非单调聚焦机制损失;

x_{gt} ——真实目标框的中心横坐标;

y_{gt} ——真实目标框的中心纵坐标;

α, δ ——超参数。

当参数 $\beta = \delta$ 时, $r = 1$ 。当锚框的离群程度满足 $\delta = C$ (C 为常数)时,锚框会取得最高的梯度增益。因为 $\overline{L_{IoU}}$ 是动态的,锚框的质量划分标准也会是动态的,这会使 WIoUv3 总是能做出最符合当前实际情况的梯度增益分配策略。同时,在训练初期为预防低质量锚框发生,预先进行部分初始化设置,例如使 $\overline{L_{IoU}} = 1$ 且 $\overline{L_{IoU}} = 1$ 的锚框具有最高的梯度增益。为了能够在初

期的训练阶段获得上述策略,还需预先设置一个动量 m 来延迟 $\overline{L_{IoU}}$ 的时间,让其接近真实值 $\overline{L_{IoU-real}}$ 的时间。动量值设置如式(7)所示。

$$m = 1 - \sqrt[n]{0.05} \quad (7)$$

式中: n ——模型训练中设置的 $batchs$ 数值;

t ——模型训练中平均精度均值 $mAP@0.5$ 的上升速度明显放缓的轮次。

由图5可知,模型训练 mAP 曲线在 $t=15$ 时发生较大的转变。

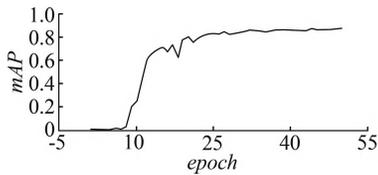


图5 数据集在YOLOv7上训练模型收敛情况

Fig. 5 Data set training model convergence on YOLOv7

2.3 模型训练

2.3.1 试验平台

进行模型训练与对模型性能进行检测的设备环境信息如表1所示,该参数是结合自身设备硬件性能与软件间的兼容性进行配置的。

表1 试验平台环境参数

Tab. 1 Experimental environment parameters

名称	参数
CPU	Intel(R) Xeon(R) Platinum 8222L CPU @ 3.00 GHz
GPU	英伟达 RTX3080Ti, 显存为 12 G
操作系统	Ubuntu20.04
深度学习框架	Pytorch1.12
加速环境	CUDA11.3
编程语言	Python3.8

2.3.2 模型训练的超参数设置

模型训练过程中的超参数设置如表2所示。根据硬件设备性能,并结合模型训练过程中损失函数值变化曲线与平均精度的收敛情况对超参信息进行设置。未使用官方训练好的预训练权重进行模型训练,主要是从建立的数据集里提取建模参数,提升训练模型的针对性,同时减少模型参数。

表2 训练参数

Tab. 2 Training parameters

参数	数值	参数	数值
图片尺寸/(像素×像素)	640×640	权重衰减系数值	0.000 5
迭代周期次数值	50	学习速率值	0.01
批次大小值	12	置信度阈值	0.5
动量参数值	0.937	非极大抑制阈值	0.3

在训练过程中,当 $epoch=30$ 时,模型已经趋于收

敛,因此结合训练过程中损失函数收敛趋势及训练过程中的时间成本将模型训练周期设置为 50 个 $epoch$; 批次大小($batchsize$)设置为 12,这是结合设备显卡性能同时节省模型训练时间而设定的最佳值,批次过小训练速度慢,占用时间,批次过大则会爆显存而导致训练中断;将学习率设置为 0.001 则会出现收敛速度缓慢,训练周期过长,将学习率设置为 0.1 则会导致训练过程损失函数出现剧烈振荡,损失函数无法收敛,而在学习率为 0.01 时,模型训练过程能够较稳定和快速收敛,故模型训练学习率设置为 0.01,其余参数使用原模型的推荐参数,在使用中未做更改。

2.3.3 评估指标

采用多项衡量指标对改进 YOLOv7s 算法模型的检测性能进行评估,指标分别是精确度(P)、召回率(R)、平均精度均值(mAP)、调和平均值($F1$)及单张图片的检测耗时(幅/ms)。 mAP 值和单张图片检测用时是目标检测算法中衡量检测模型性能比较重要的评估指标,主要是查验检测模型的精确度与速度。精确度(P)、召回率(R)、平均精度均值(mAP)以及调和平均值($F1$)计算如式(8)~式(11)所示。

$$P = \frac{TP}{TP + FP} \times 100\% \quad (8)$$

$$R = \frac{TP}{TP + FN} \times 100\% \quad (9)$$

$$mAP = \frac{\sum_{k=1}^N PR}{N} \times 100\% \quad (10)$$

$$F1 = \frac{2PR}{P + R} \times 100\% \quad (11)$$

式中: TP ——实例是正类,但是被预测成为正类;

FP ——实例是负类,但是被预测成为正类;

FN ——实例是正类,但是被预测成为负类;

N ——类别总数。

2.3.4 模型的融合

对 YOLOv7 网络模型增加 BiFormer 注意力机制,使用 WIoUv3 损失函数对不同质量的香梨与果梗检测中的预测边界框经过自学习进行实时权重的调整。主要嵌入方式操作:在 YOLOv7 特征加强网络的 93 层进行调用,通过在 common 文件夹导入注意力机制 BiFormer,同时将机制的调用指令嵌入网络模型结构文件 yml 里,并修改模型 model 文件夹中 YOLO.py 文件里的参数,完成训练过程对该注意力机制进行调用;将 WIoU 损失函数代码导入 utils 文件夹下的 general.py,同时修改 model 下的 loss.py 文件里的调用指令及参数配置,实现 2 种模块与 YOLOv7 网络模型的融合。

3 结果与讨论

3.1 YOLOv7 模型与 YOLO 系列其他模型对比分析

模型训练与性能检测均在同一个香梨与果梗的数据集上开展,训练过程中均未使用官方提供的预训练权重。由表3可知,当迭代周期相同,YOLOv7模型的 $mAP@0.5$ 高达88.0%,比YOLOv5m、YOLOv5s、YOLOv5l^[16]、YOLOv6^[17]高2.4、2.8、4.3、2.1个百分点,尤其在目标特征少、识别难度大的果梗数据集上优势更为明显。在果梗数据集上,YOLOv7的 $AP_1@0.5$ 比YOLOv5m、YOLOv5s、YOLOv5l、YOLOv6分别高出2.7、3.6、3.3、2.5个百分点。在检测速度方面,模型检测的运行速度比YOLOv5系列算法略微降低,

这是YOLOv7学习网络的层级增加、运算量增大导致的,但YOLOv7的15 ms的检测速度可以满足香梨机械化自主式实时采摘的检测需求。YOLOv7模型对香梨检测的 AP 值高达93.8%,而对果梗检测的 AP 值为82.2%。YOLOv7模型对香梨的检测效果比对果梗更佳,这是由于香梨的外形相较于果梗尺寸更大、表面特征更多,故而更容易被识别。果梗的尺寸较小,表面纹理特征较少,同时容易被枝叶遮挡,导致检测效果相对不佳。但该模型对香梨与果梗检测的 $mAP@0.5$ 达88.0%,而且对单幅图片的检测时间仅为15 ms,在满足快速检测的同时具有高精度水准。基于此,选择在YOLOv7算法基础上进行改进优化,以增强该算法在复杂采摘环境下对香梨与果梗的实时检测精度与效率。

表3 不同模型检测结果

Tab. 3 Different model test results

检测模型	$P/\%$	$R/\%$	$AP_0@0.5/\%$	$AP_1@0.5/\%$	$F1/\%$	$mAP@0.5/\%$	推理时间/ms
YOLOv5s	91.8	85.3	92.2	79.5	85.4	85.6	11
YOLOv5m	91.0	84.8	91.8	78.6	84.7	85.2	12
YOLOv5l	92.1	85.2	92.5	78.9	88.5	85.7	13
YOLOv6	91.6	85.7	92.1	79.7	85.5	85.9	13
YOLOv7	92.5	87.8	93.8	82.2	87.6	88.0	15

注: AP_0 为香梨数据集训练后的平均精度, AP_1 为果梗数据集训练后的平均精度。

3.2 BiFormer 注意力机制对模型的影响

为检测不同注意力机制加入YOLOv7模型后对模型训练的影响,分别将通道注意力机制SimAM^[18]、

空间注意力机制SAM^[19]、混合注意力机制CBAM^[20]以及自注意力机制BiFormer模块添加到YOLOv7的网络模型,训练效果如表4所示。

表4 不同注意力机制嵌入YOLOv7后模型训练效果

Tab. 4 Model training effect after different attention mechanisms are embedded in YOLOv7

注意力机制	$P/\%$	$R/\%$	$AP_0@0.5/\%$	$AP_1@0.5/\%$	$F1/\%$	$mAP@0.5/\%$	推理时间/ms
SimAM	88.2	83.1	88.5	72.2	79.5	80.4	16
SAM	88.1	82.9	87.1	71.3	78.4	79.2	17
CBAM	92.4	89.7	93.1	81.6	87.0	88.4	19
BiFormer	92.9	88.2	93.9	84.5	89.0	89.2	21

由表4可知,这4种注意力机制均降低原模型对测试图片的推理速度,相比于原YOLOv7模型,单幅图片的检测时间分别慢1 ms、2 ms、4 ms、6 ms,其中BiFormer推理速度下降幅度最大。增加注意力机制SimAM与SAM后,模型 $mAP@0.5$ 分别降低7.6、8.8个百分点,表明该通道与空间注意力机制对数据集的特征提取均出现弱化现象,且对少特征的果梗降低更为明显,而CBAM与BiFormer注意力机制均对模型的 $mAP@0.5$ 有提升作用,分别提升0.4、1.2个百分点,分析结果可知,CBAM与BiFormer注意力机制在本数据集上优于其他两种的注意力机制性能,检测模型的性能提升,这得益于两种注意力机制均采用多途径提取特征的方式提升模型对特征信息的提取能

力。BiFormer在数据集上的提升效果更为显著,尤其是对特征信息不足的果梗的改善效果更为显著。

3.3 Wise IoU 损失函数对模型的影响

为确定超参数 α 、 δ 在本数据集下对模型的影响,通过改变损失函数WIoU中的超参数组合值来分别对模型进行训练。对 α 取1.4、1.6、1.9、2.5, δ 取2、3、4、5值,模型进行训练与未改进时默认($\alpha=0$ 、 $\delta=0$,即未引入 α 、 δ 的情况)进行比较分析,由表5可得:(1)关于 mAP 值,当 $1.4 < \alpha < 1.9$ 且 $3 < \delta < 5$ 时,模型的 mAP 值逐渐降低,并且该阶段的 mAP 值均低于原函数,但当 $\alpha=1.9$ 且 $\delta=3$ 时, mAP 值出现较大的提升,比原函数高出1.4个百分点,同时果梗的检测精度提升较大,当 $1.9 < \alpha < 2.5$ 且 $2 < \delta < 3$ 时,模型的 mAP 值出

现降低趋势;(2)关于精确度,当 $\alpha=1.9$ 且 $\delta=3$ 时,准确率高于原函数,其余指标均降低;(3)关于召回率,当

$\alpha=1.9$ 且 $\delta=3$ 或 $\alpha=1.6$ 且 $\delta=4$ 时,召回率与原函数持平,其余指标均出现降低的情况。

表5 超参数 α 、 δ 对模型检测性能影响对比

Tab. 5 Comparison of the influence of hyper parameters α and δ on model detection performance

不同参数组合		P/%	R/%	AP ₀ @0.5/%	AP ₁ @0.5/%	F1/%	mAP@0.5/%	推理时间/ms
α	δ							
0	0	92.5	87.8	93.8	82.2	87.6	88.0	15
1.4	5	91.2	86.5	91.6	77.4	83.9	84.5	18
1.6	4	91.6	87.8	92.2	79.9	85.6	86.1	17
1.9	3	92.7	87.8	93.6	82.7	87.8	88.2	19
2.5	2	91.8	86.4	91.1	81.3	85.9	86.2	18

综上,当 $\alpha=1.9$ 且 $\delta=3$ 时,损失函数 WIoU 训练出来的模型检测性能提升最大。如表 5 所示,对 α 与 δ 取不同数值时的训练效果与未改进时默认 $\alpha=0$ 且 $\delta=0$ 的结果进行对比分析,当 $\alpha=1.9$ 且 $\delta=3$ 时,改进提升后的损失函数比原函数的 mAP 值提升 1.4 个百分点,在香梨与果梗的单类别 AP 值分别降低 0.2、1.5 个百分点,从所得试验数据可知, WIoU 损失函数对少特征数据样本进行一定的强化,进而改善目标特征不足导致的识别率低的情况。

3.4 改进后模型检测效果

通过利用提升后的检测模型对香梨与果梗数据集开展模型训练与性能检测的测试操作,运用测试集对模型性能展开测试,结果如图 6 所示。

从图 6 可以看出,提升后的检测模型能够在不同的光照强度、不同的遮挡程度以及果实不同密集程度等复杂环境下准确地识别出香梨与果梗 2 种目标,并且提升后的检测模型对测试集图片识别的平均准确率达到 85% 以上。说明改进后的检测模型对复杂采摘环境下香梨及果梗数据样本的识别检测性能较好。

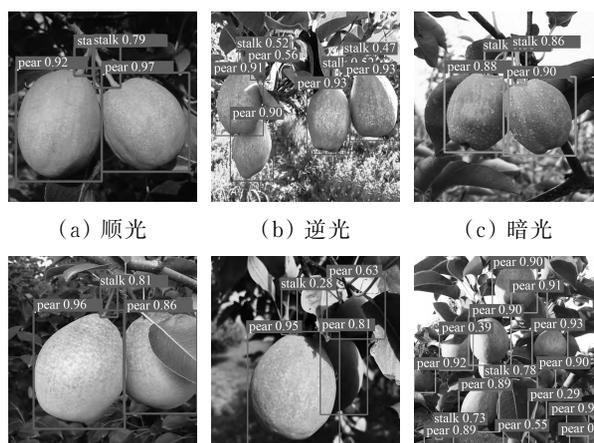


图 6 改进后的模型在真实采摘环境下的检测效果

Fig. 6 Detection effect of the improved model in the real picking environment

3.5 与其他模型的性能对比

在 YOLOv7 模型基础上添加 BiFormer 注意力机制和用来提升小目标少特征信息的损失函数 WIoU。在自建数据集上,将改进后模型分别与改进前 YOLOv7 模型、SSD、Faster R-CNN、Deformable DETR 进行对比,结果如表 6 所示。

表 6 不同检测模型在香梨与果梗数据集检测结果对比

Tab. 6 Comparison of detection results of different detection models in pear and stem datasets

检测模型	P/%	R/%	AP ₀ @0.5/%	AP ₁ @0.5/%	F1/%	mAP@0.5/%	推理时间/ms
SSD	85.2	79.3	77.2	58.5	66.7	67.9	10
Faster R-CNN	94.5	92.4	92.1	76.2	83.4	84.2	35
Deformable DETR	86.7	84.6	86.8	62.7	72.8	74.8	42
YOLOv7	92.5	87.8	93.8	79.2	85.9	86.5	15
Improved YOLOv7	94.7	91.3	94.5	85.2	89.6	89.9	23

4 结论

1) 在建立的香梨与果梗数据集上,改进模型的检测结果比起改进前,两类样本的数据集检测平均精度均值 mAP@0.5 达到 89.9%,提升 1.9%,其中干扰较多的果梗样本的平均精度均值提升 3%;与 SSD、Faster R-CNN、Deformable DETR 模型相比,平均精度均值分别

提升 22.0%、5.7%、0.1%,其中特征信息较少的果梗样本的平均精确率分别提升 26.7%、9.0%、22.5%;无论是综合检测效果还是特征信息较少的样本检测的平均检测精度,改进后模型的优势都比较明显。

2) 通过将 BiFormer 注意力机制添加到模型特征提取层,进一步提升在复杂采摘环境下目标的检测平均精准率,特别是提升对特征信息较少的果梗检测精

准率,能够减少香梨在采摘过程的漏检误检现象,虽然改进的模型检测速度略有降低,但仍可以满足智能设备进行实时采摘的需求,并且对机器采摘而言,这种略微降低检测速度而获取检测精度提升的方式更有助于提升智能设备采摘香梨的品质。

3) 通过添加 WIoU 损失函数并调整 α 、 δ 两个超参数的大小,经过试验找到与 YOLOv7 模型以及香梨和果梗数据集较匹配的参数组合,改善果梗特征信息少、识别精度低的问题,经过试验比较,确定超参数 $\alpha=1.9$ 且 $\delta=3$ 时,所得的香梨与果梗的检测模型性能提升效果最好。

参 考 文 献

- [1] 王瑾瑾,王彦发,马东芬,等. 巴州品牌农业发展模式分析——以库尔勒香梨为例[J]. 农村经济与科技, 2023, 34(4): 6—9.
- [2] Sébastien P, Tanguy L, Sandrine R, et al. Effect of carbohydrates and night temperature on night respiration in rice [J]. *Journal of Experimental Botany*, 2015, 66(13): 3931—3944.
- [3] 陈夕朦,崔菁,赵秀文,等. 技术伦理视野下人工智能的未来发展趋势[J]. 中国软科学, 2022(S1): 76—82.
Chen Ximeng, Cui Jing, Zhao Xiuwen, et al. Future development of “artificial intelligence”: A study from the perspective of technical ethics [J]. *China Soft Science*, 2022(S1): 76—82.
- [4] Hameed K, Chai D, Rassau A. Score-based mask edge improvement of Mask—RCNN for segmentation of fruit and vegetables [J]. *Expert Systems with Applications*, 2022, 190: 116205.
- [5] Yu Y, Zhang K, Yang L, et al. Fruit detection for strawberry harvesting robot in non-structural environment based on Mask—RCNN [J]. *Computers and Electronics in Agriculture*, 2019, 163: 104846.
- [6] Zhang J, Karkee M, Zhang Q, et al. Multi-class object detection using Faster R—CNN and estimation of shaking locations for automated shake-and-catch apple harvesting [J]. *Computers and Electronics in Agriculture*, 2020, 173: 105384.
- [7] Tian Y, Yang G, Wang Z, et al. Apple detection during different growth stages in orchards using the improved YOLO—V3 model [J]. *Computers and Electronics in Agriculture*, 2019, 157: 417—426.
- [8] 王新彦,易政洋. 基于改进YOLOv5的割草机器人工作环境障碍物检测方法研究[J]. 中国农机化学报, 2023, 44(3): 171—176.
Wang Xinyan, Yi Zhengyang. Research on obstacle detection method of mowing robot working environment based on improved YOLOv5 [J]. *Journal of Chinese Agricultural Mechanization*, 2023, 44(3): 171—176.
- [9] Liu T, Kang H, Chen C. ORB—Livox: A real-time dynamic system for fruit detection and localization [J]. *Computers and Electronics in Agriculture*, 2023, 209: 107834.
- [10] Kalfas I, Ketelaere B D, Bunkens K, et al. Towards automatic insect monitoring on witloof chicory fields using sticky plate image analysis [J]. *Ecological Informatics*, 2023, 75: 102037.
- [11] 宋怀波,马宝玲,尚钰莹,等. 基于YOLOv7—ECA模型的苹果幼果检测[J]. 农业机械学报, 2023, 54(6): 233—242.
Song Huaibo, Ma Baoling, Shang Yuying, et al. Young apple fruit detection based on YOLOv7—ECA model [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2023, 54(6): 233—242.
- [12] 张三林,张立萍,郑威强,等. 基于YOLOv5的核桃品种识别与定位[J]. 中国农机化学报, 2022, 43(7): 167—172.
Zhang Sanlin, Zhang Liping, Zheng Weiqiang, et al. Identification and location of walnut varieties based on YOLOv5 [J]. *Journal of Chinese Agricultural Mechanization*, 2022, 43(7): 167—172.
- [13] Wang C Y, Bochkovskiy A, Liao H Y M. YOLOv7: Train-able bag-of-freebies sets new state-of-the-art for real-time object detectors [J]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023: 7764—7475.
- [14] Zhu L, Wang X, Ke Z, et al. BiFormer: Vision transformer with Bi—level routing attention [J]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023: 10323—10333.
- [15] Tong Z, Chen Y, Xu Z, et al. Wise—IoU: Bounding box regression loss with dynamic focusing mechanism [J]. *arXiv preprint arXiv: 2301.10051*, 2023.
- [16] Jocher G, Stoken A, Chaurasia A, et al. Ultralytics/YOLOv5: v6.0—YOLOv5n “Nano” models, roboflow integration, tensorflow export, openCV DNN support [J]. *Zenodo*, 2021.
- [17] Li C, Li L, Jiang H, et al. YOLOv6: A single-stage object detection framework for industrial applications [J]. *arXiv preprint arXiv: 2209.02976*, 2022.
- [18] Yang L, Zhang R Y, Li L, et al. Simam: A simple, parameter-free attention module for convolutional neural networks [C]. *International Conference on Machine Learning*, 2021: 11863—11874.
- [19] Kirillov A, Mintun E, Ravi N, et al. Segment anything [J]. *arXiv preprint arXiv: 2304.02643*, 2023.
- [20] 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.