

DOI: 10.13733/j.jcam.issn.2095-5553.2024.03.031

黄威, 刘义亭, 李佩娟, 等. 基于改进 YOLOX-S 的苹果成熟度检测方法[J]. 中国农机化学报, 2024, 45(3): 226-232

Huang Wei, Liu Yiting, Li Peijuan, et al. Detection method of apple ripeness based on improved YOLOX-S [J]. Journal of Chinese Agricultural Mechanization, 2024, 45(3): 226-232

基于改进 YOLOX-S 的苹果成熟度检测方法^{*}

黄威¹, 刘义亭^{1, 2}, 李佩娟¹, 陈光明^{3, 4}

(1. 南京工程学院, 南京市, 211167; 2. 东南大学, 南京市, 210096;

3. 南京农业大学, 南京市, 210031; 4. 江苏省智能化农业装备重点实验室, 南京市, 210000)

摘要:准确检测果园中未成熟与成熟的苹果对果园早期作物的负荷管理至关重要, 提出一种能够实时检测苹果成熟度, 并估算出整棵果树果实数量的方法。为提高 YOLOX-S 网络在复杂场景下的检测能力, 在 FPN(特征金字塔)的残差连接处增加了 Coordinate Attention(位置注意力); 为更好地检测图像中生长密集、存在遮挡、尺寸较小的苹果, 将位置损失函数 IoU_Loss 更换为 CIoU_Loss。试验结果表明, 所提出的改进 YOLOX-S 检测算法相较于原算法, *mAP* 值提高约 1.97%, 苹果低成熟度、中等成熟度和高等成熟度的 *AP* 值分别为 90.85%、95.10% 和 80.50%。

关键词:苹果; YOLOX-S; 目标检测; 位置注意力; 成熟度检测

中图分类号:S24; TP391.4 **文献标识码:**A **文章编号:**2095-5553 (2024) 03-0226-07

Detection method of apple ripeness based on improved YOLOX-S

Huang Wei¹, Liu Yiting^{1, 2}, Li Peijuan¹, Chen Guangming^{3, 4}

(1. Nanjing Institute of Engineering, Nanjing, 211167, China; 2. Southeast University, Nanjing, 210096;

3. Nanjing Agricultural University, Nanjing, 210031, China; 4. Jiangsu Key Laboratory of Intelligent

Agricultural Equipment, Nanjing, 210000, China)

Abstract: Accurate detection of immature and mature apples in orchards is crucial to the load management of early crops in orchards. This paper proposes a method that can detect apple maturity in real time and estimate the number of fruits in the whole fruit tree. In order to improve the detection ability of the YOLOX-S network in complex scenes, Coordinate Attention (position attention) is added to the residual connection of the FPN (feature pyramid). In order to better detect the apple with dense growth, occluding and small size in the image, the position loss function is replaced from IoU_Loss to CIoU_Loss. The experimental results show that the *mAP* value of the improved YOLOX-S detection algorithm proposed in this paper is increased by about 1.97% compared with the original algorithm, and the *AP* values of apples with low maturity, medium maturity and high maturity are 90.85%, 95.10% and 80.50%, respectively.

Keywords: apple; YOLOX-S; object detection; coordinate attention; maturity detection

0 引言

中国是世界上最大的苹果出口国, 其苹果种植面积与产量均占世界总量的 45% 以上^[1], 中国苹果产业在世界苹果产业中占据主导地位, 对多数果树作

物而言, 作物的早期负荷管理十分重要, 会大幅影响果实的产量与品质。种植者通过了解果实的成熟度与数量大小等信息来估算作物的负荷量, 并借此实现高效的农业管理与合理的作物营养分配^[2]。

种植者想要在实际果园环境中精确估算出作物负

收稿日期: 2022 年 6 月 1 日 修回日期: 2022 年 7 月 13 日

^{*} 基金项目: 国家自然科学基金青年基金(61903184); 江苏省自然科学基金青年基金(BK20181017, BK2019K186); 南京工程学院引进人才科研启动基金(YKJ2018822); 中国博士后科学基金第 67 批面上项目(2020M671292); 江苏省博士后科研资助计划(B 类); 2021 年度省重点研发计划(产业前瞻与共性关键技术)(BE2021016—5)

第一作者: 黄威, 男, 1993 年生, 江苏泰兴人, 硕士; 研究方向为机器人控制与计算机视觉。E-mail: 1464754954@qq.com

通讯作者: 刘义亭, 男, 1985 年生, 湖北襄阳人, 博士, 副教授; 研究方向为导航、制导与控制、计算机视觉。E-mail: gcdlyt1985@163.com

荷量,需及时了解苹果的真实生长信息。而传统的人工计数方式存在着人力成本高、耗时长等弊端,为降低人力成本并提高检测速度,越来越多基于视觉的检测方式运用于苹果识别中^[3]。传统视觉检测方案主要是通过人工提取的特征进行判别,如颜色(颜色阈值)、形状(圆形霍夫变换)、局部纹理特征等^[4],这种方式易受环境因素影响且鲁棒性不高,在光照不足或苹果目标存在遮挡时其识别精度会很低。近年来随着人工智能技术在农业生产中的广泛应用,基于卷积神经网络的目标检测算法由于其卓越的特征提取能力(无需人工设计)被广泛运用于水果检测中^[5];熊俊涛等^[6]通过Faster R-CNN完成复杂环境下绿色柑橘的识别,其模型检测精度高,泛化性强但运行速度较慢,不能满足检测的实时性需求;Tian等^[7]通过改良的YOLOv3算法实现对不同成熟期苹果的检测估产,该算法响应速率与鲁棒性让人满意,但未对大视场(密集场景)下的苹果检测进行研究;赵德安等^[8]利用改进的YOLOv3算法,实现各类复杂环境下的苹果识别,该算法的检测速度可达60帧/s,具有良好的实时性,但在检测密集场景下的苹果时其效果较差。

为解决上述研究人员所提出模型存在的检测速度慢、密集场景下检测效果弱等问题,本文以苹果作为研究对象,在YOLOX-S算法^[9]的基础上进行相关改进,为增强模型对目标图像的特征提取能力,在FPN^[10]中的网络残差模块处融入位置注意力机制;为提高密集场景下小目标的检测精度,将原位置损失函数IoU_Loss

更换为对于小目标回归定位更精准的CIoU_Loss,提出一种改进YOLOX-S的苹果成熟度检测方法,为果园早期作物的负荷管理提供助力。

1 YOLOX-S 算法

当前主流的目标检测算法依据其原理可大致分为两类:一类为二阶段检测算法,该算法先生成包含可能目标物品的候选区并提取出每个候选区的相应特征,再完成对候选区的分类与回归,最后得到检测结果,此类算法以R-CNN系列为代表,精度较高但实时性较差;另一类为一阶段检测算法,此算法能够直接在特征图上实现对所测目标的分类与回归,以YOLO系列为代表,其检测效率高但精度稍逊于二阶段检测算法。在农业自动观测领域,在保持一定检测精度的情况下还需要良好的实时性,所以一阶段检测算法更适用于观测苹果的真实情况。

1.1 YOLOX-S 网络架构

YOLOX算法是2021年所提出的一阶段检测算法,其将目标检测算法领域中的众多杰出工作:解耦头、数据增强、无锚点、标签分配策略等与YOLO系列巧妙结合。该算法无论是AP数值还是推理速度,都实现了对YOLO系列算法的全面超越。本文采用的是YOLOX系列中的YOLOX-S网络,该网络是在YOLOv5-S的基础上进行改良的,整体网络结构如图1所示,共分为四部分:输入端、Backbone(主干网络)、Neck(颈部)、Prediction(预测端)。

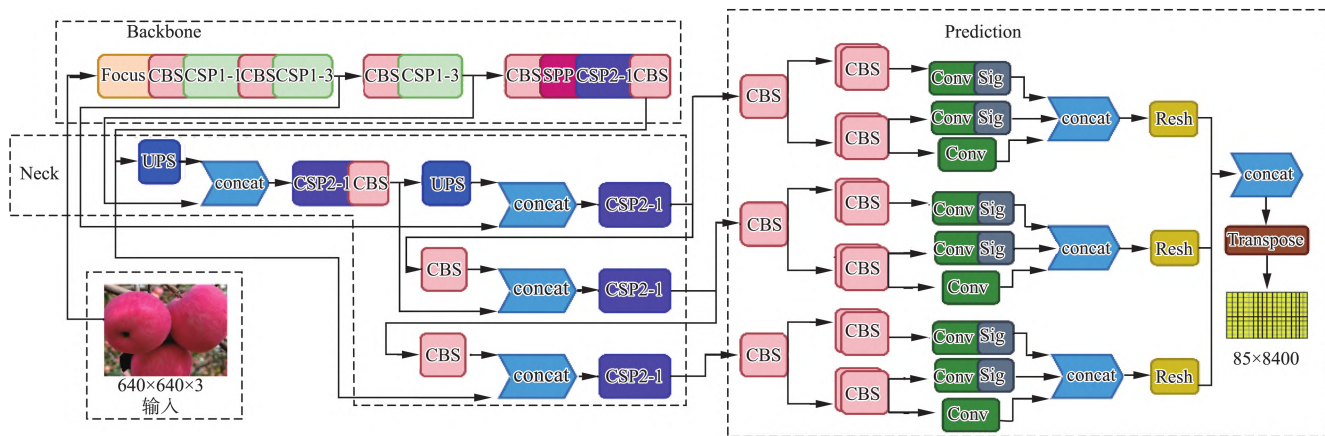


图1 YOLOX-S网络结构图

Fig. 1 YOLOX-S network structure diagram

整个网络的检测过程具体为:先在输入端应用Mosaic策略对输入的图片进行数据增强,丰富其背景;然后再将输入的图片统一裁剪处理为640像素×640像素的RGB格式传送至Backbone;Backbone主要包含了Focus模块、CSP(跨阶段局部)模块、SPP(空间金字塔池化)模块,先通过Backbone完成对目标浅

层特征的提取,再通过主要构成模块为FPN的Neck完成对目标更深层次特征的提取;最后来到Prediction部分,YOLOX舍弃了原本的YOLO Head,而是采用了Decoupled Head(解耦头),先前得到的特征信息分别传入三个不同尺寸的Decoupled head(解耦头)进行目标框预测,最终实现目标检测。

1.2 YOLOX-S 在检测苹果成熟度时遇到的问题

虽然 YOLOX-S 在检测苹果成熟度方面拥有不错的运行速率与检测性能,但在检测密集场景下的苹果目标时仍有不足。

1) 密集场景图像中的苹果目标较小,且存在严重遮挡,小目标的检测更依赖于浅层特征图,而浅层特征的提取主要通过 Backbone 来实现,由于 Backbone 中存在大量卷积会导致小目标特征信息的丢失。

2) YOLOX-S 的边界框回归损失函数采取的是 IoU_Loss,该回归损失函数会带来一些问题:首先是当蓝色虚线矩形框所代表的预测框与红色实线矩形框所代表的目标框不存在交集时,其损失值恒为 1;其次当二者相交为某一定值时,预测框与目标框会存在不同的重叠方式,如图 2 所示,既无法区分相对位置信息,也无法完成精确定位。



图 2 预测框落入目标框内的不同情况

Fig. 2 Different situations where the prediction box falls into the target box

2 改进 YOLOX-S 算法

针对上述 YOLOX-S 在密集场景下检测苹果成熟度所存在的问题,本文拟采用如下改进措施来提高模型在密集场景下的检测性能:一是在 FPN 模块的残差处引入位置注意力机制;二是采用 CIoU_Loss 作为边界框的回归损失函数。

2.1 引入位置注意力机制

在苹果成熟度检测任务中,常需对密集场景下的苹果进行检测,该场景下的小目标苹果极易在深层次网络中丢失其特征信息,随后发生漏检、误检的现象。而注意力机制的引入常常会改善模型的性能,但之前所提出的一些注意力机制往往会忽略位置信息^[11]。在视觉检测模型中加入 Coordinate Attention mechanism(位置注意力机制),这样做的优点是能够使得输入特征具有更丰富的信息,并协助模型更易精准定位与识别所感兴趣区域,有效增加小目标在特征图中权重,使其更易被模型学习,其主要结构如图 3 所示。

本文拟在 YOLOX-S 的 Neck 部分中的 FPN 模块里的 CSPLayer 结构中加入 Coordinate Attention,具体网络设计如图 4 所示。

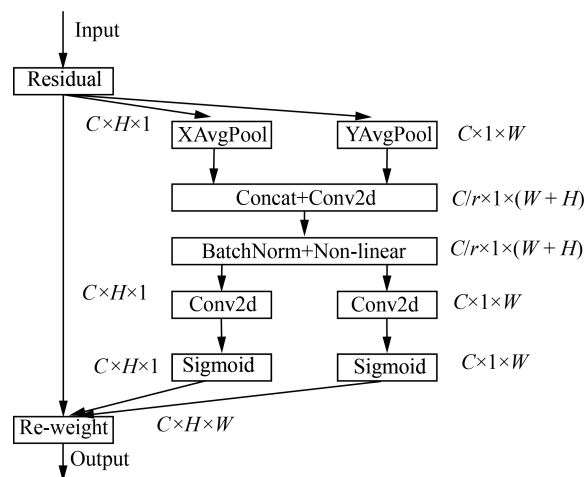


图 3 坐标注意力机制结构图

Fig. 3 Coordinate attention mechanism structure diagram

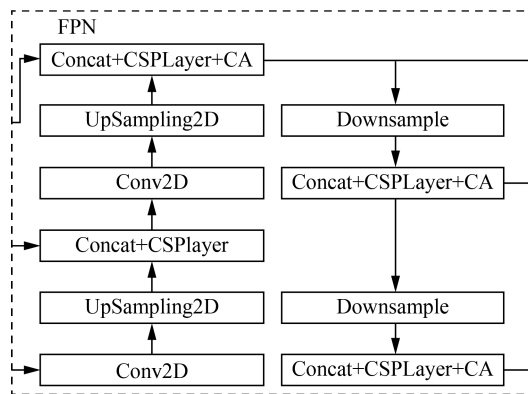


图 4 在 FPN 结构加入 CA

Fig. 4 Add CA to FPN structure

2.2 改进损失函数

YOLOX-S 的 Loss(损失)函数可由三部分组成:位置损失、置信度损失、类别损失。其中 YOLOX-S 采用的位置损失函数是 IoU_Loss^[12],该损失函数先求解出预测框与真实框之间交并比,然后再求其负对数。先假设预测框为 A ,目标框为 B ,则 IoU_Loss 的具体公式如式(1)所示。

$$IoU_Loss = 1 - \frac{(A \cap B)}{(A \cup B)} \quad (1)$$

由于 YOLOX 选取正样本的机制是依据预测结果是否落入真实框或真实框周围,而当实际检测目标为小目标时,常常会出现预测结果与真实标签毫不相交的场景,此时 IoU_Loss 的损失值恒为 1,为了改善这一情况,应根据预测结果与真实标签距离远近来做判断。本文拟更换位置损失函数 IoU_Loss 为 CIoU_Loss,其计算公式详见式(2)。CIoU_Loss 函数充分考虑了预测框 A 与目标框 B 之间的重叠面积与中心距离,如图 4 所示,通过度量预测框 A 与目标框 B 边缘框之间的距离,获得边界框中心点距离与宽高比信息以及 A 与 B 之间的长宽比,最终得到更好的边界回归效果。

$$CIoU_Loss = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + av \quad (2)$$

式中： b ——预测框 A 的中心点；

b^{gt} ——目标框 B 的中心点；

ρ ——欧氏距离；

c —— A 与 B 相交时所构成的最小外接矩阵对角线长度；

a ——权重系数；

v ——衡量长宽比一致性的参数。

将 IoU_Loss 更换成 $CIoU_Loss$ 能够使得模型在回归定位时更加准确,并提高检测性能。

3 试验与分析

3.1 数据采集与预处理

本文所采用的苹果数据集来源于 Hani^[13]所分享的苹果公开数据集 MinneApple,该数据集中包含了三种成熟度与多种背景环境下的苹果:有被枝叶遮挡的苹果、有被其他苹果遮挡的苹果、有迎着光的苹果、有逆着光的苹果、有近距离拍摄的苹果、也有远距离拍摄的苹果、有稀疏场景下的苹果、也有密集场景下的苹果。最终采用的苹果图像为 2 001 幅,每幅图像像素皆为 $1\,280 \times 720$ 的 PNG 格式的图片,部分苹果场景如图 5 所示,其中 1 502 幅划分为训练集,168 幅划分为验证集,331 幅划分为测试集。



(a) 近距离且密集的低成熟度苹果 (b) 远距离稀疏场景下的中等成熟度苹果 (c) 逆光远距离场景下高成熟度苹果

图 5 实际果园环境下的苹果图像

Fig. 5 Apple image in the actual orchard environment

由于户外场景复杂,光照强度多变,为提高模型在实际果园环境中的检测性能与鲁棒性,数据集的丰富性必不可少;此外由于当前深度学习还是由数据驱动的科学,所以需要大量的数据来学习。故需对原训练集中的 1 502 幅图像进行相应的数据增强,在进行数据增强时发现其中低成熟度的苹果与中等成熟度的苹果数据量较少,与高成熟度苹果数据量相比存在数据不平衡现象,所以着重对这两类数据进行了数据增强,采用了添加随机像素、高斯噪声、运动模糊、水平翻转、不等比例缩放、随机亮度、随机剪裁与填充等数据增强

的方式,部分增强效果如图 6 所示。



(a) 不等比缩放与高斯噪声 (b) 随机剪裁与随机亮度 (c) 不等比缩放与运动模糊

图 6 数据增强后的部分图片效果

Fig. 6 Part of the picture effect after data enhancement

原 1 502 幅图像中,低成熟度苹果图像数为 443 幅,中等成熟度苹果图像数为 335 幅,高成熟度苹果图像数为 724 幅。经过上述数据增强后,低成熟度与中等成熟度苹果的图像数目达到 6 268 幅,再加上 3 620 幅高成熟度苹果图像,总计训练图像为 9 888 幅。

完成训练集的数据增强后,采用 LabelImg 软件将数据图像中不同成熟度的苹果^[14]分别以 LR_apple(低成熟度)、MR_apple(中等成熟度)、HR_apple(高成熟度)这三个类别标注出来,最后将数据集保存为 PASCALVOC 格式。

3.2 模型训练

本文进行模型训练时采用的硬件为 DELL 的 P5820 塔式工作站,其配置为:NVIDIA GeForce RTX2080 显卡,Intel[®] Xeon CPU E3-1231v3 @ 3.40 GHz 处理器,32 GB 内存。工作站所采用的操作系统为 Windows 10 Version 21H2,模型搭建以及运行的框架为 Pytorch1.10.1,并配合使用 CUDA11.3 与 CUDNN v8.2.1 实现对深度神经网络的加速训练。

模型在训练时采用 Freeze train(冻结训练),能够加速模型收敛并减少训练时间^[15],训练时采用的 epoch(训练世代)为 1 500,使得模型训练时有足够多的 step(迭代次数)来进行学习;在模型训练时采用 Mosaic 数据增强策略,该策略的原理是将训练数据集中四张图片随意裁剪缩放并拼接成一张图片,具体效果见图 7,这样做的目的一是能够把原图上的目标尺寸进一步缩小,增强对小尺寸目标的检测能力;二是能够丰富检测目标的背景,由于该策略会使得训练图片过于失真不符合真实场景,所以会在最后 30% 的 epoch 关闭它。

EMA(指数平均数)的引入让 Loss 曲线更加平滑;训练所采用的优化器为 SGD(随机梯度下降);由于 RTX2080 显存的大小只有 8 GB,经实验测试将冻结训练时的 Batch size(批量大小)设为 32,非冻结训练时的 Batch size 设为 16。具体的训练参数设置如表 1 所示。



图 7 采用 Mosaic 后的图片效果

Fig. 7 Picture effect after using Mosaic

表 1 模型的训练参数设置

Tab. 1 Model training parameter setting

参数	Freeze	Unfreeze
Epoch	300	1 200
Batch size	32	16
lr	0.000 1	0.000 01
优化器	SGD	SGD

3.3 评估指标

为评价模型对不同成熟度苹果的检测结果,采用了目标检测算法中常使用的评估指标: mAP (平均精度均值)、 $Recall$ (召回率)、 FPS (每秒传输帧数)。 AP 是指计算单类别的模型平均精度,该值由每类目标的 P (查准率)和 R (召回率)计算得到的, mAP 指的是所有类别的平均 AP 值。计算如式(3)~式(7)所示。

$$P = \frac{TP}{TP + FP} \quad (3)$$

$$R = \frac{TP}{TP + FN} \quad (4)$$

$$AP = \sum_{k=1}^N P(k) \Delta R(k) \quad (5)$$

$$\Delta R(k) = R(k) - R(k-1) \quad (6)$$

$$mAP = \frac{\sum AP}{N(class)} \quad (7)$$

式中: TP ——正样本预测的正确数量;

FP ——正样本预测的错误数量;

FN ——负样本预测的错误数量;

k ——每个样本点的索引;

$class$ ——类别;

N ——数据总数量。

3.4 试验结果分析

为综合评估本方法所提出改良模型的检测性能,共设计了两组试验:第一组是对比 YOLOX-S 模型改良前后的检测性能;第二组是将本方法所提出的改良模型与其他主流目标检测模型进行性能对比。

3.4.1 模型改良前后的性能分析

本方法采取的改进策略可归类为两种:一种是在模型训练时采取的改进措施,如更换优化器、采取 Mosaic 数据增强等;另一种就是对模型本身结构进行改进,如引入 CA 注意力模块以及更换边界框的回归损失函数为 CIOU_Loss。

为验证本方法所采取的改进措施是否有效,在同一数据集下,采用相同的训练参数进行消融实验,来判断该措施是否有效。具体实验流程就是在原有模型的基础上依次采用如下改进措施,如训练时采用 SGD 优化器与 Mosaic 数据增强、FPN 模块残差连接处加入 CA 注意力模块、更换 CIOU_Loss,然后逐一比较每次改变后模型的性能,如采用某一改进策略就用“√”表示,未采用就用“—”来表示,具体试验效果见表 2。

表 2 YOLOX-S 消融试验结果

Tab. 2 YOLOX-S ablation experiment results

SGD	Mosaic	CA	CIOU	mAP /%	时间/ms	帧数/fps	AP^L /%	AP^M /%	AP^H /%	R^L /%	R^M /%	R^H /%
—	—	—	—	85.3	21	47.65	85.79	87.18	82.56	74.86	78.76	72.26
—	√	—	—	86.85	21	47.65	87.41	91.93	81.21	78.15	83.12	73.65
√	√	—	—	87.41	22	44.05	88.72	92.47	81.05	79.97	84.96	73.91
√	√	√	—	88.64	24	42.56	90.44	94.96	80.52	85.46	90.62	75.51
√	√	√	√	88.82	24	42.06	90.85	95.1	80.5	85.74	90.93	75.96

表 2 中 AP^L 、 AP^M 、 AP^H 分别代表的是当阈值设定为 50 时,类别为 LR_apple、MR_apple、HR_apple 的 AP 值。 R^L 、 R^M 、 R^H 则依次表示 LR_apple、MR_apple、HR_apple 这三个类别的召回率。

由表 2 可知,在不更改模型结构的前提下,模型在训练时采取 Mosaic 数据增强后, mAP 值增长了 1.55%,在此基础上更换原 Adam 优化器为 SGD 优化器, mAP 增长了 0.56%。由此可见训练时采用数据增强策略效果拔群,更换优化器也有不少提升。经分

析认为,Adam 优化器虽然能够自适应调整学习率,在训练过程中表现更好,但在测试集中表现不如 SGD 优化器,所以选用 SGD 优化器能够训练出效果更好的模型。维持上述模型训练所采取的优化策略不变,模型引入 CA 注意力机制,可观察到其 mAP 增长了 1.23%。而再引入 CIOU 损失函数, mAP 只提升了 0.18%,提升幅度有限。经分析后认为注意力机制的引入大幅提高了模型的特征提取能力,能够更加关注小目标苹果的语义信息,具体表现为 mAP 值提升明

显,而更换位置损失函数,则提高了目标边界框的回归精度,具体表现为其精度有小幅提升。

原模型相比,本文方法在单幅图片的识别时间只增加了 3 ms,识别帧数也降低了 5.59 fps,但是其 mAP 较原版增加了 1.97%,尤其是低成熟度苹果、中等成熟度苹果这两类别的目标识别率分别提高了 3.44%、3.17%;此外三个类别的目标召回率相较原模型分别提高 7.59%、7.81%、3.31%。为直观感受模型改进前后的苹果成熟度检测效果,其在测试集上的部分检测输出效果如图 8~图 10 所示。

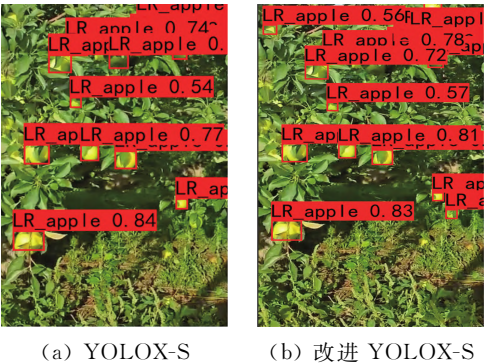


图 8 模型改进前后在验证集中识别低成熟度苹果的效果图

Fig. 8 Effects of identifying apples of low maturity in the verification set before and after model improvement

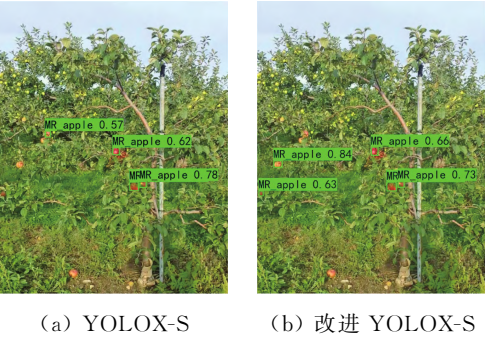


图 9 模型改进前后在验证集中识别中等成熟度苹果的效果图

Fig. 9 Effects of identifying apples of moderate maturity in the verification set before and after model improvement

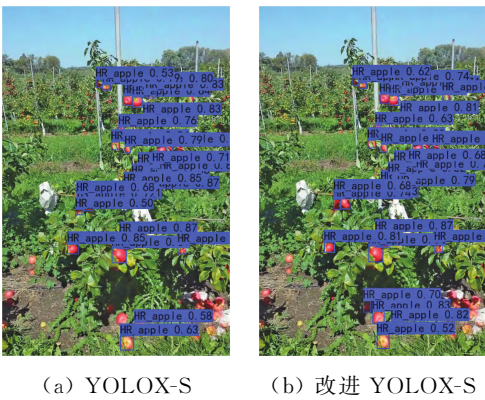


图 10 模型改进前后在验证集中识别高成熟度苹果的效果图

Fig. 10 Effects of identifying apples of high maturity in the verification set before and after model improvement

图 8~图 10 中检测到不同成熟度的苹果目标以不同的颜色方框表示,其中红色方框代表低成熟度苹果,绿色方框表示中等成熟度苹果,蓝色方框表示高成熟度苹果。方框的上方显示所检测目标类别以及置信度。

通过试验对比可看出:YOLOX-S 网络模型对图像边缘目标、小目标以及暗背景下的目标检测能力较弱;而改进后的 YOLOX-S 网络模型(本文)在复杂场景下的检测能力更强,并能检测到更多的目标物体:在暗场景下识别出高成熟度苹果目标,在图像边缘识别出低成熟度与中等成熟度的小苹果目标。由此可见本文所提出的改进 YOLOX-S 模型在苹果成熟度检测方面有着更加良好的性能与实用性。

3.4.2 改良后模型与主流目标检测模型性能对比

为进一步验证本文所提出的改进算法的实际性能表现,将其与其他主流视觉检测模型:Faster R-CNN、SSD、CenterNet、YOLOv4、YOLOv5-S,在同一测试集上进行识别对比试验,验证各模型性能表现的评估指标采用 mAP 、单幅图片识别时间以及识别帧率、低中高三种成熟度苹果各自的 AP 以及 $Recall$,最终对比试验的结果如表 3 所示。

表 3 主流视觉检测模型试验对比结果

Tab. 3 Experimental comparison results of mainstream visual inspection models

模型	mAP / %	Size / MB	识别时间/ms	识别帧数/fps	AP^L / %	AP^M / %	AP^H / %	R^L / %	R^M / %	R^H / %
FasterR-CNN	87.68	136.730	98	9.87	85.67	90.05	87.32	85.41	88.23	73.24
SSD	26.26	23.879	15	66.43	25.32	27.33	26.14	7.75	5.05	4.76
CenterNet	79.94	32.665	21	46.42	81.07	89.61	69.15	72.38	77.96	67.31
YOLOv4	38.5	63.948	38	25.88	34.63	32.43	48.42	15.25	11.37	38.04
YOLOv5-S	82.67	7.069	22	44.52	82.16	84.55	81.31	66.53	70.38	66.58
YOLOX-S(原版)	86.85	8.938	21	47.65	87.41	91.93	81.21	78.15	83.12	73.65
YOLOX-S(本文)	88.82	9.386	24	42.06	90.85	95.1	80.5	85.74	90.93	75.96

从表 3 中可看出,相较于其他主流视觉检测模型, YOLOX-S(本文)对于三种不同成熟度苹果的检测性

能表现,无论是在 AP 、 $Recall$ 、 mAP 上都有着极大的性能优势;这些模型中单幅图像识别速度最快以及帧数最高的是 SSD,其识别时间为 15 ms,帧数 66.43 fps;而识别速度最慢以及帧数最低的是 Faster R-CNN,其识别时间为 95 ms,帧数为 10.45 fps;YOLOX-S(本文)的单幅图像识别时间为 24 ms,识别帧数为 42.06 fps,能够满足真实果园场景下目标检测算法实时性的要求;综上可知本文改进的 YOLOX-S 模型在拥有高精度检测能力的同时还具有不错的实时性,与主流模型相比其检测速度与帧数均具有一定的优势。

4 结论

为更好地进行果园早期的作物负荷管理,本文提出采用改良后的 YOLOX-S 检测模型来实时观测果树上苹果真实情况,该模型能够快速且精准地识别低成熟度、中等成熟度、高成熟度的苹果并完成相应计数。

1) 改良后的 YOLOX-S 检测模型无论是与原版 YOLOX-S 检测模型还是与其他主流视觉检测模型比较,其拥有最高的 mAP 值为 88.82%。

2) 所提模型在识别图像边缘及暗场景目标时表现良好,能在保证精度的同时还具有良好的实时性,其平均识别时间只有 24 ms,识别帧数高达 42.06 fps。

后续会将该检测模型部署到嵌入式设备上,并进一步优化网络结构,以实现在移动端实时观测苹果的真实情况。

参 考 文 献

- [1] 赵春江. 智慧农业发展现状及战略目标研究[J]. 智慧农业, 2019, 1(1): 1—7.
Zhao Chunjiang. State-of-the-art and recommended developmental strategic objectives of smart agriculture [J]. Smart Agriculture, 2019, 1(1): 1—7.
- [2] 黑龙江省佳木斯农业学校, 江苏省苏州农业学校. 果树栽培学总论[M]. 北京: 中国农业出版社, 2009.
- [3] 王立扬, 张瑜, 沈群, 等. 基于改进型 LeNet-5 的苹果自动分级方法[J]. 中国农机化学报, 2020, 41(7): 105—110.
Wang Liyang, Zhang Yu, Shen Qun, et al. Automatic apple grading method based on improved LeNet-5 [J]. Journal of Chinese Agricultural Mechanization, 2020, 41(7): 105—110.
- [4] 王丹丹, 宋怀波, 何东健. 苹果采摘机器人视觉系统研究进展[J]. 农业工程学报, 2017, 33(10): 59—69.
Wang Dandan, Song Huaibo, He Dongjian. Research advance on vision system of apple picking robot [J]. Transactions of the Chinese Society of Agricultural Engineering, 2017, 33(10): 59—69.
- [5] Zhou R, Damerow L, Sun Y, et al. Using color features of cv. 'Gala' apple fruits in an orchard in image processing to predict yield [J]. Precision Agriculture, 2012, 13(5): 568—580.
- [6] 熊俊涛, 刘振, 汤林越, 等. 自然环境下绿色柑橘视觉检测技术研究[J]. 农业机械学报, 2018, 49(4): 45—52.
Xiong Juntao, Liu Zhen, Tang Linyue, et al. Visual detection technology of green citrus under natural environment [J]. Transactions of the Chinese Society for Agricultural Machinery, 2018, 49(4): 45—52.
- [7] Tian Yunong, Yang Guodong, Wang Zhe, et al. Apple detection during different growth stages in orchards using the improved YOLO-V3 model [J]. Computer & Electronics in Agriculture, 2019, 157: 417—426.
- [8] 赵德安, 吴任迪, 刘晓洋, 等. 基于 YOLO 深度卷积神经网络的复杂背景下机器人采摘苹果定位[J]. 农业工程学报, 2019, 35(3): 164—173.
Zhao De'an, Wu Rendi, Liu Xiaoyang, et al. Apple positioning based on YOLO deep convolutional neural network for picking robot in complex background [J]. Transactions of the Chinese Society of Agricultural Engineering, 2019, 35(3): 164—173.
- [9] Ge Z, Liu S, Wang F, et al. YOLOX: Exceeding YOLO series in 2021 [J]. arXiv, 2021.
- [10] Lin T Y, Dollar P, Girshick R, et al. Feature pyramid networks for object detection [C]. Proceedings The IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, New York, USA: IEEE, 2017: 2117—2125.
- [11] Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design [C]. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13713—13722.
- [12] Rezatofighi H, Tsoi N, Gwak JY, et al. Generalized intersection over union: A metric and a loss for bounding box regression [C]. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019.
- [13] Hani N, Roy P, Isler V. Minne apple: A benchmark dataset for apple detection and segmentation [J]. IEEE Robotics and Automation Letters, 2020, 5(2): 852—858.
- [14] 齐锐丽, 陈曼龙, 杨宗浩, 等. 基于 HSV 模型与改进的 OTSU 算法花椒图像分割[J]. 中国农机化学报, 2019, 40(11): 155—160.
Qi Ruili, Chen Manlong, Yang Zonghao, et al. Image segmentation of peppercorns based on HSV model and improved OTSU algorithm [J]. Journal of Chinese Agricultural Mechanization, 2019, 40(11): 155—160.
- [15] 许迎春, 刘英, 范开欣, 等. 基于线扫描及其参数空间的圆形果实识别[J]. 中国农机化学报, 2018, 39(9): 51—55.
Xu Yingchun, Liu Ying, Fan Kaixin, et al. Circular fruit recognition based on line scan and its parameter space [J]. Journal of Chinese Agricultural Mechanization, 2018, 39(9): 51—55.